Belief ascription under bounded resources *

Natasha Alechina and Brian Logan
School of Computer Science
University of Nottingham
Nottingham NG8 1BB, UK
({nza,bsl}@cs.nott.ac.uk)

Abstract. There exists a considerable body of work on epistemic logics for resource-bounded reasoners. In this paper, we concentrate on a less studied aspect of resource-bounded reasoning, namely, on the ascription of beliefs and inference rules by the agents to each other. We present a formal model of a system of bounded reasoners which reason about each other's beliefs, and investigate the problem of belief ascription in a resource-bounded setting. We show that for agents whose computational resources and memory are bounded, correct ascription of beliefs cannot be guaranteed, even in the limit. We propose a solution to the problem of correct belief ascription for feasible agents which involves ascribing *reasoning strategies*, or preferences on formulas, to other agents, and show that if a resource-bounded agent knows the reasoning strategy of another agent, then its ascription of beliefs to the other agent is correct in the limit.

1. Introduction

There has been considerable work on epistemic logics for reasoners with bounded inferential abilities or bounded memory (or both), e.g., (Hintikka, 1962; Rantala, 1982; Fagin and Halpern, 1985; Konolige, 1986; Fagin et al., 1990; Elgot-Drapkin and Perlis, 1990; Halpern et al., 1994; Duc, 1995; Alechina and Logan, 2002; Ågotnes, 2004; Ågotnes and Alechina, 2006; Albore et al., 2006; Alechina et al., 2008; Artëmov and Kuznets, 2009). However this work has not considered the problem of *belief ascription* for time- and memory-bounded reasoners. In this paper we investigate multi-agent epistemic logics which result from taking seriously the idea that agents have bounded memory and their reasoning takes time. We consider agents which communicate and reason about their own beliefs and the beliefs they ascribe to other agents. Each agent has some initial beliefs (which may include beliefs about the beliefs of other agents) and some inference rules which allow it to derive new beliefs. In addition to ascribing beliefs to other agents, agents also ascribe inference rules to each other: e.g., an agent may assume that another agent can perform inference using the rule of resolution. Using such ascribed inference rules agents can derive further beliefs about the beliefs of other agents; for example, if agent 1 believes that agent 2 believes $p \vee r$ and $\neg p \vee q$, and agent

* This work was supported by the UK Engineering and Physical Sciences Research Council [grant number EP/E031226].

1 believes that agent 2 reasons using resolution, then agent 1 may ascribe belief in $r \vee q$ to agent 2.

We investigate the problem of correct belief ascription in this setting under various bounds on the agents' computational and memory resources. We consider several classes of structures which are models of a multi-agent epistemic temporal logic. The first class of structures, which we call T_0, corresponds to systems where agents can perform an unbounded number of inferences in a single moment of time and have unbounded memory. The second class of structures, T_1, bounds the agents' computational ability to applying only a single inference rule at a single moment of time, but still assumes unbounded memory. In the third class of structures, T_2, agents have both bounded computational ability and bounded memory. We give a straightforward axiomatisation of those structures. Finally, we introduce a class of structures T_3 which is similar to T_2 but has in addition a preference relation on the agent's beliefs, corresponding to the *reasoning strategy* of an agent.

We define *correct* belief ascription by agent i to agent j as follows: in each state, if agent i believes that j believes a formula α, then j indeed believes α. We show that under some natural conditions, correct ascription of beliefs requires an agent with unbounded computational ability and unbounded memory (that is, correct belief ascription is possible for the agents modelled by T_0, but not T_1 or T_2). We also show that a weaker correctness property, which we call *correctness in the limit*, holds for the agents modelled by the structures in T_1. Correctness in the limit means that for every state, if in some future state i believes that j believes α, then in some (possibly different) future state j indeed believes α. However, it turns out to be difficult to define reasonable conditions on T_2 structures which guarantee correct belief ascription (even in the limit), even if the agents' initial ascription of beliefs and inference rules to each other are correct. The problem is that the agents apply inference rules one step at a time, as in T_1 structures, and in T_2 structures they also may overwrite or forget formulas since their memory is of bounded size.

We propose a solution to the problem of belief ascription under bounded computational and memory resources which involves ascribing *reasoning strategies*, or preferences on formulas, to agents. The intuition behind preference ascription is simple; if I know that another agent can infer a formula which more closely accords with its preferences or one which is less preferred then I will assume that the other agent will derive the more preferred formula at the next step. For example, given a choice between deriving a more important formula p (such as $P \neq NP$) or a formula q (which says something trivial such as $1 = 1$), we would expect an agent which prefers more important formulas to derive p rather than q. We show that if an agent i has at least partial information regarding the reasoning strategy of agent j, then the ascription of beliefs by agent i to agent j is correct in the limit. The main contribution of this paper is in precisely formulating the problem of belief ascription under differing resource bounds and in defining a notion of reasoning strategy which allows which allows belief ascription by time- and memory-bounded reasoners which is correct in the limit.[1]

The remainder of this paper is organised as follows. In section 2, we present our model of reasoning agents, and in section 3 we say what it means for one agent to correctly ascribe beliefs to another agent. In sections 4, 5 and 6 we define structures which characterise various bounds on the agent's computation and memory and investigate their implications for belief ascription. We show that under bounded computation and memory, correct belief ascription cannot be guaranteed, even in the limit. In section 7, we sketch a solution to the problem of belief ascription under bounded resources which involves giving the agents (partial) information about the reasoning strategies of other agents. We consider related work in section 8 and conclude in section 9.

2. Model of reasoning agents

We consider a finite set of agents $A = \{1, \ldots, n\}$ which reason about their own beliefs and their beliefs about the beliefs of other agents. We assume that all agents have the same internal language where formulas are either *clauses* (disjunctions of propositional variables and their negations) denoted $c, c_1, \ldots$, or clauses prefixed by a sequence of *belief operators* B_i, e.g. $B_1 B_2 (p \vee q)$. We use $\alpha, \alpha_1, \ldots$ to range over formulas of the agents' internal language. We assume that the clauses are built from a finite set of propositional variables $PROP$ and that the set of clauses is finite (clauses do not contain duplicate disjuncts). We also assume that the depth of nesting of belief operators is bounded by some fixed number b. More precisely, the internal language Ω contains the following formulas α:

$$\alpha = \; l \mid l \vee \ldots \vee l \mid B_{i_1} \ldots B_{i_k} (l \vee \ldots \vee l)$$

where l is either a propositional variable from $PROP$ or its negation, $i_1,\ldots,i_k \in A$, and $k \leq b$. This language may seem (and in fact is) rather limited, but it has clear analogues in existing agent programming languages, where instead of clauses, beliefs are simply positive literals and the depth of the belief prefix is 0 or 1 . The clauses are assumed to be represented in a fixed normal form (no repetition of literals etc.) and relate to a fixed number of facts relevant to

[1] In (Alechina et al., 2009), a different definition of a reasoning strategy is proposed, based on an ordering of inference rules. However, while that definition is appropriate for, e.g., rule firing strategies in rule-based systems, it leads to somewhat counterintuitive results. For example, the decision whether to derive $P \neq NP$ or $1 = 1$ would depend not on the relative importance of the statements but on which inference rules are employed in the derivation of each statement.

the agent, and also have an ‘annotation’ (corresponding to the belief prefix) which allows the agent to distinguish its own beliefs from beliefs of other agents. In many agent programming languages, such annotation will have just length 1 (to represent beliefs of other agents about some fact). However, if the agent needs to introspect or to reason about other agent’s beliefs about beliefs, the prefix will need to be longer. We introduce the bound on the depth of the belief prefix in order to keep Ω finite. This assumption is crucial for the axiomatisation results later in the paper but does not affect the general point of the paper concerning the possibility of correct belief ascription.

Agents may have *a priori* beliefs about the beliefs of other agents or they may acquire beliefs about the beliefs of other agents via communication. Communication between the agents is assumed to be error free, and the agents are assumed to be sincere. For example, if agent 1 tells agent 2 that p, then agent 1 really does believe p and agent 2 believes that agent 1 believes p (agent 2 believes B_1p).

Agents reason about their own and others’ beliefs using a fixed set of inference rules. We assume for the sake of concreteness that agents reason using the rules of resolution and positive introspection and ascribe similar inference capabilities to other agents. For example, if agent 1 believes p, then it can acquire the belief B_1p by positive introspection. Similarly, if agent 1 believes that agent 2 believes $p \vee q$ and $\neg p$ (and agent 1 believes that agent 2 reasons using resolution), then agent 1 may believe that agent 2 believes q.[2] We assume that inference and communication takes time, and that the system of agents evolves synchronously. At each step, each agent can perform a (potentially unbounded) amount of computation (applying inference rules) and/or send message(s) to other agents. Inferred and communicated beliefs are added to the agents’ beliefs at the beginning of the next cycle.

Each agent has a memory, which is essentially the set of formulas the agent believes at any point in time. The meta-language which we use to talk about the agents’ beliefs contains belief operators B_i (which are the same as the operators in the agents’ internal language). We interpret $B_i\alpha$ as true if and only if α is contained in the agent i’s memory. For example, B_1B_2p is true if agent 1 has the formula B_2p in its memory. In addition to applying belief operators to formulas of the agents’ internal language, the meta-language contains the usual boolean connectives and the temporal operators EX and EF which we use to describe the evolution of the system. We assume a discrete branching model of time; $EX\phi$ means that there is a possible future where in the next moment of time, ϕ holds, and $EF\phi$ means that there is a possible future

[2] Note that the assumption that the agents reason using resolution and positive introspection (and ascribe them to each other) is not essential for the main argument of this paper. This particular set of inference rules has been chosen to make the logic concrete; we could have assumed, for example, that the agents reason using modus ponens and conjunction introduction instead of resolution, or that agents ascribe different sets of inference rules to other agents.

where at some moment of time ϕ holds. The formulas of the meta-language L are defined as follows:

$$\phi = B_i\alpha \mid \neg\phi \mid \phi \wedge \phi \mid EX\phi \mid EF\phi$$

where α is in Ω. Other boolean connectives are definable in the standard way. We also define $AX\phi$ (in all successor states) as $\neg EX\neg\phi$ and $AG\phi$ (in all future states) as $\neg EF\neg\phi$. Note that the logic only allows us to reason about the dynamics of agent's beliefs; we do not have any propositional variables referring to the state of the environment. It would not be technically difficult to introduce them, but we omit them since the point of the logics is to reason about the correctness of belief ascription by the agents, not about the correctness of agents' beliefs about the environment.

The structures corresponding to the system sketched above are pairs (S, R), where S is a non-empty set of states and R is a transition relation. The structures give rise to models of branching time temporal logic when they are unravelled into a tree structure. In what follows, we assume, without loss of generality, that the models are tree structures with a distinguished root node s_0. The (global) states of the system are n-tuples of local states of the agents. For simplicity, we can identify each local state s^i with a finite subset of Ω - intuitively, the set of formulas believed by the agent. The definition of R depends on the actions the agents can perform (given below), and on the assumptions we make about the agents' memory and computational resources. We introduce four different versions of the structures in the sequel corresponding to different assumptions about resource bounds, all of which share the same truth definition.

A formula ϕ is true in a structure M and state s, $M, s \models \phi$ iff:

$M, s \models B_i\alpha$ iff $\alpha \in s^i$

$M, s \models \neg\phi$ iff $M, s \not\models \phi$

$M, s \models \phi_1 \wedge \phi_2$ iff $M, s \models \phi_1$ and $M, s \models \phi_2$

$M, s \models EX\phi$ iff there exists an s' such that $R(s, s')$ and $M, s' \models \phi$

$M, s \models EF\phi$ iff there exists an R-path $s = s_0, \ldots, s_k$, where $k \geq 0$, such that $M, s_k \models \phi$.

The transition relation R is determined by the actions which the agents can perform. For an agent i to be able to perform an action a, the preconditions of a must hold in i's current state s^i. Executing the action updates the successor state of i and possibly the states of other agents. There are five actions corresponding to communication between agents, and applying the inference rules of resolution and positive introspection to the agent's own beliefs and the beliefs it ascribes to other agents.

The actions have the following preconditions:

Tell $tell(i, \alpha, j)$ (i telling j that α) can be executed by agent $i \neq j$ in a state s if $\alpha \in s^i$. Note that this condition means that the agents are truthful.

Resolution $res(i, c_1, c_2, l)$ can be executed by agent i in state s if $c_1, c_2 \in s^i$ and c_1 and c_2 resolve on l.

Positive introspection $int(i, \alpha)$ can be executed by agent i in state s if $\alpha \in s^i$ and the belief prefix of α has depth less than b.

Ascribed resolution $ares(i, B_{i1} \ldots B_{ik}c_1, B_{i1} \ldots B_{ik}c_2, l)$ can be executed by agent i in state s if $B_{i1} \ldots B_{ik}c_1, B_{i1} \ldots B_{ik}c_2 \in s^i$ and c_1 and c_2 resolve on l.

Ascribed positive introspection $aint(i, B_j\alpha)$ can be executed by agent $i \neq j$ in state s if $B_j\alpha \in s^i$ and the belief prefix of $B_j\alpha$ has depth less than b.

The effect of each of the several actions that an agent may perform is that some formula is added to the state of an agent. The effect of executing an action a in state s, $\mathit{eff}(a, s)$, is a pair (α, j) which has the intuitive meaning that α is added to the state of agent j in the successor state. For convenience, we will denote the resolvent of c_1 and c_2 on l by $resv(c_1, c_2, l)$. The effects of actions are then as follows:

Tell $\mathit{eff}(tell(i, \alpha, j), s) = (B_i\alpha, j)$ (in the successor state, $B_i\alpha$ is added to the state of j)

Resolution $\mathit{eff}(res(i, c_1, c_2, l), s) = (resv(c_1, c_2, l), i)$

Positive introspection $\mathit{eff}(int(i, \alpha), s) = (B_i\alpha, i)$

Ascribed resolution $\mathit{eff}(ares(i, \bar{B}c_1, \bar{B}c_2, l), s) = (\bar{B}\ resv(c_1, c_2, l), i)$ (where $\bar{B} = B_{i1} \ldots B_{ik}$)

Ascribed positive introspection $\mathit{eff}(aint(i, B_j\alpha), s) = (B_jB_j\alpha, i)$

The intuition is that in the state s' resulting from the agents executing some set of actions Ac in a state s, the local state of agent i will be exactly $s^i \cup \{\alpha \mid (\alpha, i) = \mathit{eff}(a, s) \text{ where } a \in Ac\}$.

3. Ascribing beliefs

In this section we consider the problem of belief ascription in a resource-bounded setting. First, we need some terminology:

DEFINITION 1. *The ascription of beliefs by agent i to agent j is* correct *in state s if for every formula α, if i believes $B_j\alpha$ in s, then j believes α in s. The ascription of beliefs by agent i to agent j is* correct *if it is correct in every state s.*

This is the property we would like the system to have, namely provided the agents have correct initial beliefs about other agent's beliefs, and know the inference rules other agents use, that they continue to ascribe beliefs to them correctly. Note that this is different from requiring a *complete* belief ascription, that is that one agent knows (has a correct belief about) every single belief of another agent. Such a property would clearly be unrealistic.

If plain correctness is impossible, correctness in the limit is also a useful property:

DEFINITION 2. *The ascription of beliefs by agent i to agent j is* correct in the limit *in state s,*

$$\{\alpha \mid M, s \models EFB_iB_j\alpha\} \subseteq \{\alpha \mid M, s \models EFB_j\alpha\}$$

The ascription of beliefs by agent i to agent j is correct in the limit *if it is correct in the limit in the initial state s_0.*

In other words, if in some future possible state i believes that j believes α, then in some (possibly different) future state j indeed believes α. Note that correctness entails correctness in the limit, but not vice versa.

Our model of acquiring beliefs about beliefs of other agents is simple: one agent can tell another agent something about its beliefs. We assume that the agents are sincere in that they only tell other agents that α if they really believe α. Given this model of belief acquisition, the question we want to investigate is as follows. Suppose that the agents 'start' in a state where they have correct beliefs about the beliefs of other agents, can they continue to ascribe beliefs correctly? It turns out that this depends on the assumptions we make about the bounds on the agents' computational resources.

4. Unbounded computation, unbounded memory

In this section we define structures where the agents' memory is assumed to be unbounded and where the agents have unbounded computational ability in the sense that each agent can execute all its available actions in a single transition of the system (i.e., in a single tick of time). This is not completely unbounded computational power, but is still an idealised notion given that any number of inferences can be performed in constant time. Of the three systems considered in this paper, this model of resource bounded agents is the closest to modelling the agents as logically omniscient. It is very similar to the step logic model (Elgot-Drapkin and Perlis, 1990).

We denote the set of structures where the agents can execute all of their actions in one step by T_0.

Let $Act(s^i)$ be the set of actions of i with preconditions enabled in s^i. Let $Act(s) = \bigcup_{i \in A} Act(s^i)$. Let the set of the effects of all the actions in $Act(s)$ be $\mathit{Eff}(s)$. Let $\mathit{Eff}(s^i) = \{\alpha \mid (\alpha, i) \in \mathit{Eff}(s)\}$.

DEFINITION 3. *T_0 is the set of structures where the transition relation R satisfies the following condition in each state $s = (s^1, \ldots, s^n)$:*

$$R(s, s') \text{ iff } s' = (s^1 \cup \mathit{Eff}(s^1), \ldots, s^n \cup \mathit{Eff}(s^n))$$

Note that in T_0, each state has a unique successor, so the corresponding temporal structures are linear. Note also that although we refer to the set of structures defined above as corresponding to 'unbounded' computation, the states are not closed under inference. The number of inference rules applied at each step is unbounded, but the rules are not immediately applied again to the newly derived formulas at the same time step. So these models correspond to the step logic of Perlis et al. rather than to for example Konolige's deduction model of belief (Konolige, 1986) or algorithmic knowledge (Pucella, 2006) where the set of beliefs is assumed to be deductively closed under inference, and the time taken to achieve this closure is not taken into account. It would be straightforward to introduce systems with deductively closed belief sets as well, but we do not believe that they constitute an interesting case for the study of resource-bounded belief ascription problem.

4.1. Axiomatisation of T_0

If we assume that Ω is finite (e.g., $PROP$ is finite and the depth of nesting of belief operators is bounded) it is possible to axiomatise T_0 by adding to the temporal logic axioms the *successor axioms* which exhaustively describe the transition relation.

Let $s = (s^1, \ldots, s^n)$ be a state. We define the set of belief atoms true in s as $P(s) = \{B_i\alpha \mid i \in A,\ \alpha \in s^i\}$ and the set of belief atoms which are false in s as $N(s) = \{B_i\beta \mid i \in A,\ \beta \in \Omega \setminus s^i\}$. The complete description $d(s)$ of state s is

$$d(s) = \bigwedge P(s) \wedge \bigwedge_{B_i\beta \in N(s)} \neg B_i\beta$$

Clearly, the successor state s' of s in T_0 should satisfy $\bigwedge P(s)$, but in addition, it should contain all the effects of all the actions available to each of the agents in s. For every $(\alpha, j) \in \mathit{Eff}(s)$, we need to add $B_j\alpha$ to the successor state. Let us define as $E(s)$ the set of those formulas:

$$E(s) = \{B_j\alpha \mid (\alpha, j) \in \mathit{Eff}(s)\}$$

Finally, the set of belief formulas *not* true in s' are all those formulas which were not true in s and are not the effects of the actions.

Hence the successor axiom for s in T_0 becomes:

$$d(s) \rightarrow EX(\bigwedge P(s) \wedge \bigwedge E(s) \wedge \bigwedge_{B_i\beta \in (N(s) \setminus E(s))} \neg B_i\beta)$$

THEOREM 1. *For a finite Ω, the set of valid formulas of T_0 is completely axiomatised by adding the set of successor axioms for T_0 to the axioms and inference rules below:*

A0 *Classical propositional logic*

A1 $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$ *for* $\Box \in \{AX,\ AG\}$

A2 $EF\phi \leftrightarrow (\phi \vee EX\ EF\ \phi)$

A3 $AG(\phi \rightarrow AX\phi) \rightarrow (\phi \rightarrow AG\phi)$

A4 $EX\phi \rightarrow AX\phi$

MP $\vdash\ \phi,\ \vdash\ \phi \rightarrow \psi\ \Rightarrow \vdash\ \psi$

N $\vdash \phi\ \Rightarrow \vdash\ \Box\phi$ *for* $\Box \in \{AX,\ AG\}$

Proof. We are proving weak completeness, namely that for every formula ϕ, $\vdash \phi\ \Leftrightarrow \models \phi$ (ϕ is provable iff it is valid).

Soundness is straightforward. Note that the axioms **A1** – **A3** and the rule **N** axiomatize two normal modalities AX and AG with two accessibility relations (the successor relation and 'in the future' relation), the latter being the reflexive transitive closure of the former (see, e.g., (Segerberg, 1977). Axiom **A4** is valid because the successor relation is deterministic.

To prove completeness, we build a satisfying model M_ϕ for a consistent formula ϕ. First we define a finite set of formulas $Cl(\phi)$; the states in a satisfying model are going to be maximal consistent subsets of $Cl(\phi)$. The set $Cl(\phi)$ contains all subformulas of ϕ, all formulas of the form $B_i\alpha$ where $\alpha \in \Omega$ and $i \in A$, is closed under single negations and the condition that if $EF\psi \in CL(\phi)$, then $EX\ EF\psi \in Cl(\phi)$. We define the set of states in M_ϕ to be the set of all maximal consistent subsets of $Cl(\phi)$. For every such state $s \subseteq Cl(\phi)$, we set $\alpha \in s^i$ iff $B_i\alpha \in s$. The accessibility relation R_ϕ holds between two sets of formulas s and s' iff the formula $\bigwedge_{\psi \in s} \psi \wedge EX \bigwedge_{\psi \in s'} \psi$ is consistent; essentially, the formula says that s' is the successor state of s. Similarly to the Existence lemma in the PDL completeness proof given in for example (Blackburn et al., 2001), we can show that if a formula of the form $EF\psi$ is in s, then there is a state s' such that the pair (s, s') is in the reflexive transitive closure of R_ϕ and $\psi \in s'$. This allows us to prove a truth lemma for

M_ϕ, namely that for every $\psi \in Cl(\phi)$ and every state s, $\phi \in s$ iff $M_\phi, s \models \psi$. Finally, the axiom **A4** ensures that every state s has a single successor, and the successor axioms ensure that the results of applying inference rules in s belong to this successor, so the condition

$$R_\phi(s, s') \text{ iff } s' = (s^1 \cup \mathit{Eff}(s^1), \ldots, s^n \cup \mathit{Eff}(s^n))$$

holds in M_ϕ. Since ϕ is consistent and $\phi \in Cl(\phi)$, there is a state in M_ϕ which satisfies ϕ. Although M_ϕ is not a tree model, it is trivial to unravel it into a tree model while preserving the truth of ϕ and the properties of R_ϕ. So we have built a T_0 satisfying model for ϕ. □

Note that for axiomatising T_0, the successor relation alone and the axioms **A0**, **A1**, **A4** and the rule **N** for AX would be sufficient. We include EF in the logic only in order to be able to express properties of belief ascription precisely.

4.2. Belief ascription properties of T_0

THEOREM 2. *For every M in T_0, if the ascription of beliefs by agent i to agent j is correct in s_0, then the ascription of beliefs by agent i to agent j is correct in all states.*

Proof. Assume that in s_0, for every α, if $B_iB_j\alpha$ is true then so is $B_j\alpha$. Consider some state s in the future, where $B_iB_j\beta$ holds. Agent i either was told by j that β (and we assume that the agents are sincere, so $B_j\beta$ is true), or i derived $B_j\beta$ from other correctly ascribed beliefs using ascribed resolution and ascribed positive introspection. It is clear that if preconditions of $ares$ and $aint$ hold for i, then preconditions of res and int hold for j in the same state; so it is straigtforward to show by induction on the length of path from s_0 to s that if i derives $B_j\beta$ at step k then j derives β at the same step. □

5. Bounded computation, unbounded memory

In this section we keep the assumption that the agents' memories are unbounded, but introduce a more realistic model of the amount of computation an agent can perform in a single tick of time. We define the transition relation to be composed of a tuple of agent's actions, with only one action for each agent (other bounds on the number of actions that can be performed in a single step can be formalised similarly).

For a tuple of actions $\bar{a} = (a_1, \ldots, a_n)$ by all the agents in state s, let us denote by $\mathit{Eff}(\bar{a}, s)$ the set of effects of the actions $a_1, \ldots, a_n$ executed in s. Let $\mathit{Eff}(\bar{a}, s^i) = \{\alpha \mid (\alpha, i) \in \mathit{Eff}(\bar{a}, s)\}$.

DEFINITION 4. *T_1 is the set of structures where the transition relation R satisfies the following condition in each state $s = (s^1, \ldots, s^n)$: for every tuple of actions $\bar{a} = (a_1, \ldots, a_n)$ available in s, there exists an R-successor s' of s, such that*

$$s' = (s^1 \cup \mathit{Eff}(\bar{a}, s^1), \ldots, s^n \cup \mathit{Eff}(\bar{a}, s^n)),$$

and these are the only R-successors of s.

Note that in T_1, each state may have multiple successors, so the corresponding temporal structures are branching.

5.1. AXIOMATISATION OF T_1

The axiomatisation of T_1 is similar to the axiomatisation of T_0. We can describe each state $s = (s^1, \ldots, s^n)$ by a conjunction of belief atoms and negations of belief atoms

$$d(s) = \bigwedge P(s) \wedge \bigwedge_{B_i\beta \in N(s)} \neg B_i\beta$$

Each successor state s' of s in T_1 should satisfy $\bigwedge P(s)$, but in addition, it should contain the effects of some tuple $\bar{a} = (a_1, \ldots, a_n)$ of actions available to the agents in s. For every such $\bar{a} \in \mathit{Act}(s)$ we have a different successor axiom. For every $(\alpha, j) \in \mathit{Eff}(\bar{a}, s)$, we need to add $B_j\alpha$ to the successor state. Let us define as $E(\bar{a}, s)$ the set of those formulas:

$$E(\bar{a}, s) = \{B_j\alpha \mid (\alpha, j) \in \mathit{Eff}(\bar{a}, s)\}$$

Finally, the set of belief formulas *not* true in s' are all those formulas which were not true in s and are not the effects of actions in $\bar{a}$.

Hence the successor axiom for the action $\bar{a}$ in s in T_1 is

$$d(s) \rightarrow EX(\bigwedge P(s) \wedge \bigwedge E(\bar{a}, s) \wedge \bigwedge_{B_i\beta \in (N(s) \setminus E(\bar{a},s))} \neg B_i\beta)$$

We also need to say that each successor of s is the result of applying one of the actions possible in s (recall that $\mathit{Act}(s)$ is finite):

$$d(s) \rightarrow AX \bigvee_{\bar{a} \in \mathit{Act}(s)} (\bigwedge P(s) \wedge \bigwedge E(\bar{a}, s) \wedge \bigwedge_{B_i\beta \in (N(s) \setminus E(\bar{a},s))} \neg B_i\beta)$$

THEOREM 3. *For a finite Ω, the set of valid formulas of T_1 is completely axiomatised by adding the set of successor axioms for T_1 to the axioms* **A0** – **A3** *and rules* **MP** *and* **N**.

Proof. The proof is very similar to the completeness proof for T_0. The successor axioms constrain the successor relation so that it satisfies the properties of the successor relation in T_1, and the axioms **A2** and **A3** force the accessibility relation for AG to be the reflexive transitive closure of the successor relation. □

5.2. BELIEF ASCRIPTION PROPERTIES OF T_1

Suppose the agents i and j start in a state s_0 where i correctly ascribes some beliefs to j: for every formula α, if $M, s_0 \models B_i B_j \alpha$, then $M, s_0 \models B_j \alpha$. Then if in some future state s, i derives, from correctly ascribed beliefs and using correctly ascribed reasoning rules, that j believes α, then (since we require that every possible action of j is realised in the model) at some other state s', j indeed will derive α.

THEOREM 4. *For every M in T_1, if the ascription of beliefs by agent i to agent j is correct in s_0, then the ascription of beliefs by agent i to agent j is correct in the limit.*

Proof. Assume that in s_0 for every α if $B_i B_j \alpha$ is true, then $B_j \alpha$ is true. As before, it is easy to show that if a sequence of actions for deriving $B_j \beta$ exists for agent i, then there is also a sequence of actions which agent j can perform to derive β. So from any state where a state satisfying $B_i B_j \beta$ is reachable, there is also a path to a state where $B_j \beta$ is satisfied. However, the latter may be a state corresponding to a later moment of time, or on a different branch altogether. □

Note that the agents are not guaranteed to have pointwise correct belief ascription: i may guess wrongly which inference rule j will apply next, and derive that j believes α before j actually believes α. In order to guess correctly, the agents in T_1 need to know not just the initial beliefs and the inference rules of other agents, but also their reasoning strategy; we will address this point later in the paper.

6. Bounded computation, bounded memory

In this section, we study agents that can perform a bounded amount of computation at each step and that have memory of a fixed bounded size. For simplicity, we assume that all agents have the same memory bound m, and, since the internal language of the agents does not contain conjunction, we identify m with the number of formulas the agent can believe simultaneously. This means that none of the agents can believe more than m formulas simultaneously.

When memory is bounded, we need to modify the actions available to the agents. We have two choices: either the agents cannot derive any new formulas if their memory is full; or they have to overwrite (forget) some formula which they derived earlier. We have chosen the second alternative, because it allows us to consider more interesting and realistic agents which do not stop after the first m inference steps but continue their reasoning indefinitely by re-using their memory. The preconditions and effects of the actions must change as a consequence. For example, if the number of agents n is greater than the memory size m, it is impossible for each agent to tell something to agent j, for agent j to derive some new formula, and to retain all of the resulting formulas in j's state.

We now give preconditions and effects for *tuples* of agents' actions, and provide each tuple of actions with an extra argument: a set of 'overwriting effects' of the joint action, which essentially says which formulas are removed from each agent's memory. A set of overwriting effects o is a set of pairs of the form (β, i), which intuitively says that β is to be removed from s^i. A joint action is enabled if the normal preconditions of each individual action hold, and the overwriting ensures that no agent's memory overflows as a result. For example, a resolution action by agent i with effect (c, i) together with overwriting effects $(c_1, i), (c_2, i)$ means that c will be added and c_1 and c_2 will be removed from agent i's state. On the other hand, an overwriting effect $(B_j\alpha, i)$ will cancel the effect of a tell action $tell(j, i, \alpha)$ by agent j to agent i. Note that the transition systems will have transitions corresponding to every action combined with each possible overwriting effect (provided the pair is enabled); this can be seen as modelling non-deterministic or random overwriting.

Formally, an action tuple $\bar{a} = (a_1, \ldots, a_n)$ together with a set of overwriting effects o is *enabled* if each a_i is enabled and in addition, for each s^i,

$$|(s^i \cup \{\alpha \mid (\alpha, i) \in \mathit{Eff}(\bar{a}, s)\}) \setminus \{\beta \mid (\beta, i) \in o\}| \leq m$$

The effects of $\bar{a}$ in combination with the overwriting effects o are $\mathit{Eff}(\bar{a}, o, s) = \mathit{Eff}(\bar{a}, s) \setminus o$. We define $\mathit{Eff}(\bar{a}, o, s^i) = \{(\alpha, i) \in \mathit{Eff}(\bar{a}, o, s)\}$.

DEFINITION 5. *T_2 is the set of structures where the transition relation R satisfies the following condition in each state $s = (s^1, \ldots, s^n)$: for every pair $(\bar{a}, o)$ which is enabled in s, there exists an R-successor s' of s, such that*

$$s' = (s^1 \cup \mathit{Eff}(\bar{a}, o, s^1), \ldots, s^n \cup \mathit{Eff}(\bar{a}, o, s^n)),$$

and these are the only R-successors of s.

6.1. AXIOMATISATION OF T_2

We can axiomatise T_2 using successor axioms, one for each state s, and a pair consisting of a joint action $\bar{a}$ and overwriting effects o which is enabled in s.

The belief atoms which are true in the next state after executing the action with overwriting $(\bar{a}, o)$ are: $In(\bar{a}, o, s) = (P(s) \cup E(s)) \setminus O(s)$, where $O(s) = \{B_i\beta \mid (\beta, i) \in o(s)\}$. The set of belief atoms which are false are $Out(\bar{a}, o, s) = (N(s) \setminus E(s)) \cup O(s))$.

$$d(s) \rightarrow EX(\bigwedge In(\bar{a}, o, s) \wedge \bigwedge_{B_j\beta \in Out(\bar{a},o,s)} \neg B_j\beta)$$

Let us denote the set of all joint actions with overwriting which are enabled in state s by $Act^o(s)$. To say that only those transitions are possible which correspond to actions in $Act^o(s)$, we add the axiom

$$d(s) \rightarrow AX \bigvee_{(\bar{a},o) \in Act^o(s)} (\bigwedge In(\bar{a}, o, s) \wedge \bigwedge_{B_j\beta \in Out(\bar{a},o,s)} \neg B_j\beta)$$

THEOREM 5. *For a finite Ω, the set of valid formulas of T_2 is completely axiomatised by adding the set of successor axioms for T_2 to the axioms* **A0** *–* **A3** *and rules* **MP** *and* **N***.*

Proof. Exactly the same as for T_1. (An axiomatisation for a more complex language containing modalities corresponding to memory counters is given in (Alechina et al., 2009).) □

6.2. BELIEF ASCRIPTION PROPERTIES OF T_2

For the structures in T_2, neither correct ascription not correct ascription in the limit holds, even if the agents correctly ascribe beliefs in the initial state s_0.

Consider the following example:

Agent 1 has beliefs p (that it invited agent 2 for dinner, for example) and $\neg p \vee q$ (that if agent 2 is invited, then agent 1 promises to be at home and cook the dinner), and no other beliefs which entail q. Agent 1 tells agent 2 that p and $\neg p \vee q$. In some state s, therefore, agent 2 believes $B_1 p$ and $B_1(\neg p \vee q)$ and this belief ascription is correct. However, there is a successor state s' of s where agent 1 forgets one of its original beliefs, for example p, while agent 2 remembers $B_1 p$ and $B_1(\neg p \vee q)$ and derives $B_1 q$. Then belief ascription by agent 2 to agent 1 in s' is first of all incorrect (but this may also happen in T_1 structures) but it is also incorrect in the limit: there is no time line starting in s' such that at some time in the future agent 1 believes q as agent 1 has forgotten one of the premises necessary to derive it.

This example highlights an inherent problem with modelling agents which reason about each other's beliefs in a step-wise, memory-bounded fashion. The disparity between agent 1's beliefs and the beliefs agent 2 ascribes to agent 1 at each step is due both to the fact that at most one formula is derived by each agent at any given step (and agent 2 may guess incorrectly which inference rule agent 1 is going to use), and to memory limitations which cause agents to forget formulas. An obvious alternative would be to do tentative ascription of beliefs to other agents, i.e., to conclude that the other agent will be in *one of several* possible belief states in the next state, e.g.

$$B_2B_1p \wedge B_2B_1(\neg p \vee q) \rightarrow EX(B_2((B_1p \wedge B_1(\neg p \vee q) \wedge B_1q)\vee$$

$$(\neg B_1p \wedge B_1(\neg p \vee q) \wedge \neg B_1q) \vee \ldots))$$

However, this implies that one of the agents (agent 2 in this case) has a much larger (exponentially larger!) memory and a more expressive internal language to reason about the other agent's beliefs.

It is clearly not sufficient for correct belief ascription in the resource-bounded setting described above for the reasoners to ascribe to other agents just a set of inferences rules or a logic such as KD45. They also need to be able to ascribe to other agents a *reasoning strategy* or a preference order on the set reasoning actions used by the other agents, which constrains the possible transitions of each reasoner, and which directs each agent's reasoning about the beliefs of other agents.

7. Ascribing strategies

In this section, we introduce another class of structures where the agents' choices of actions are determined by their reasoning strategies. Informally, a reasoning strategy characterises the inferences that an agent may make when in a given state.

First we need to define strategies formally. A *reasoning strategy for agent* i, $\succ_i$, is an order on the set Ω of possible beliefs. $\alpha \succ_i \beta$ means that agent i prefers deriving α to deriving β. We use the preference relation on the sets of formulas to compare potential successor states with the current state. Intuitively, agents do not make transitions to states which are less preferred than the current state — this applies both to making inferences and to 'paying attention to' formulas communicated by other agents. Only transitions resulting in a state which is at least as preferred as the current state are present in T_3 models. Recall the example given in the introduction; a state containing an interesting new formula $P \neq NP$ may be preferred to the current state, while a state where we derive $1 = 1$ at the cost of overwriting some other formula may not be. Similarly, if a fellow reasoner communicates $1 = 1$, a state where we acquire the formula $B_j(1 = 1)$ at a cost of overwriting something useful may not be preferred to the current state.

This intuition leads to the following definition of the preference relation on the local states of agents.

DEFINITION 6. *A (successor) state $s^{i\prime}$ is* preferred *to a (current) state s^i by agent i ($s^{i\prime} \succ_i s^i$) if either $s^i \subseteq s^{i\prime}$ or $|s^i \cup s^{i\prime}| > m$ (meaning that some formulas have to be overwritten), $|s^i \setminus s^{i\prime}| = |s^i \cup s^{i\prime}| - m$ (only as many formulas are overwritten as is necessary to make the resulting set of size m), and*

$$\forall \beta \in (s^i \setminus s^{i\prime})\, \forall \alpha \in s^{i\prime}\, (\alpha \succ_i \beta)$$

(all overwritten formulas are less preferred than any formula in the successor state $s^{i\prime}$).

Finally we define a class of structures where the transition relation is constrained by the preference on derived beliefs.

DEFINITION 7. *T_3 is the set of structures where each agent i has an associated preference order $\succ_i$ on Ω and the transition relation R is as in T_2, but is further restricted as follows. An action with ovewriting (a, o) is only enabled in a state $s = (s^1, \ldots, s^n)$ if it is enabled in T_2 and in addition, the state $s' = (s^{1\prime}, \ldots, s^{n\prime})$ resulting from applying (a, o) to s satisfies the condition that for every i, $s^{i\prime} \succ_i s^i$.*

Note that knowing an agent's reasoning strategy does not imply perfect information about the inferences the agent will make. However, more informative orders $\succ_i$, e.g., total orders, place greater constraints on the inferences available to an agent.

We now define what it means for an agent to ascribe a strategy to another agent. Observe that agent i can only ascribe future beliefs to agent j correctly if the beliefs agent i is reasoning about will never be overwritten. Hence, for correct (in the limit) ascription of beliefs to j, agent i needs to know some of agent j's 'essential', or most preferred beliefs: the ones which never get overwritten. Let us denote by $p(j)$ the set of agent j's most preferred beliefs such that $|p(j)| \leq m$ (where m is the bound on the size of the agents' memory) and for every $\alpha \in p(j)$ and $\beta \in \Omega \setminus p(j)$, $\alpha \succ_j \beta$. Note that if j's preference order is not total, $p(j)$ may be empty. We denote by $p(i, j)$ the set of those 'unoverwritable beliefs' of agent j that agent i knows about; note that this is not necessarily the complete set of j's most preferred beliefs, or even a non-empty set.

The structures where agent i correctly ascribes a reasoning strategy to agent j forms a subset of all T_3 structures.

DEFINITION 8. *Agent i* correctly ascribes a reasoning strategy *to agent j if:*

– $p(i, j) \subseteq p(j)$

– *$ares(i, B_j\bar{B}c_1, B_j\bar{B}c_2, l)$ by i is only enabled if $\bar{B}c \in p(i, j)$, where c is the resolvent of c_1 and c_2 on l.*

– *$aint(i, B_j\alpha)$ by i is only enabled if $B_j\alpha \in p(i, j)$.*

– *the effects of $tell(i, \alpha)$ by j are included in the overwriting effects of a joint action if $\alpha \notin p(i, j)$.*

THEOREM 6. *Let M be a structure in T_3 where agent i correctly ascribes a reasoning strategy to agent j, and correctly ascribes beliefs to j in the initial state. Then agent i's ascription of beliefs to agent j is correct in the limit.*

Proof. We need to prove that if $M, s_0 \models EF\ B_iB_j\alpha$ then $M, s_0 \models EF\ B_j\alpha$. The proof is the same as for T_1, since if $p(i, j) \subseteq p(j)$, we can ignore the possibility of formulas used in the 'derivation' of $B_j\alpha$ by i on the path witnessing $EF\ B_iB_j\alpha$ being overwritten on any derivation path taken by j. Hence there is a matching sequence of actions by j which witnesses $EF\ B_j\alpha$. □

Note that for the correctness of agent i's ascription of beliefs to agent j to be non-vacuous, we need an additional requirement that inferences involving j's beliefs be preferred to any other possible inferences in at least some states.

The intuition that in addition to applying inference rules, agents use some kind of reasoning strategy which excludes pointless inferences is, we believe, uncontroversial. It is however surprisingly difficult to formalise this intuition. The discussion above shows that given some information about another agent's strategy, a resource-bounded agent can ascribe beliefs correctly (in the limit) to the other agent. It remains an open question whether the notion of correct strategy ascription considered above is in some sense the weakest one which can guarantee correctness in the limit for belief ascription.

8. Related work

To give the reader an idea where the current proposal fits into the body of research on epistemic logics for bounded reasoners, we include a brief survey of existing approaches, concentrating on those which have influenced the work presented here.

In standard epistemic logic (see e.g., (Fagin et al., 1995; Meyer and van der Hoek, 1995) for surveys) an agent's (implicit) knowledge is modelled as closed under logical consequence. This can clearly pose a problem when using an epistemic logic to model resource-bounded reasoners, whose set of beliefs is not generally closed with respect to their reasoning rules. Various proposals to modify possible worlds semantics in order to solve this problem of logical omniscience (e.g., introducing impossible worlds as in (Hintikka, 1962; Rantala, 1982), or non-classical assignment as in (Fagin et al., 1990)) result in agent's beliefs still being logically closed, but with respect to a weaker logic.

Our work builds on another approach to solving this problem, namely treating beliefs as syntactic objects rather than propositions (sets of possible worlds). The approach goes back to Eberle (1974) and Moore and Hendrix (1979). In (Fagin and Halpern, 1985), Fagin and Halpern proposed a model of limited reasoning using the notion of awareness: an agent explicitly believes only the formulas which are in a syntactically defined awareness set. Implicit beliefs are still closed under consequence, but explicit beliefs are not, since a consequence of explicit beliefs is not guaranteed to belong to the awareness set. However, the awareness model does not give any insight into the connection between the agent's awareness set and the agent's resource limitations, which is what we try to do in this paper.[3] Konolige (1986) proposed a different model of non-omniscient reasoners, the deduction model of belief. In Konolige's approach, reasoners are parameterised with sets of rules which could, for example, be incomplete. However, the deduction model of belief still models the beliefs of a reasoner as closed with respect to reasoner's deduction rules; it does not take into account the time it takes to produce this closure, or any limitations on the agent's memory. Step logic, introduced in (Elgot-Drapkin and Perlis, 1990), gives a syntactic account of beliefs as theories indexed by time points; each application of inference rules takes a unit of time.

Yet another account of epistemic logic called algorithmic knowledge, which treats explicit knowledge as something which has to be computed by an agent, was introduced in (Halpern et al., 1994), and further developed in, e.g., (Fagin et al., 1995; Pucella, 2006). In the algorithmic knowledge approach, agents are assumed to possess a procedure which they use to produce knowledge. In later work (Pucella, 2006), this procedure is assumed to be given as a set of rewrite rules which are applied to the agent's knowledge to produce a closed set, so, like Konolige's approach, algorithmic knowledge is concerned with the result rather than the process of producing knowledge.

In (Duc, 1995; Duc, 1997) Duc proposed logics for non-omniscient epistemic reasoners which will believe all consequences of their beliefs *eventually*, after some interval of time. It was shown in (Ågotnes and Alechina, 2007) that Duc's system is complete with respect to semantics in which the set of agent's beliefs is always finite. Duc's system did not model the agents' reasoning about each others' beliefs.

[3] We also completely dispense with the notion of implicit beliefs.

Other relevant approaches where epistemic logics were given a temporal dimension and each reasoning step took a unit of time include, for example, Sierra et al. (1996), where each inference step is modelled as an action in the style of dynamic logic, and Alechina et al. (2006) which proposes a logic for verification of response-time properties of a system of communicating rule-based agents (each rule firing or communication takes a unit of time). In a somewhat different direction, Fisher and Ghidini (1999) proposed a logic where agents reason about each others beliefs, but have no explicit time or memory limit; however there is a restriction on the depth of belief nestings (context switching by the agents). Epistemic logics for bounded-memory agents were investigated in, for example, (Ågotnes, 2004; Ågotnes and Alechina, 2006; Albore et al., 2006; Alechina et al., 2008), and the interplay between bounded recall and bounded memory (ability to store strategies of only bounded size) was studied in (Ågotnes and Walther, 2007). However none of these papers concentrated on belief ascription by one resource-bounded agent to another. In (Alechina et al., 2009) the problem of belief ascription for time- and memory-bounded reasoners was introduced, however as stated earlier, they proposed a different solution for this problem, namely the introduction of a preference order on the set of all possible rule instances in the language.

It should also be pointed out that the problem studied in this paper is of correctly predicting what another agent will believe given limited information about what it currently believes (of deriving correct conclusions from correct premises), rather than a problem of belief revision (Alchourrón et al., 1985), i.e., what an agent should do if it discovers the beliefs it has ascribed to another agent are incorrect. It is also distinct from the problem of determining the consequences of information updates as studied in dynamic epistemic logic (e.g. (Baltag et al., 1998)). Adding new true beliefs in a syntactic approach such as ours is straightforward compared to belief update in dynamic epistemic logic, which interprets beliefs as sets of possible worlds. Essentially, in dynamic epistemic logic an agent acquires a new logically closed set of beliefs at the next 'step' after an announcement is made, while we model the gradual process of deriving consequences from a new piece of information (and the agent's previous beliefs).

9. Conclusion

We presented four classes of models for epistemic temporal logic, where the agents have different restrictions on computational ability and memory capacity. The agents can communicate; they can also ascribe beliefs and inference rules to each other. We have shown that even with correct initial beliefs and correct ascription of inference rules, it is impossible for resource-bounded agents to maintain correct belief ascription. We have also shown that agents can maintain belief ascription which is 'correct in the limit' if they have some information about the other agent's preferences on the set of possible beliefs.

In future work, we plan to extend out framework to consider agents reasoning about resource limitations of other agents. At the moment the agents have no way of forming beliefs about another agent's memory limit m (note that we can easily make this limit different for different agents). If the agents could represent those limitations, then one agent could infer that another agent does not believe some formula on the grounds that the latter agent's memory is bounded.

Acknowledgements

We thank the anonymous referees for their detailed and helpful comments.

References

Ågotnes, T.: 2004, 'A Logic of Finite Syntactic Epistemic States'. Ph.D. thesis, Department of Informatics, University of Bergen, Norway.

Ågotnes, T. and N. Alechina: 2006, 'Knowing minimum/maximum n formulae'. In: G. Brewka, S. Coradeschi, A. Perini, and P. Traverso (eds.): *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006)*. pp. 317–321, IOS Press.

Ågotnes, T. and N. Alechina: 2007, 'The Dynamics of Syntactic Knowledge'. *Journal of Logic and Computation* **17**(1), 83–116.

Ågotnes, T. and D. Walther: 2007, 'Towards a Logic of Strategic Ability under Bounded Memory'. In: *Proceedings of the Workshop on Logics for Resource-Bounded Agents*.

Albore, A., N. Alechina, P. Bertoli, C. Ghidini, B. Logan, and L. Serafini: 2006, 'Model-checking memory requirements of resource-bounded reasoners'. In: *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006)*. pp. 213–218, AAAI Press.

Alchourrón, C. E., P. Gärdenfors, and D. Makinson: 1985, 'On the logic of theory change: Partial meet functions for contraction and revision'. *Journal of Symbolic Logic* **50**, 510–530.

Alechina, N., M. Jago, and B. Logan: 2006, 'Modal logics for communicating rule-based agents'. In: G. Brewka, S. Coradeschi, A. Perini, and P. Traverso (eds.): *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006)*. pp. 322–326, IOS Press.

Alechina, N. and B. Logan: 2002, 'Ascribing beliefs to resource bounded agents'. In: *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, Vol. 2. Bologna, pp. 881–888, ACM Press.

Alechina, N., B. Logan, H. N. Nguyen, and A. Rakib: 2008, 'Verifying time, memory and communication bounds in systems of reasoning agents'. In: L. Padgham, D. Parkes, J. Müller, and S. Parsons (eds.): *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Vol. 2. Estoril, Portugal, pp. 736–743, IFAAMAS.

Alechina, N., B. Logan, H. N. Nguyen, and A. Rakib: 2009, 'Reasoning about other agents' beliefs under bounded resources'. In: J.-J. Meyer and J. Broersen (eds.): *Post-proceedings of KR2008-workshop on Knowledge Representation for Agents and Multi-Agent Systems (KRAMAS), Sydney, September 2008*, Vol. 5605 of *Lecture Notes in Artificial Intelligence*. pp. 1–15, Springer.

Artëmov, S. N. and R. Kuznets: 2009, 'Logical omniscience as a computational complexity problem'. In: A. Heifetz (ed.): *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2009)*. pp. 14–23.

Baltag, A., L. S. Moss, and S. Solecki: 1998, 'The logic of public announcements, common knowledge, and private suspicions'. In: *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK'98)*.

Blackburn, P., M. de Rijke, and Y. Venema: 2001, *Modal Logic*, Vol. 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.

Duc, H. N.: 1995, 'Logical Omniscience vs. Logical Ignorance on a Dilemma of Epistemic Logic.'. In: C. A. Pinto-Ferreira and N. J. Mamede (eds.): *Progress in Artificial Intelligence, 7th Portuguese Conference on Artificial Intelligence, EPIA '95, Funchal, Madeira Island, Portugal, October 3-6, 1995, Proceedings*, Vol. 990 of *Lecture Notes in Computer Science*. pp. 237–248, Springer.

Duc, H. N.: 1997, 'Reasoning about rational, but not logically omniscient, agents'. *Journal of Logic and Computation* **7**(5), 633–648.

Eberle, R. A.: 1974, 'A logic of believing, knowing and inferring'. *Synthese* **26**, 356–382.

Elgot-Drapkin, J. J. and D. Perlis: 1990, 'Reasoning Situated in Time I: Basic Concepts'. *Journal of Experimental and Theoretical Artificial Intelligence* **2**, 75–98.

Fagin, R. and J. Y. Halpern: 1985, 'Belief, Awareness and Limited Reasoning: Preliminary Report'. In: *Proceedings of the 9th International Joint Conference on Artificial Intelligence*. pp. 491–501.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi: 1995, *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Fagin, R., J. Y. Halpern, and M. Y. Vardi: 1990, 'A non-standard approach to the logical omniscience problem'. In: R. Parikh (ed.): *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Third Conference*. pp. 41–55, Morgan Kaufmann.

Fisher, M. and C. Ghidini: 1999, 'Programming Resource-Bounded Deliberative Agents'. In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI 1999)*. pp. 200–205, Morgan-Kaufmann.

Halpern, J. Y., Y. Moses, and M. Y. Vardi: 1994, 'Algorithmic Knowledge'. In: R. Fagin (ed.): *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Fifth Conference (TARK 1994)*. San Francisco: Morgan Kaufmann, pp. 255–266.

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

Konolige, K.: 1986, *A Deduction Model of Belief*. San Francisco, Calif.: Morgan Kaufmann.

Meyer, J.-J. and W. van der Hoek: 1995, *Epistemic Logic for Computer Science and Artificial Intelligence*. Cambridge University Press.

Moore, R. C. and G. Hendrix: 1979, 'Computational models of beliefs and the semantics of belief sentences'. Technical Note 187, SRI International, Menlo Park, Calif.

Pucella, R.: 2006, 'Deductive Algorithmic Knowledge'. *Journal of Logic and Computation* **16**(2), 287–309.

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

Segerberg, K.: 1977, 'A completeness theorem in the modal logic of programs'. *Notices of the American Mathematical Society* **24**, A–552.

Sierra, C., L. Godo, R. L. de Mántaras, and M. Manzano: 1996, 'Descriptive dynamic logic and its application to reflective architectures'. *Future Generation Computer Systems* **12**(2-3), 157–171.