Logical Omniscience and the Cost of Deliberation

Natasha Alechina and Brian Logan

School of Computer Science and IT,
University of Nottingham,
Nottingham NG8 1BB, UK,
{nza,bsl}@cs.nott.ac.uk

Abstract. Logical omniscience is a well known problem which makes traditional modal logics of knowledge, belief and intentions somewhat unrealistic from the point of view of modelling the behaviour of a resource bounded agent. We propose two logics which take into account 'deliberation time' but use a more or less standard possible worlds semantics with classical possible worlds.

1 Introduction

There has been considerable recent interest in *agent-based systems*, systems based on autonomous software and/or hardware components which perceive their environment and act in that environment in pursuit of their goals. The paradigmatic example of an agent is an autonomous robot situated in a physical environment, but there are other kinds of agents, including software agents whose environment is the Internet and synthetic characters in games and computer entertainments. Agents integrate a range of (often relatively shallow) competences, e.g., goals and reactive behaviour, emotional state and its effect on behaviour, natural language, memory and inference. As such they are central to the study of many problems in Artificial Intelligence, including modelling human mental capabilities (e.g., emotions) and performing complex tasks (e.g., those combining perception, planning, and opportunistic plan execution).

An agent can be viewed as a mapping from percepts to actions (see Fig. 1). The agent constantly monitors its environment and selects actions which allow it to achieve its goals given the current state of the environment. For example, a robot with the goal of delivering a package to an office at the end of the hall may modify its path to avoid someone who has just stepped out of an office half way down the hall.

An agent consists of three main components (e.g., [9]):

- the *agent program* implements a mapping from percepts to actions (this is sometimes called the action selection function or action composition);
- the *agent state* includes all the internal representations on which the agent program operates (this may include representations of the agent's environment and goals, the plans it has for achieving those goals, which parts of the plan have been executed and so on); and

– the *agent architecture*, a (possibly virtual) machine that makes the percepts from the agent's sensors available to the agent program, runs the agent program, updates the agent state, and executes the primitive action(s) chosen by the agent program.

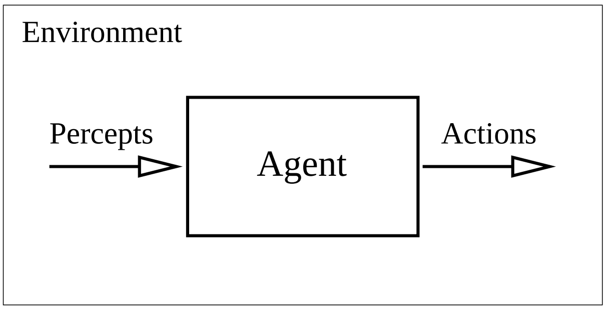

Fig. 1. An agent

Our main concern is with the agent architecture. The architecture defines the atomic operations of the agent program, and implicitly defines the components of the agent. Building a successful agent system consists largely in finding the correct architecture. There is no one correct architecture for all problems; the correct architecture depends on the task and environment.

A major focus of research in intelligent agents has therefore been to understand the implications of different agent architectures. One way to do this is empirically, by building a range of agent systems with differing architectures and conducting controlled experiments (often using simulations) to assess the relative advantages and disadvantages of each architecture. Such experiments allow the agent designer to learn more about the behaviour of a proposed system, and the agent researcher to probe the relationships between agent architectures, environments and behaviour [10]. However, conducting experiments, even in simulation, is time consuming and costly. Existing work on agent simulation is largely ad-hoc, with little re-use of simulation components and scenarios, and often fails to distinguish clearly between models of the agent and the test environment, and between these models and the simulations themselves. Agent and environment models and the simulation mechanisms are typically developed and implemented from scratch for each project or application. This limits the re-use of test scenarios, makes it difficult to reproduce previous experimental results and makes it difficult to compare architectures and implementations.

Another approach is to prove properties of the agent architecture. This means that we formalise a particular architecture in some logic and prove theorems about agent behaviour resulting from the architecture, for example: an agent with architecture X will solve a given problem faster than an agent with architecture Y; an agent with architecture Z will not be able to solve a given problem, or will not be able to solve it before the environment changes and the solution becomes irrelevant.

However most logical approaches to reasoning about agents are based on idealisations which make reasoning about agent architectures problematic. Chief among those is *logical omniscience*. The concept of logical omniscience was introduced by Hintikka in [4] and is usually defined as the agent knowing all logical tautologies and all the consequences of its knowledge. For example, the influential Belief, Desire, Intention (BDI) framework of Georgeff and Rao [8] models agents as logically omniscient. However, logical omniscience is problematic when attempting to build realistic models of agent behaviour, as closure under logical consequence implies that deliberation takes no time. If processes within the agent such as belief revision, planning and problem solving are modelled as derivations in a logical language, such derivations require no investment of computational resources by the agent. To return to the example of the package delivery robot above, when the robot becomes aware of an obstacle in the hall (e.g., from sonar data) it instantaneously revises its beliefs to update its representation of the world, making decisions about whether the obstacle is real or the result of noisy sensor data, and instantaneously decides which steps in its current plan need to be revised and derives a new plan to avoid the obstacle.

There is a significant body of work which has addressed the problem of logical omniscience from a number of different perspectives including: limiting the agent's deductive capabilities by introducing non-classical worlds in the possible worlds semantics [7, 3]; distinguishing between beliefs which can be ascribed to the agent and the agent's actual beliefs [6]; and explicitly incorporating the notion of resources [1, 13].

In this paper, we propose an alternative approach which incorporates a notion of 'delayed belief'. This has some similarities to the notion of resources in [1, 13] but our approach is developed within the context of standard possible worlds semantics. We believe that this makes it more transparent and computationally tractable. In section 2 we develop the notion of delayed belief and define two logics which formalise this notion. We prove that both logics have complete and sound axiomatisations and are decidable. In section 3 we briefly survey related work and point out similarities and differences with our approach. In section 4 we outline some open problems and sketch a program of further work.

2 Delayed Belief

In this section we consider two logics, $L_{\Delta}^{\equiv}$ and $L_{\Delta}^{\rightarrow}$. Both logics contain an operator $\Box$ which can be interpreted as standing for belief or knowledge[1]. These logics are our first attempt to incorporate the notion of computational cost (time) in reasoning about the agent's beliefs or knowledge. In $L_{\Delta}^{\equiv}$, if at the current moment an agent believes ϕ, then after a fixed delay Δ it will believe in all propositions

[1] Or any other propositional attitude where closure under logical equivalence or consequence could be expected from an ideally rational and computationally unbounded agent, but not from a realistic agent.

equivalent to ϕ. In $L_\Delta^{\rightarrow}$, after the same delay it will believe in all consequences of its beliefs at the previous step. The intuition underlying this notion of delayed belief is that an agent is able to draw inferences but needs time to derive consequences of its beliefs, so it does not believe the consequences *instantaneously.*

In both logics, none of the principles usually identified with logical omniscience (see for example [12]) is valid:

$\models \Box\phi \wedge \Box(\phi \rightarrow \psi) \rightarrow \Box\psi$ (the agent's beliefs are closed under modus ponens)
$\models \phi \Longrightarrow \models \Box\phi$ (the agent believes all tautologies)
$\models \phi \rightarrow \psi \Longrightarrow \models \Box\phi \rightarrow \Box\psi$ (the agent believes all logical consequences of its beliefs)
$\models \phi \equiv \psi \Longrightarrow \models \Box\phi \equiv \Box\psi$ (if ϕ and ψ are logically equivalent, the agent believes ϕ if and only if it believes ψ)
$\models \Box\phi \wedge \Box\psi \rightarrow \Box(\phi \wedge \psi)$ (the agent's beliefs are closed under conjunctions)
$\models \Box\phi \rightarrow \Box(\phi \vee \psi)$ (the agents beliefs are closed under weakening)
$\models \neg(\Box\phi \wedge \Box\neg\phi)$ (the agent's beliefs are consistent).

The logics $L_\Delta^{\equiv}$ and $L_\Delta^{\rightarrow}$ contain an operator Δ which stands for 'After a delay'. We make several simplifying assumptions concerning Δ. We assume that the world (atomic facts) does not change while the agent is deriving consequences of its beliefs (during the delay). We also assume that although new beliefs can be added, no beliefs can be removed from the agent's belief set. These are very strong assumptions. We discuss possible ways of overcoming them in Section 4.

2.1 $L_\Delta^{\equiv}$

The language of both logics $L_\Delta^{\equiv}$ and $L_\Delta^{\rightarrow}$ consists of a set $Prop$ of propositional variables $p, q, r, p_1, \ldots$, usual boolean connectives $\neg, \wedge, \rightarrow, \ldots$ and two unary modalities: $\Box$ which could be informally read as 'Believes' and Δ, standing for 'After a delay'. A well formed formula is defined as usual: $p|\neg\phi|\phi \wedge \psi|\Box\phi|\Delta\phi$ however we require that $\Box\phi$ is a well formed formula only if ϕ does not contain $\Box$ and Δ. We denote the set of all well formed formulas as $Form$. We denote the set of formulas which do not contain $\Box$ and Δ as $NonModForm$.

Definition 1. *The models of $L_\Delta^{\equiv}$ are structures of the form $M = \langle W, V, R, \delta\rangle$ where W is a non-empty set of possible worlds, $V : Prop \longrightarrow 2^W$ assigns subsets of W to propositional variables, $R \subseteq W \times NonModForm$ is a relation used to interpret $\Box$ and $\delta : W \longrightarrow W$ is a function (a sort of successor function) which is used to describe the next state of the world (after a delay) and interpret Δ. The satisfaction relation of a formula being true in a world in a model ($M, w \in W \models \phi$) is as follows:*

$M, w \models p \Longleftrightarrow w \in V(p)$;
$M, w \models \neg\phi \Longleftrightarrow M, w \not\models \phi$;
$M, w \models \phi \wedge \psi \Longleftrightarrow M, w \models \phi$ and $M, w \models \psi$;
$M, w \models \Box\phi \Longleftrightarrow R(w, \phi)$;
$M, w \models \Delta\phi \Longleftrightarrow M, \delta(w) \models \phi$;

There are two conditions on δ:

Frozen world *For every $p \in Prop$, $w \in V(p) \Longleftrightarrow \delta(w) \in V(p)$*
Equivalences *For every $\phi \in NonModForm$, if $R(w, \phi)$ or there exists a formula ψ such that $R(w, \psi)$ and $\vdash \phi \equiv \psi$ in classical propositional logic, then $R(\delta(w), \phi)$.*

The notions of $L^{\equiv}_{\Delta}$-valid and satisfiable formulas are standard: a formula ϕ is $L^{\equiv}_{\Delta}$-satisfiable if there exists an $L^{\equiv}_{\Delta}$-model M and a world w such that $M, w \models \phi$. A formula ϕ is $L^{\equiv}_{\Delta}$-valid ($\models \phi$) if all worlds in all models satisfy ϕ.

Consider the following axiom system (we will refer to it as $L^{\equiv}_{\Delta}$, too, in the light of the completeness theorem which follows):

Cl Classical propositional logic;
A1 $\phi \equiv \Delta\phi$ for all $\phi \in NonModForm$;
A2 $\Delta\phi \vee \Delta\neg\phi$
A3 $\neg\Delta(\phi \wedge \neg\phi)$
A4 $\Delta(\phi \to \psi) \to (\Delta\phi \to \Delta\psi)$
MP If ϕ and $\phi \to \psi$ derive ψ
R1 If $\phi \equiv \psi$ derive $\Box\phi \to \Delta\Box\psi$
R2 If ϕ derive $\Delta\phi$

We say that ϕ is derivable in $L^{\equiv}_{\Delta}$ if there is a sequence of formulas $\phi_1, \ldots, \phi_n$, each of which is either an instance of an axiom schema from $L^{\equiv}_{\Delta}$ or is obtained from the previous formulas using the inference rules of $L^{\equiv}_{\Delta}$, and $\phi_n = \phi$.

Theorem 1. *$L^{\equiv}_{\Delta}$ is complete and sound, namely $\vdash_{L^{\equiv}_{\Delta}} \phi \Longleftrightarrow \models_{L^{\equiv}_{\Delta}} \phi$*

Proof. First we give a proof of soundness: $\vdash_{L^{\equiv}_{\Delta}} \phi \Longrightarrow \models_{L^{\equiv}_{\Delta}} \phi$. All instances of the axiom schemas are obviously valid. **A1** expresses the fact that the world is 'frozen' as far as non-modal statements are concerned. **A2**-**A4** state that after a delay the world is still a classical boolean universe.

Note that $\neg(\Delta\phi \wedge \Delta\neg\phi)$ follows from **A3, A4**, $\Box\phi \to \Delta\Box\phi$ follows from **R1** and $\Delta\neg\Box\phi \to \neg\Box\phi$ follows from the previous formulas.

Next we need to show that if the premises of the rules are valid, then the conclusions are. Rule **R1** expresses the main point of $L^{\equiv}_{\Delta}$: if an agent believes ϕ and ϕ is equivalent to ψ, then after a delay the agent believes ψ. This follows from the second condition on δ. **R2** states that after a delay all tautologies are still valid.

Next we prove completeness: $\models_{L^{\equiv}_{\Delta}} \phi \Longrightarrow \vdash_{L^{\equiv}_{\Delta}} \phi$. We show that for every ϕ if $\nvdash_{L^{\equiv}_{\Delta}} \neg\phi$ then ϕ is satisfiable, that is $\not\models_{L^{\equiv}_{\Delta}} \neg\phi$.

Assume that ϕ is an $L^{\equiv}_{\Delta}$-consistent formula. In a standard way, we can show that ϕ can be extended to a maximally consistent set of formulas w_ϕ, which is a consistent set closed under $L^{\equiv}_{\Delta}$ derivability and containing either ψ or $\neg\psi$ for each $\psi \in Form$. We construct a model $M^{canonical}$ (M^c for short) satisfying ϕ as follows:

W^c is the set of all maximally consistent sets; we also require that each world is unique, in other words there are no copies of the same set;
$w \in V^c(p) \Longleftrightarrow p \in w$;
$R^c(w, \psi) \Longleftrightarrow \Box\psi \in w$;
$\delta^c(w) = \{\psi | \Delta\psi \in w\}$. In other words, $\forall \psi \in Form(\Delta\psi \in w \Longleftrightarrow \psi \in \delta(w))$.

In order to complete the proof, we need to show:

Truth Lemma: for every $\psi \in Form$ and every $w \in W^c$, $M^c, w \models \psi \Longleftrightarrow \psi \in w$.
Correctness of δ^c: for every $w \in W^c$, $\delta^c(w)$ is unique and is a maximally consistent set.
Frozen world: for every $p \in Prop$, $w \in V^c(p) \Longleftrightarrow \delta^c(w) \in V^c(p)$.
Equivalences: For every $\phi \in NonModForm$, if $R^c(w, \phi)$ or there exists a formula ψ such that $R^c(w, \psi)$ and $\vdash \phi \equiv \psi$ in classical propositional logic, then $R^c(\delta(w), \phi)$.

From the Truth Lemma, it follows that ϕ is true in w_ϕ, hence ϕ is satisfiable.
The proofs of these statements are given below.

Truth lemma. The proof goes by induction on subformulas of ψ. It is very easy for $\psi = p | \neg\psi_1 | \psi_1 \wedge \psi_2$.
Suppose $\psi = \Delta\psi_1$. Then $M^c, w \models \Delta\psi_1 \Longleftrightarrow M^c, \delta^c(w) \models \psi_1 \Longleftrightarrow \psi_1 \in \delta^c(w)$ (induction hypothesis) $\Longleftrightarrow \Delta\psi_1 \in w$ (definition of δ^c).
Suppose $\psi = \Box\psi_1$. Then $M^c, w \models \Box\psi_1 \Longleftrightarrow R^c(w, \psi_1) \Longleftrightarrow \Box\psi_1 \in w$ (definition of R^c).
Correctness of δ^c. Consistency of $\delta^c(w)$ follows from **A3**. Maximality follows from **A2, A4** and **R2**. Uniqueness follows from the fact that each $w' \in W^c$ is unique.
Frozen world $w \in V^c(p) \Longleftrightarrow p \in w \Longleftrightarrow \Delta p \in w$ (**A1**) $\Longleftrightarrow p \in \delta^c(w) \Longleftrightarrow \delta^c(w) \in V^c(p)$.
Equivalences Suppose $R^c(w, \phi)$. Then $\Box\phi \in w$. By **R2**, $\Delta\Box\phi \in w$. By definition of δ^c, $\Box\phi \in \delta^c(w)$. Hence $R^c(\delta^c(w), \phi)$.
Suppose there exists a formula ψ such that $R^c(w, \psi)$ and $\phi \equiv \psi$ is provable in classical propositional logic and hence in $L^{\equiv}_{\Delta}$. Then $\Box\psi \in w$ and $\Delta\Box\phi \in w$ by **R2**. This implies $\Box\phi \in \delta^c(w)$ so $R^c(\delta^c(w), \phi)$.

2.2 $L^{\rightarrow}_{\Delta}$

It is easy to modify $L^{\equiv}_{\Delta}$ so that an agent, instead of being able to derive all formulas equivalent to its beliefs, after a delay can derive all *consequences* of its beliefs.

Definition 2. *A model for $L^{\rightarrow}_{\Delta}$ is defined in the same way as a model for $L^{\equiv}_{\Delta}$, but replacing the* **Equivalences** *condition with the following stronger condition:*

Consequences *For every $\phi \in NonModForm$, if $R(w, \phi)$ or there exists a formula ψ such that $R(w, \psi)$ and $\vdash \psi \rightarrow \phi$in classical propositional logic, then $R(\delta(w), \phi)$.*

Theorem 2. *The following axiom system is sound and complete for* $L_{\Delta}^{\rightarrow}$*: the axioms and rules for* $L_{\Delta}^{\equiv}$ *plus*

R3 *If* $\phi \rightarrow \psi$ *derive* $\Box\phi \rightarrow \Delta\Box\psi$

(Note that **R1** *becomes derivable).*

The proof is very similar to the proof of completeness and soundness of $L_{\Delta}^{\equiv}$.

Both logics $L_{\Delta}^{\equiv}$ and $L_{\Delta}^{\rightarrow}$ are decidable and have the bounded model property. Before proving this, we need a simple lemma. Below, $Subf(\phi)$ denotes the set of all subformulas of ϕ, and $ModSubf(\phi) = \{\psi \in Subf(\phi) : \Box\psi \in Subf(\phi)\}$ are modal subformulas of ϕ.

Lemma 1. *For every* $\phi \in Form$*, and every two* $L_{\Delta}^{\equiv}$ *(*$L_{\Delta}^{\rightarrow}$*) models* $M_1 = \langle W_1, V_1, R_1, \delta_1 \rangle$ *and* $M_2 = \langle W_2, V_2, R_2, \delta_2 \rangle$*, if* $W_1 = W_2$*,* $\delta_1 = \delta_2$*,* V_1 *and* V_2 *agree on* $p \in Prop \cap Subf(\phi)$ *and* R_1 *and* R_2 *agree on* $\psi \in ModSubf(\phi)$*, then for every* w*,*

$$M_1, w \models \phi \Longleftrightarrow M_1, w \models \phi$$

Proof. The proof is just a simple induction on subformulas of ϕ.

Let us call the number of nestings of Δ operator in ϕ *Δ-depth of* ϕ, $d(\phi)$. More precisely,

$d(p) = 0$ for $p \in Prop$;
$d(\neg\psi) = d(\psi)$;
$d(\Box\psi) = d(\psi)$;
$d(\psi_1 \wedge \psi_2) = max(d(\psi_1), d(\psi_2))$;
$d(\Delta\psi) = d(\psi) + 1$.

Clearly $d(\phi) \leq |\phi|$ where $|\phi|$ is the size (number of subformulas) of ϕ. So the result below is better than usual results for modal logics obtained by filtrations which produce models of size less or equal to $2^{|\phi|}$.

Theorem 3. $L_{\Delta}^{\equiv}$ *and* $L_{\Delta}^{\rightarrow}$ *have the bounded model property, that is, if a formula* ϕ *is satisfiable then it has a model where the set of worlds is less or equal to* $d(\phi)$ *(hence less or equal to* $|\phi|$*).*

Proof. The proof is similar for both logics. We can show that if a formula ϕ of Δ-depth $d(\phi) = k$ is satisfied in a world w of a model M then it is satisfied in a model M' where the set of worlds contains only w and the worlds reachable from w in k δ-steps, i.e. $W' = \{w, \delta(w), \delta(\delta(w)), \ldots, \delta^k(w)\}$. Obviously W' is of size at most $d(\phi)$ even if W is infinite ($|W'|$ could be less than k if for some $m < k, \delta^m(w) = \delta^{m+1}(w)$).

The proof that $M, w \models \phi \Longleftrightarrow M', w \models \phi$ is standard (see for example [11], Lemma 2.8) and is omitted here.

Theorem 4. *The satisfiability problem for* $L_{\Delta}^{\equiv}$ *and* $L_{\Delta}^{\rightarrow}$ *is decidable.*

Proof. Suppose we would like to check whether a formula ϕ is satisfiable in $L_{\Delta}^{\equiv}$ ($L_{\Delta}^{\rightarrow}$). By the previous theorem, it suffices to check whether ϕ is satisfiable in any $L_{\Delta}^{\equiv}$ ($L_{\Delta}^{\rightarrow}$) model of size less or equal to $|\phi|$. The set of models of size less or equal to $|\phi|$ is strictly speaking infinite since R is defined on the set of all formulas which is infinite, so there are infinitely many models of a fixed finite size which differ in R. However, by the previous lemma the only part of R in every model which really matters for checking whether ϕ is satisfied or not is the part dealing with all subformulas of ϕ of the form $\Box\psi$. There are only finitely many different relations R with respect to the set $ModSubf(\phi)$, so we need to check only finitely many cases. Being an $L_{\Delta}^{\equiv}$ ($L_{\Delta}^{\rightarrow}$) model is a decidable property since the equivalence relation (consequence relation) on classical propositional formulas is decidable.

3 Related Work

In this section, we briefly survey previous approaches to the problem of logical omniscience and point out similarities and differences with our approach.

Hintikka [4, 5] and Rantala [7] saw the problem of logical omniscience mostly as a result of unrealistic principles in a formal model of knowledge. The solution they favoured was to make the principles invalid by changing the possible worlds semantics so that logically equivalent formulas do not necessarily hold in the same sets of possible worlds. This was achieved by introduction of 'impossible worlds' ([7]) where classical logic does not hold. Similar in spirit is the work of Fagin et al. [3] where possible worlds model a flavour of relevance logic. There again classical logical omniscience does not hold, although the agents are perfect reasoners in a weaker logic. Levesque [6] makes an important distinction between the beliefs which the agent actually has (explicit beliefs) and beliefs which can be attributed to it. The explicit beliefs do not conform to the principle of logical omniscience. Levesque's approach involves using incomplete worlds (situations). A similar but simpler and more intuitive semantics for explicit beliefs was proposed by Fagin et al. [2]. Elgot-Drapkin & Perlis [1] and Weyhrauch et al. [13] take a different approach which is concerned more with modelling the bounded resources which prevent the agent from deriving all consequences from its beliefs rather than modelling its irrationality or lack of awareness.

Our motivation is closer to the bounded-resources approach of Elgot-Drapkin and Perlis and Weyhrauch et al., in that we would like to model a rational but resource-bounded agent. However, our solution is in a traditional possible worlds setting rather than in a complex first-order theory of resources or step-logic. Unlike many other epistemic logic approaches, we distinguish between beliefs at the current moment and beliefs after the reasoner had time to consider their consequences, rather than distinguishing between implicit and explicit beliefs.

4 Discussion and Further Work

The logics $L_{\Delta}^{\equiv}$ and $L_{\Delta}^{\rightarrow}$ are simple and have attractive formal properties. However, they are far from what we actually would like to achieve. We describe them here as a proof of concept, which requires further elaboration to achieve a realistic model of agent behaviour. In this section, we briefly outline some of the ways in which the approach presented above could be extended.

First of all, we would like to make the connection between delay time and computational effort involved in deducing a formula more explicit. Although nestings of the delay operator Δ can express some of the intuitions (e.g $\Box\phi \wedge \Delta\Box(\phi \rightarrow \psi) \rightarrow \Delta\Delta\Box\psi$), it may be useful to introduce finer structure on what kind of derivations can be made after a fixed amount of time. For example, after a single unit of delay we could add all statements derivable from current beliefs in one application of an inference rule. Another possibility is to add extra expressive power to the language to allow us to explicitly mention moments of time as in [1] or available resources (e.g., inference rules) as in [13].

Another serious limitation of $L_{\Delta}^{\equiv}$ and $L_{\Delta}^{\rightarrow}$ is that we assume that the world does not change while the agent is reasoning and that the agent never has to revise its beliefs. This could be overcome by explicitly tagging particular beliefs with moments of time.

For some applications, the agent's inability to reason about its beliefs is a limitation. For example, an agent should be able to realise that it does not know whether ϕ and attempt to derive it (see [1] for more examples).

The logics we proposed only consider deductive reasoning, not default reasoning or planning. However, we believe that our approach can be extended to other kinds of deliberation.

References

[1] Elgot-Drapkin, J. J., Perlis, D.: Reasoning situated in time I: Basic concepts. *Journal of Experimental and Theoretical Artificial Intelligence*, 2:75–98, 1990.

[2] Fagin, R., Halpern, J. Y.: Belief, awareness, and limited reasoning. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 480–490, 1985.

[3] Fagin, R., Halpern, J. Y., Vardi, M. V.: A non-standard approach to the logical omniscience problem. *Artificial Intelligence*, 79(2):203–240, 1996.

[4] Hintikka, J.: *Knowledge and belief.* Cornell University Press, Ithaca, NY, 1962.

[5] Hintikka, J.: Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.

[6] Levesque, H. J.: A logic of implicit and explicit belief. In *Proceedings of the Fourth National Conference on Artificial Intelligence, AAAI-84*, pages 198–202. AAAI, 1984.

[7] Rantala, V.: Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica*, 35:106–115, 1982.

[8] Rao, A. S., Georgeff, M. P.: Modeling rational agents within a BDI-architecture. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484, 1991.

[9] Russell, S., Norvig, P.: *Artificial Intelligence: a modern approach.* Prentice Hall, 1995.

[10] Scheutz, M., Logan, B.: Affective vs. deliberative agent control. In *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing*, pages 1–10. AISB, The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, March 2001.

[11] van Benthem, J.: *Modal logic and classical logic.* Bibliopolis, 1983.

[12] van der Hoek, W., van Linder, B., Meyer, J-J. Ch.: An integrated modal approach to rational agents. In Wooldridge, M. and Rao, A. editors, *Foundations of Rational Agency*, pages 133–168. Kluwer Academic, Dordrecht, 1999.

[13] Weyhrauch, R. W., Cadoli, M., Talcott, C. L.: Using abstract resources to control reasoning. *Journal of Logic, Language and Information*, 7(1):77–101, 1998.