Electronic Notes in Theoretical Computer Science 85 No. 2 (2004)
URL: http://www.elsevier.nl/locate/entcs/volume85.html 17 pages

A Logic For Ignorance

Wiebe van der Hoek

Department of Computer Science
University of Liverpool
Liverpool L69 7ZF, UK
wiebe@csc.liv.ac.uk

Alessio Lomuscio

Department of Computer Science
King's College London
London WC2R 2LS, UK
alessio@dcs.kcl.ac.uk

Abstract

We introduce and motivate a non-standard multi-modal logic to represent and reason about ignorance in Multi-Agent Systems. We argue that in Multi-agent systems being able to reason about what agents *ignore* is just as important as being able to reason about what agents know. We show a sound and complete axiomatisation for the logic. We investigate its applicability by restating the feasibility condition for the FIPA communication primitive of inform.

As we know, there are known knowns; there are things we know that we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns — the ones we don't know we don't know.

Donald Rumsfeld (US Secretary of State for Defence)

1 Introduction

Following Dennet's influential work [5] MAS are traditionally modelled by taking an *intentional stance*. This amounts to ascribing notions such as knowledge, beliefs, desires, intentions etc. to agents in order to model, and specify their behaviour. Concepts such as knowledge, beliefs, are not easy to model by means of first order logic. On the one hand they are referentially opaque, on the other they require a formalism in which operators can be arbitrarily

[1] This work was supported by the Nuffield Foundation grant NAL/00690/G.

nested one into another. It has long been argued that modal logic provides a possible solution for these problems. Indeed, many of the most important and widely used approaches to model Multi-Agent Systems (MAS) are now based on various modal logics [1].

A considerable amount of research has gone in the past 20 years into exploring the formalisations of concepts such as knowledge and beliefs in MAS. Many of the most successful theories we now have (such as the area of epistemic logic, variations of the BDI model, etc.) are based on earlier work in AI, or philosophical logic. For example, the foundations of the modern use of epistemic logic (such as the one proposed in [6]) can be found in the work of Hintikka and Aumann of the 1950s. The basis for the BDI work ([16,17]) take inspiration from the work of Bratman, and Cohen [2], [4]. This is not to say that work in MAS theories consists simply in a rediscovery exercise of previously explored ideas. The theories as they are used now are considerably more refined than they were at the time, and they are now integrated with specification and verification techniques from software engineering (witness recent progress in verification of MAS theories [9,15].

Still, while it is encouraging that the field of MAS has taken inspiration from successful theories first appeared elsewhere, it would be interesting to see whether MAS call for the use of previously unexplored concepts. One way this may happen in MAS theories is for a logic arising directly from MAS studies, and applications. In this paper we argue that this may be the case for the concept of ignorance.

Consider the typical scenario in Agent Communication Languages in which one agent queries another for information, perhaps by using the FIPA construct of *query-if*. Assuming honesty, the agent will reply it is unable to answer the query if it is *ignorant* about the value of the information it is being asked. Indeed, in a model where full cooperation is assumed, the fact that it is actually ignorant about the value of what it is being asked may be a precondition for a negative answer of an agent. Consider a similar example in which agents are exchanging data over a channel on which an intruder may be listening. A desirable property of the interaction is that the state of *ignorance* of the intruder with respect to the content of the messages is preserved. We argue that in these and other examples a key property that we want to reason about is *states of ignorance*. Note that by ignorance we do not mean the mere lack of knowledge, but something stronger. When an agent does not know a fact p, it may be that it does not know p, because in fact it knows that $\neg p$. For instance, by using the usual properties of knowledge in MAS in a particular example it is true to say that an agent does not know that the printer is connected, because in fact it knows that printer is *not* connected. We would not call ignorance this simple lack of knowledge. By state of ignorance about φ in the following we shall refer to a mental state in which the agent is unsure about the truth value of φ. So, not only the agent does not know the truth value of φ but also that of $\neg\varphi$.

The reader acquainted with the epistemic logic literature will note that it is possible to express this concept in epistemic logic by stipulating that one agent is ignorant about φ if it does not know φ and it does not know $\neg\varphi$. We argue that this is cumbersome to express in many interesting examples. For instance, the reader may try and express the concept of one agent being ignorant about the ignorance of another agent:

$$\neg K_i(\neg K_j\varphi \wedge \neg K_j\neg\varphi) \wedge \neg K_i\neg(\neg K_j\varphi \wedge K_j\neg\varphi)$$

This constitutes a typical notion in use in security when the recipient of a message is reasoning about whether or not the intruder has been able to decode the content of the message. The above looks unnecessarily complicated. Further, this complexity makes it difficult to investigate what properties ignorance should have. For example, if an agent is ignorant about φ should it be ignorant about its ignorance? Still, ignorance must be related to epistemic states — when an agent is ignorant about φ, intuitively it is because it contemplates some alternatives in which φ is true, and others in which it is false.

What we do in this paper is to build upon these very simple observations. We aim at defining ignorance as a first class citizen, investigate its properties and explore a logic that is able to represent this concept formally. Technically this will be done by means of modal logic. We shall be using a syntax that allows for a modal operator to express the notion of ignorance of an agent with respect to a formula. Semantically we shall be using the standard possible worlds epistemic interpretation — the satisfaction definition for the operator of ignorance will obviously need to be introduced. We try and provide an in-depth analysis of the logic by giving a completeness result, and we apply this analysis to a concrete example from the literature.

The rest of this paper is organised as follows. In Sections 2 and 3 we give a formal account of ignorance, establishing some of its properties. In Section 4, we investigate a richer framework where ignorance is paired with the classical operator for knowledge. In Section 5 we show how the operator of ignorance can be used to simplify the semantic definition of the communication act of inform as defined in FIPA semantics. We conclude in Section 6.

2 Ignorance: Language and Semantics

We assume familiarity with basic concepts of modal logic. We refer the reader to [1] for details. We base our discussion on the monomodal case.

2.1 Syntax

We use a very simple mono-modal language $\mathcal{L}_I$ defined as follows in BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid I\varphi$$

Other propositional operators can be defined as standard, and are to be read as usual. A formula $I\varphi$ is to be read as 'the agent is ignorant about φ', i.e., he is not aware of whether or not φ is true. As an example the formula $I\varphi \rightarrow \neg II\varphi$ is to be read as "If the agent is ignorant about φ, then it is not ignorant about it being ignorant about φ".

We ground our discussion on classically minded agents, so we assume that statements cannot both be true or false at the same time. So, to anticipate semantical considerations made clear below, the agent in order to be ignorant about φ will have to conceive at least two epistemic alternatives, one in which φ is true, and one in which φ is false.

In our concept of ignorance we do not intend to capture degrees of ignorance with respect to a formula. There is a whole spectrum of concepts that seem worth exploring. On the one side we have agents which have absolutely no information about a fact, so are in a way "truly ignorant" about it. On the other side of the spectrum we have agents which may regard a fact to be a lot more likely to be true (or false) but still contemplate the possibility of the fact being false. In our formalism we shall not be able to differentiate between these. This, and various variants of probabilistic reasoning seem worth exploring, but are left for future work.

2.2 Semantics

We use standard possible worlds semantics to give an interpretation to the language above. A model will be built on a set of epistemic alternatives (or worlds), and a relation built on these. Intuitively, like in standard modal epistemic logic, we consider two epistemic alternatives to be related if up to the agent's information they may both be models of the real situation.

Definition 2.1 [Frames, Models, and Satisfaction] A Kripke Frame $F = (W, R)$ is a tuple where W is a set of epistemic alternatives for the agent, and $R \subseteq W \times W$ is an accessibility relation. A Kripke Model $M = (F, \pi)$, is a tuple where F is a Kripke frame and $\pi : P \rightarrow 2^W$ is an interpretation for a set of propositional variables P.

Given a model M and a formula φ, we say that φ is true in M at world w, written $M, w \models \varphi$ if:

- $M, w \models p$ if $w \in \pi(P)$,
- $M, w \models \neg\varphi$ if it is not the case that $M, w \models \varphi$,
- $M, w \models \varphi \wedge \psi$ if $M, w \models \varphi$ and $M, w \models \psi$,
- $M, w \models I\varphi$ if there exist w', w'' such that $Rww', Rww'', M, w' \models \varphi,$ and $M, w'' \models \neg\varphi$.

A formula φ is *valid*, written $\models \varphi$, if it is true in every world in every model. We write $F, w \models \varphi$ to represent $M, w \models \varphi$ where M is an arbitrary model whose underlaying frame is F.

We assume the standard definitions for metalogical properties such as axiomatisation, completeness, etc. We refer to [1] for details.

Lemma 2.2 *The following formulas are valid on the class of arbitrary Kripke models.*

$A1 \quad I\varphi \leftrightarrow I\neg\varphi$

$A2 \quad I(\varphi \wedge \psi) \rightarrow (I\varphi \vee I\psi)$

Proof. We prove A1 here; for A2, we refer to Lemma 3.1. Consider an arbitrary model M. Assume $M, w \models I\varphi$. Then there must be two points w', w'' such that Rww', Rww'' and $M, w' \models \varphi, M, w'' \models \neg\varphi$. But then by definition this means that $M, w \models I\neg\varphi$. □

The properties above seem rather reasonable for ignorance. Axiom $A1$ says that being ignorant about φ is logically equivalent to being ignorant about $\neg\varphi$. Since by ignorance we mean no definite information about the truth of the object of ignorance, intuitively this is correct.

Property $A2$ is maybe best understood in its contrapositive form $(\neg I\varphi \wedge \neg I\psi) \rightarrow \neg I(\varphi \wedge \psi)$. If an agent is neither ignorant about φ, nor about ψ it is surely not ignorant about the conjunction $\varphi \wedge \psi$. This also seems reasonable.

3 A Logic for Ignorance

In this section we aim at presenting a completeness result (Theorem 3.10) for a logic of ignorance. In order to do this we need some preliminary results.

3.1 Preliminary remarks

Axiom $A2$ regulates how to distribute I over a boolean connective. Similarly, note that the following is also valid.

$$I(\varphi \rightarrow \psi) \rightarrow I\varphi \vee I\psi \tag{1}$$

As a proof, note that $I(\varphi \rightarrow \psi) \equiv I\neg(\varphi \rightarrow \psi) \equiv I(\varphi \wedge \neg\psi)$. Also, $I(\varphi \wedge \neg\psi) \rightarrow (I\varphi \vee I\neg q)$, the latter being equivalent to $I\varphi \vee I\psi$.

$$I(\varphi \vee \psi) \rightarrow (I\varphi \vee I\psi) \tag{2}$$

Since negation and implication (or, for that matter, negation and disjunction) are functionally complete, we can generalise (1) and (2) into the following lemma:

Lemma 3.1 *Let $A \subseteq P$ be a set of propositional atoms, and $b(A)$ be a Boolean function on A. Then there exist literals $\ell_1 \ldots \ell_k$ over A, such that $I(b(A)) \rightarrow \bigvee_{i \leq k} I(\ell_i)$ is valid.*

Proof. First write $b(A)$ in conjunctive normal form $\bigwedge_i d_i$, where every d_i is a disjunction of literals over A. Observe, by $A2$, that $I(\bigwedge_i d_i) \rightarrow \bigvee_i I(d_i)$. Now

note that $I(d_{i1} \vee \cdots \vee d_{i_{k_i}}) \rightarrow (Id_{i_1} \vee \cdots \vee Id_{i_{k_i}})$, where every d_{i_j} is a literal over A. □

Corollary 3.2 *Let $A \subseteq P$ be a set of propositional atoms, and $b(A)$ be Boolean function on A. Then*

$$\models I(b(A)) \rightarrow \bigvee_{a \in A} I(a)$$

Corollary 3.2 states that, in order to be ignorant about a complex formula, one must be ignorant about one of its atoms. Again, this corresponds to our intuition.

Let us now turn our attention to possible inference rules for a system of ignorance. We begin by observing that the following is sound.

$$\text{from } \vdash \varphi, \text{ infer } \vdash \neg I\varphi \tag{3}$$

Indeed, this corresponds to the intuition that an agent cannot be ignorant about propositional tautologies, and formulas following from them. Note the rule above is related to the commonly accepted rule of necessitation in epistemic logic that states that the agent knows all propositional tautologies.

A more complex inference rule that can also be shown to be sound is the following.

$$\vdash (\psi_1 \rightarrow \chi) \wedge (\chi \rightarrow \psi_2) \Rightarrow \vdash \neg I\chi \rightarrow (\neg I\psi_1 \vee \neg I\psi_2) \tag{4}$$

In fact, note that rule (3) follows from rule (4) when $\neg I\top$ is a validity (like it is in our case) by taking $\psi_1 = \psi_2 = \varphi$ and $\chi = \top$. It is easy to see that Equation 4 provides for a sound inference rule for the semantics described above. It says that if an agent is not ignorant about an interpolant χ for a stronger ψ_1 and a weaker ψ_2, that it cannot be ignorant about both the ψ_i's. If the agent is not ignorant about χ, it either thinks χ is true (in which case it also should consider ψ_2 as being true), or that χ is false (in which case it should be convinced of ψ_1's falsity as well).

3.2 A Complete System for Ignorance

We now present a system that we can show to be complete with respect to the semantics above.

Definition 3.3 The modal system **Ig** for ignorance is defined as follows:

$I0$	All instances of propositional tautologies
$I1$	$I\varphi \leftrightarrow I\neg\varphi$
$I2$	$I(\varphi \wedge \psi) \rightarrow (I\varphi \vee I\psi)$
$I3$	$(\neg I\varphi \wedge I(\alpha_1 \wedge \varphi)) \wedge \neg I(\varphi \rightarrow \psi) \wedge I(\alpha_2 \wedge (\varphi \rightarrow \psi))$ $\rightarrow (\neg I\psi \wedge I(\alpha_1 \wedge \psi))$
$I4$	$(\neg I\psi \wedge I\alpha) \rightarrow (I(\alpha \wedge \psi) \ \vee \ I(\alpha \wedge \neg\psi))$
RI	$\vdash_{\mathbf{Ig}} \varphi \ \Rightarrow \vdash_{\mathbf{Ig}} \neg I\varphi \wedge (I\alpha \ \rightarrow \ I(\alpha \wedge \varphi))$
MP	Modus Ponens
Sub	Substitution of equivalences

Observation 1 *Before we prove soundness of* **Ig***, we make the following remark about occurrences in axiom $I3$ of the form $\neg I\varphi \wedge I(\alpha \wedge \varphi)$. If such a formula is true, either the agent is sure about the truth value of φ, or of $\neg\varphi$. Moreover, the truth of $I(\alpha \wedge \varphi)$ implies that the agent regards as possible an $\alpha\wedge\varphi$-world*[2] *, which is both a φ- and an α-world. But this then implies that all the agents' conceivable worlds verify φ. Therefore the agent is sure about the value of φ. But $I(\alpha\wedge\varphi)$ also implies that the agent considers a $\neg(\alpha\wedge\varphi)$-world possible, which is also a $(\neg\alpha\vee\neg\varphi)$-world. But since all the conceivable worlds verified φ, this implies that the agent considers one world possible in which $\neg\alpha$ is true. This, together with the fact that an α-world was an alternative to the agent, gives us that the agent is ignorant about α, but knows that φ.*

Lemma 3.4 (Soundness) *The system* **Ig** *is sound with respect to the class of arbitrary Kripke models.*

Proof. Validity of $I1$ has been argued in the previous section. For $I2$, first note that this is equivalent to $\neg I\varphi \wedge I(\psi \wedge \varphi) \rightarrow I\psi$. Consider any model M. Suppose that $M, w \models (\neg I\varphi \wedge I(\psi \wedge \varphi))$. By the argument given in Observation 1, we know then that all successors of w verify φ. Since $I(\varphi \wedge \psi)$ holds at w we must also have $I\psi$ there.

We prove $I3$ and the inference rule RI. For $I3$, as before, suppose we have a state for which $M, w \models (\neg I\varphi \wedge I(\alpha_1 \wedge \varphi)) \wedge \neg I(\varphi \rightarrow \psi) \wedge I(\alpha_2 \wedge (\varphi \rightarrow \psi))$. $\neg I\varphi$ is true in w when all R-successors of w either satisfy φ or $\neg\varphi$. Since $I(\alpha_1 \wedge \varphi)$ is also true at w, we know that w must have at least an $(\alpha_1 \wedge \varphi)$-successor u. These two statements together imply that all successors of w verify φ. Moreover, we know that w has a $(\neg\alpha_1 \wedge \varphi)$-successor v. The same line of reasoning applied to $M, w \models \neg I(\varphi \rightarrow \psi) \wedge I(\alpha_2 \wedge (\varphi \rightarrow \psi))$ gives us that all successors of w satisfy ψ. Hence, there is no $\neg\psi$-successor of w, which gives $M, w \models \neg I\psi$. Moreover, in u we have $\alpha_1\wedge\psi$, and in v, $\neg\alpha_1\wedge\psi$; therefore $M, w \models I(\alpha_1 \wedge \psi)$.

[2] A world w is an α-world if α is true at w in the model under consideration.

To prove the soundness of RI, suppose $\models \varphi$. Then there cannot be a world w with a $\neg\varphi$-successor, so $\models \neg I\varphi$. Moreover, let w be an arbitrary world in an arbitrary model M for which $M, w \models I\alpha$. Since φ is true in all worlds, we get that w has a successor u for which $M, u \models (\alpha \wedge \varphi)$ and a successor v with $M, v \models (\neg\alpha \wedge \varphi)$, implying $M, v \models \neg(\alpha \wedge \varphi)$. Thus, we have $M, w \models I\alpha \rightarrow I(\alpha \wedge \varphi)$.

We begin by noting that the following is a theorem of **Ig**.

Lemma 3.5 *We have that* $\vdash_{\mathbf{Ig}} Cb$*, where* Cb *is defined as:*

$$Cb \quad (\neg I\gamma_1 \wedge I(\delta_1 \wedge \gamma_1)) \wedge (\neg I\gamma_2 \wedge I(\delta_2 \wedge \gamma_2)) \rightarrow I(\delta_2 \wedge (\gamma_1 \wedge \gamma_2))$$

Proof. Note that $\gamma_1 \rightarrow (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))$ is a theorem (*). Furthermore, let us assume the antecedent of Cb (**). We use the rule MP implicitly.

1	$\vdash \gamma_1 \rightarrow (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))$	(*)
2	$\neg I\gamma_1 \wedge I(\delta_1 \wedge \gamma_1) \ \wedge \neg I\gamma_2 \wedge I(\delta_2 \wedge \gamma_2)$	(**)
3	$\neg I(\gamma_1 \rightarrow (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))) \wedge$ $I\delta_2 \rightarrow I(\delta_2 \wedge (\gamma_1 \rightarrow (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))))$	$RI, 1$
4	$\neg I\gamma_1 \wedge I(\delta_1 \wedge \gamma_1) \wedge$ $\neg I(\gamma_1 \rightarrow (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2)))$	$2, 3$
5	$(\neg I\gamma_2 \wedge I(\delta_2 \wedge \gamma_2)) \rightarrow I\delta_2$	$I0, I2$
6	$I\delta_2$	$2, 5$
7	$I(\delta_2 \wedge (\gamma_1 \rightarrow (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))))$	$6, 3$
8	$I(\delta_1 \wedge (\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2)))$	$I3, 4, 5, 7$ (see below)
9	$\neg I(\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))$	$I3, 4, 5, 7$ (see below)
10	$\neg I\gamma_2 \wedge I(\delta_2 \wedge \gamma_2) \wedge \neg I(\gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2))$	$2, 9$
11	$I(\delta_2 \wedge (\gamma_1 \wedge \gamma_2))$	$I3, 8, 9, 10$ (see below)

In steps 8 and 9, the instances from $I3$ to be chosen are $\varphi ::= \gamma_1, \alpha_1 ::= \delta_1, \alpha_2 ::= \delta_2, \psi ::= \gamma_2 \rightarrow (\gamma_1 \wedge \gamma_2)$. In the last step, choose the following: $\alpha_2 ::= \delta_1, \psi ::= \gamma_1 \wedge \gamma_2, \varphi ::= \gamma_2, \alpha_1 ::= \delta_2$.

Note that it can be proven that the disjunctive conclusion in $I4$ is in fact exclusive.

In order to show completeness we build maximal consistent sets of formulas and show that the canonical model for the logic **Ig** can be built on these. Although the canonical model construction will be non-standard, much of the background definitions that we use are standard. In particular we assume the usual definitions for logical consistency, and maximal consistent sets. We refer to [1] for details. Since we only refer to system **Ig** in this section we shall refer

to consistency to mean **Ig**-consistency.

Lemma 3.6 *Let Γ be a maximal consistent set, containing some formula $I\alpha$. Then the set $K^\alpha(\Gamma) = \{\psi \mid \neg I\psi, I(\alpha \wedge \psi) \in \Gamma\}$ has the following properties:*

(i) *If $\neg I\psi \in \Gamma$, then either $\psi \in K^\alpha(\Gamma)$ or $\neg\psi \in K^\alpha(\Gamma)$;*

(ii) *$K^\alpha(\Gamma) \cup \{\beta\}$ is consistent, for every $I\beta \in \Gamma$.*

Proof. Note that $I\alpha \in \Gamma$, and consider $K^\alpha(\Gamma)$ as defined above.

(i) Suppose $\neg I\psi \in \Gamma$. Then by axiom $I4$ and the maximality of Γ, we have either $I(\psi \wedge \varphi) \in \Gamma$, or $I(\neg\psi \wedge \varphi) \in \Gamma$. Hence, by construction of $K^\alpha(\Gamma)$, we are done.

(ii) To arrive at a contradiction, suppose that for some β with $I\beta \in \Gamma$, $K^\alpha(\Gamma) \cup \{\beta\}$ is inconsistent. So, $\vdash (\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \neg\beta$, with $\psi_i \in K^\alpha(\Gamma), i = 1, \ldots, n$. Let us first write Ψ for $(\psi_1 \wedge \cdots \wedge \psi_n)$. Then, by RI we have $\vdash \neg I(\Psi \rightarrow \neg\beta)$ and $\vdash I\alpha \rightarrow I(\alpha \wedge (\Psi \rightarrow \neg\beta))$. Since $I\alpha \in \Gamma$. Then by maximal consistency we have that:

$$\neg I(\Psi \rightarrow \neg\beta) \wedge I(\alpha \wedge (\Psi \rightarrow \neg\beta)) \in \Gamma \tag{5}$$

In order to apply MP and axiom $I3$ to this, we need to prove that:

$$\neg I\Psi \wedge I(\alpha \wedge \Psi) \in \Gamma \tag{6}$$

Indeed this follows from property Cb. Let $\Psi_k = (\psi_1 \wedge \ldots \psi_k)$. So $\Psi = \Psi_n$. We prove by induction on k that for all $k \leq n$, $\neg\Psi_k \in \Gamma$ and $I(\alpha \wedge \Psi_k) \in \Gamma$. For $k = 1$, we know that both $\neg\psi_1 \in \Gamma$ and $I(\alpha \wedge \psi_1) \in \Gamma$ (by construction of Γ). Now let $k < n$ and suppose that $\neg\Psi_k \in \Gamma$ (i) and $I(\alpha \wedge \Psi_k) \in \Gamma$ (ii). By construction of Γ we also have $\neg I\psi_{k+1} \in \Gamma$ (iii) and $I(\alpha \wedge \psi_{k+1}) \in \Gamma$ (iv). Therefore by applying Cb we to the four statements to derive that $I(\alpha \wedge \Psi_{k+1}) \in \Gamma$ (Note that $\Psi_{k+1} = \Psi_k \wedge \psi_{k+1}$). To see that also $\neg\Psi_{k+1} \in \Gamma$, for a contradiction suppose that $\Psi_{k+1} \in \Gamma$. Axiom $I2$ would then yield that either $\Psi_k \in \Gamma$ or $\psi_{k+1} \in \Gamma$; the first conclusion is contradictory with the induction hypothesis, and the second with our conclusion (iii). This proves $\Psi_{k+1} \in \Gamma$.

Now, using (6), (5), and axiom I3 we can conclude

$$\neg I\neg\beta \wedge I(\alpha \wedge \neg\beta) \in \Gamma. \tag{7}$$

So $\neg I\neg\beta \in \Gamma$, and, by axiom $I1$, $\neg I\beta \in \Gamma$. But since by assumption $I\beta \in \Gamma$, Γ is inconsistent, contrary to the hypothesis. Therefore $K^\alpha(\Gamma)$ is consistent. □

Lemma 3.7 *The set $K^\alpha(\Gamma) = \{\psi \mid \neg I\psi, I(\alpha \wedge \psi) \in \Gamma\}$ is independent of the choice of α: if $I\alpha_1, I\alpha_2 \in \Gamma$, then $K^{\alpha_1}(\Gamma) = K^{\alpha_2}(\Gamma)$.*

Proof. Suppose $I\alpha_1, I\alpha_2 \in \Gamma$, and that $\neg I\psi, I(\alpha_1 \wedge \psi) \in \Gamma$. From this, it follows that $\neg I\psi \wedge I\alpha_1 \wedge I(\alpha_1 \wedge \psi) \in \Gamma$ (1). It remains to show that $I(\alpha_2 \wedge \psi) \in \Gamma$. Suppose not; then, by maximal consistency $\neg(I(\alpha_2 \wedge \varphi)) \in \Gamma$. But $\neg I\psi \in \Gamma, I\alpha_2 \in \Gamma$, so by $I4$ and some calculations, $I(\alpha_2 \wedge \neg\psi) \in \Gamma$. Then

by $I1$ we have that $\neg I\neg\psi \wedge I\alpha_2 \wedge I(\alpha_2 \wedge \neg\psi) \in \Gamma$ (2). We can now apply Cb to (1) and (2), giving $I(\alpha_1 \wedge (\psi \wedge \neg\psi)) \in \Gamma$, which implies $I\bot \in \Gamma$, By $I1$ this gives $I\top \in \Gamma$. But $\vdash \top$, so $\top \in \Gamma$. So by rule RI, we have $\neg I\top \in \Gamma$, which would make Γ inconsistent. □

Observe now that since system **Ig** is consistent (this follows immediately from the soundness theorem), the existence of maximal consistent sets is guaranteed by Lindembaum's Lemma. So, for any **Ig**-consistent set of formulas there exists at least a maximal extension that properly includes the elements of the set. To prove completeness we now define a rather ad-hoc canonical model, as follows.

Definition 3.8 [Canonical Model] The canonical model $M^C = (W^C, R^C, \pi^C)$ for the logic **Ig** is defined as follows:

- W^C is the set of all maximal **Ig**-consistent sets of formulas (denoted as $\Sigma, \Delta, \Gamma \ldots$).
- $R^C \subseteq W^C \times W^C$ is defined by $R^c\Gamma\Sigma$ if $\Sigma \supseteq K^\alpha(\Gamma)$ for some $I\alpha \in \Gamma$.
- $p \in \pi(\Gamma)$ if $p \in \Gamma$.

Note that in the canonical model, every world Γ has either no successors (if there is no $I\alpha \in \Gamma$), or at least two successors (if $I\alpha \in \Gamma$, then also $I\neg\alpha \in \Gamma$, and, according to Lemma 3.6 part b) together with Lindembaum's Lemma, there are at least two successors $\Sigma_1 \supseteq K^\alpha(\Gamma) \cup \{\alpha\}$ and $\Sigma_2 \supseteq K^\alpha(\Gamma) \cup \{\neg\alpha\})$.

Lemma 3.9 (Truth-Lemma) *For all formulas φ, and all* **Ig***-maximal consistent sets Γ, we have:*

$$M^c, \Gamma \models \varphi \;\Leftrightarrow\; \varphi \in \Gamma$$

Proof. We prove the above by induction on the structure of φ. We leave to the reader to verify the basic propositional connectives. We prove the induction hypothesis for the case of $I\varphi$, from this the lemma is proved.

$M^c, \Gamma \models I\psi \;\Leftrightarrow\; I\psi \in \Gamma$.

From left to right: suppose that $M^c, \Gamma \models I\psi$. By the truth-definition of I, there exist Γ_1 and Γ_2, with $R^c\Gamma\Gamma_1$ and $R^c\Gamma\Gamma_2$, such that $M^c, \Gamma_1 \models \psi$ and $M^c, \Gamma_2 \models \neg\psi$. Hence by induction hypothesis we have $\psi \in \Gamma_1, \neg\psi \in \Gamma_2$. For a contradiction suppose that $I\psi \notin \Gamma$. Then, by the maximal consistency of Γ, $\neg I\psi \in \Gamma$. So $\neg I\neg\psi \in \Gamma$ by axiom I1. But under our hypothesis Γ has successors, so by definition there exists an $\alpha : I\alpha \in \Gamma$. Then by axiom $I4$, either $I(\alpha \wedge \psi) \in \Gamma$ or $I(\alpha \wedge \neg\psi) \in \Gamma$. But this would imply that either $\psi \in K^\alpha(\Gamma)$ or $\neg\psi \in K^\alpha(\Gamma)$. But $\Gamma_1 \supseteq K^\alpha(\Gamma), \Gamma_2 \supseteq K^\alpha(\Gamma)$, and the first possibility contradicts $R^c\Gamma\Gamma_2$ (since $\neg\psi \in \Gamma_2$), and the second $R^c\Gamma\Gamma_1$ (since$\psi \in \Gamma_1$).

From right to left. By contradiction, suppose that $I\psi \in \Gamma$ and $M^c, \Gamma \not\models I\psi$. Then by the truth definitions all R^c-successors of Γ would satisfy either ψ or $\neg\psi$, hence by induction they either contain ψ, or $\neg\psi$. Suppose all R^c successors of Γ contain ψ. Since $I\psi \in \Gamma$, by axiom $I1$, we also have $I\neg\psi \in \Gamma$. By Lemma

3.7, $K^\alpha(\Gamma) \cup \{\neg\psi\}$ would then be consistent. But then, by construction of R^c, there should be a maximal consistent set $\Sigma \supseteq K^\alpha(\Gamma)$ with $\neg\psi \in \Sigma$ and $R^c\Gamma\Sigma$. However,this would mean that we found a R^c successor of Γ containing both ψ and $\neg\psi$, which leads to a contradiction. The other case can similarly be argued. So we can conclude that $M^c, \Gamma \models I\psi$.

Theorem 3.10 (Completeness) *Given system* **Ig**, *for any formula* φ *we have the following:* $\vdash_{\mathbf{Ig}} \varphi$ *if and only if* $\models \varphi$.

Proof. Soundness was shown in Lemma 3.4. Completeness follows in the usual way from the Truth-Lemma: suppose that $\nvdash \varphi$, then $\neg\varphi$ is **Ig**-consistent, giving a **Ig**-maximal consistent set Γ with $\neg\varphi \in \Gamma$. By Lemma 3.9 then, we have $M^c, \Gamma \models \neg\varphi$, so that $\not\models \varphi$. □

Note that the completeness result above applies to the general class of Kripke frames. This contrasts with system K being the standard system to axiomatise arbitrary frames for a standard modality.

4 Ignorance and Knowledge

Now that we have a result for a basic system for ignorance we can ask the question of how this relates to what is known already in epistemic logic [6]. After all the semantics that we have used is based on the one for epistemic logic: we regard two points as related if the agent considers the two as epistemically indistinguishable. What we have done so far amounts to using this semantic concept to express ignorance as opposed to knowledge. But since intuitively it must be possible to build a correspondence between the two concepts, the curious reader must then be left wondering whether ignorance can in fact be precisely expressed in terms of knowledge. Crucially, one must consider the question of whether one could have ultimately proven Theorem 3.10 by a careful translation of epistemic operators from the usual modal systems used for epistemic logic such as S5. We explore this and other questions in the rest of this section.

Let us first ask the question of whether our ignorance operator can in fact be expressed in terms of knowledge, not just by using the semantics as we have done but also syntactically. We have taken being ignorant about φ to mean that the agent conceives two opposite alternatives for φ. In the usual epistemic language this not just implies, but is semantically equivalent to saying that the agent does not know φ, and does not know $\neg\varphi$ (an alternative definition would come from exploiting $K\varphi \equiv \neg I\varphi \wedge \varphi$, but this would assume reflexivity of the epistemic relations as will become clearer later on):

$$I\varphi \equiv (\neg K\varphi \wedge \neg K\neg\varphi).$$

We have been unable to prove a completeness result simply by using this equivalence. Indeed, surprisingly little is known about logics for modalities that are defined from others, the only constructions available resulting from

work in algebraic modal logic. One may try coding epistemic axioms for K in view of the definition above to deduce properties for I but this attempt is hindered with technical difficulties. Even if this exercise were to be successful, we would be left with a logic for I that assumes S5 as the underlaying model for K. $S5$ is often a good model for knowledge, but at other times it is useful to consider weaker models. In the way we proceeded we have made no assumption on the properties of the underlaying relations between points, and the resulting framework is a rather weak one. Indeed, we believe this to be a good feature of the logic — we have a rather weak system to begin with and we can add more properties to it if required. Given this one can define a logic for both K and I by taking the standard satisfaction definition for both operators.

We also point out that K cannot be defined in terms of I. Consider the frames of Figure 1; note that, by induction, for all formulas φ in the language for ignorance $\mathcal{L}_I$, we have

$$F_1, w_1 \models \varphi \text{ iff } F_2, w_2 \models \varphi \text{ iff } F_3, w_3 \models \varphi$$

Indeed, note that for every w_i ($i \leq 3$), no $I\varphi$ can be true: the agent considers too few alternatives possible to have any doubt whatsoever. From this, it immediately follows that K cannot be defined in terms of I, since we have for instance $F_3, w_3 \models K\bot$, but $F_1, w_1 \not\models K\bot$. In other words: K can distinguish between frames that I cannot.

4.1 Nested Ignorance

We now analyse the consequences of considering additional properties for the accessibility relations, thereby producing stronger systems then **Ig**. Let us start with transitive relations; these define positively introspective agents (in the sense of the knowledge they have, i.e., agents whose knowledge satisfies axiom 4: $K\varphi \rightarrow KK\varphi$ holds). By definition a positively introspective agent cannot be ignorant about what it knows. In fact we would have the axiom:

$I4 \quad \neg I\psi \rightarrow \neg I \neg I\psi.$

Note that axiom $I4$ is, in the context of **Ig** equivalent to $II\psi \rightarrow I\psi$.

Can an agent be ignorant about one's ignorance? Clearly, a negatively introspective agent cannot be. This would suggest the validity of the axiom:

$I5 \quad I\psi \rightarrow \neg II\psi.$

In line with epistemic logic we would expect properties $I4$ and $I5$ to impose transitivity and Euclidicity, respectively, on the canonical model. Let us first note that $I4$ is indeed true on transitive models.

Lemma 4.1 *$\mathcal{F}_4 \models I4$, where $\mathcal{F}_4$ is the class of transitive frames.*

Proof. *Consider an arbitrary model built on an arbitrary frame, and suppose*

$M, w \models I\neg I\psi$, or, equivalently, $M, w \models II\psi$. Then there are u and v for which Rwu, Rwv and $M, u \models I\psi$ and $M, v \models \neg I\psi$. The former implies that there are u_1 and u_2 with Ruu_1, Ruu_2 and $M, u_1 \models \psi$, $M, u_2 \models \neg\psi$. Since R is transitive, we have Rwu_1, Rwu_2, and hence $M, w \models I\psi$. □

However, it is illustrative to see that the converse does not hold: validity of $I4$ on a frame does not guarantee transitivity. Consider the frame with three worlds w, u, v, such that Rwu and Ruv. In w, for no formula ψ, $I\psi$ is true (since w has only one successor). Hence, $I4$ is valid in w, a point in a non-transitive frame. As for $I5$ and Euclidicity, one can do a similar analysis: $I5$ is valid on all Euclidean frames, but validity of $I5$ does not force the underlying frame to be Euclidean.

One way to proceed to achieve completeness results for stronger systems then **Ig** is to start from the semantic properties and see whether there are formulas that correspond to transitivity and Euclidicity. Here, we can use the insight of Section 3.2, i.e., that we can interpret $\neg I\varphi \wedge I\alpha \wedge I(\alpha \wedge \varphi)$ as $\neg K\alpha \wedge \neg K\neg\alpha \wedge K\varphi$. To give an indication of the kind of axiomatisations that one would have by doing so, we show the result for transitivity.

Lemma 4.2 *Consider the following axiom scheme:*

$$G4 \quad I\alpha \rightarrow [(\neg I\varphi \wedge I(\varphi \wedge \alpha)) \rightarrow \neg I(\neg I\varphi \wedge I(\alpha \wedge \varphi)) \wedge I(\neg I\varphi \wedge I(\varphi \wedge \alpha) \wedge \alpha)]$$

Then, **Ig** $\cup\{G4\}$ *is sound and complete with respect to transitive models.*

Proof. We show that the canonical model is transitive. Completeness follows by standard consideration; soundness can routinely be checked. Suppose $R^c\Gamma\Delta$ and $R^c\Delta\Sigma$. Since Γ has successor, there must be some $I\alpha \in \Gamma$. Suppose $\varphi \in K(\Gamma)$: to prove that $\varphi \in \Sigma$. By definition of $K(\Gamma)$, we have $(\neg I\varphi \wedge I(\varphi \wedge \alpha)) \in \Gamma$. By $G4$, we conclude that $\neg I(\neg I\varphi \wedge I(\alpha \wedge \varphi)), I(\neg I\varphi \wedge I(\varphi \wedge \alpha) \wedge \alpha) \in \Gamma$. By definition of $K(\Gamma)$ then, $\neg I\varphi \wedge I(\alpha \wedge \varphi) \in K(\Gamma)$ and hence $\neg I\varphi \wedge I(\alpha \wedge \varphi) \in \Delta$. By $A4$, we have $I\alpha \in \Delta$, hence $\varphi \in K(\Delta)$, and thus $\varphi \in \Sigma$. □

We leave axiomatisations of other classes of frames for further work.

We conclude by stressing that we are using a modal logic quite dissimilar to the one that the reader may be familiar with. For example we have that reflexivity is not definable by means of operator I only. This can be checked by using Figure 1 again: were reflexivity definable with the $\mathcal{L}_I$-formula ψ, then we would have $F_2, w_2 \models \psi$. However, we already observed that then also $F_1, w_1 \models \psi$, which would imply that F_1 is reflexive as well, at w_1, which it obviously is not.

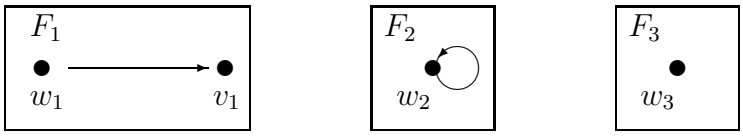

Fig. 1. Three worlds without any ignorance

5 Another look at FIPA's feasibility precondition

Logic is used in MAS and AI in a variety of topics ranging from negotiation to specification of behaviour. Of particular interest is the role that theories of beliefs and intention play in the area of communication languages. It has been argued long ago that one possible way to give a semantics to the speech act `inform` in agent communication languages is by means of pre-conditions and post-conditions. In particular, let us consider FIPA's pre-condition (called "feasibility condition" in [7]) for an inform speech act $< i, inform(j, \varphi) >$, i.e., an act in which agent i informs agent j of the formula φ. The feasibility condition is given as:

$$B_i\varphi \wedge \neg B_i(Bif_j\varphi \vee Uif_j\varphi)$$

There are three different operators at play here. B stands for "belief". Bif stands for "believe whether". Uif stands for "uncertain whether". All these are indexed with the agent for which they are referred to. So the formula above is read as: "agent i believes that φ, and it is not the case that agent i believes that either agent j believes whether or not φ is the case, or that agent j is uncertain about φ". The last term in the original formulation is meant to have a fuzzy interpretation of the kind "j suspects φ may be true but he is not sure about it.".

FIPA's specification does not define formally the way these operators are to be interpreted. In the following we try to do this by using the formal machinery of this paper. Roughly speaking, the above makes two requirements.

(i) Agent i believes φ.

(ii) Agent i believes that agent j is not aware of the truth value of φ.

1) is a sincerity condition - the agent would not send false information without violating its specification. 2) requires that agents are not sending information that they believe is redundant. Given the difficulties of expressing formally the notion of being biased towards φ we do not incorporate it into our reading. This model is best suited for cooperative systems where benevolent agents aim at distributing correct information. We do recognise our translation above may be simply an approximation of what is intended in the specification; but given that the semantics of the operators of Bif and Uif is not given, this could still prove beneficial. In this interpretation, the feasibility precondition above can be formally expressed in the logical language of this

paper as:

$$B_i\varphi \wedge B_i I_j \varphi$$

In the above B is a KD45 operator defined like knowledge, but for which the relation does not enjoy reflexivity, and I is the operator built on the same relation, and discussed in this paper. While this attempt does not incorporate the operator of uncertainty, the above does seem to capture the intuition of FIPA's specifiers, and its semantics can be given formally by using the machinery of this paper — something that is currently not possible to do for any precondition of FIPA's speech acts. Completeness results for logics in which interaction between the two operators occur can now be investigated.

6 Conclusions

We have argued that the concept of an agent being ignorant is worth investigating further, and suggested examples from MAS as to why this is the case. We showed that this analysis can be carried out independently from the commonly adopted logics for knowledge. Semantic definitions for a non-standard modal operator, and completeness results have been presented. We would contend that the technical results of the paper offer some insights into the possibility of expressing other operators that are not defined on the set of accessible points as it is traditionally done in mainstream modal logic.

Some technical questions are still left open at this stage. As we have seen in Figure 1 there may be frames that satisfy the same formulas but are not bisimilar. It would be interesting to study what technical concept is relevant here for this notion. Since our language is in many respects similar to that of *graded modalities* (cf. [10]), the notion of generalised n-bisimulation introduced there might provide a hint to find such a similarity. From a purely epistemic logic point of view, the connection with logics for *only knowing* ([8]) seems promising. The idea of saying that an agent only knows φ is that he knows φ and all of its consequences ($K_i\psi_1$, if $\vdash \varphi \rightarrow \psi_1$), but he is ignorant about any stronger formula $I\psi_2$, if $\vdash \psi_2 \rightarrow \varphi$.

More broadly we feel there is much scope for further work with respect to connections to specific MAS areas. One avenue we like to investigate is the application of this work to MAS security. The formalisms resulting from the refinements of BAN logic [3] are concerned with proving that particular protocols are secure, i.e., that any intruder would not be able to decrypt the messages being exchanged. In the language of this paper this means that its state of *ignorance* is an *invariant* in the execution. BAN logic in its standard form suffer from the lack of semantics and is purely an axiomatic system. Maybe the machinery of this paper can be used to solve this problem.

The semantics of the speech act `inform` presented above seems also a promising area for further development. Irrespective of whether the particular translation given here captures the actual intuition of FIPA's specifiers, the

operator of ignorance seems to be a useful ingredient for a definition that can actually be interpreted on a formal semantics. The interest here is not purely theoretical — having a clear semantics is an essential ingredient to move to compliance testing and verification [18]. This is something that is lacking in all FIPA's implementations at present.

References

[1] P. Blackburn, M. de Rijke, and Y. Venema. Modal logic. Cambridge University Press, 2001.

[2] M. E. Bratman. What is intention? In P. R. Cohen, J. L. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 15–32. The MIT Press: Cambridge, MA, 1990.

[3] M. Burrows, M. Abadi, and R. Needham. A logic of authentication. *ACM Transactions on Computer Systems*, 8(1):18–36, Feb. 1990.

[4] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–261, Mar. 1990.

[5] D. Dennet. *The Intentional Stance*. MIT Press, 1987.

[6] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.

[7] FIPA: Foundation for intelligent physical agents. http://www.fipa.org.

[8] J. Halpern. Theory of knowledge and ignorance for many agents. *Journal of Logic and Computation*, 7(1):79–108, 1997.

[9] W. van der Hoek and M. Wooldridge. Model checking knowledge and time. In *SPIN 2002 – Proceedings of the Ninth International SPIN Workshop on Model Checking of Software*, Grenoble, France, April 2002.

[10] W. van der Hoek. On the semantics of graded modalities. *Journal of Applied Non Classical Logics*, 2(1):81–123, 1992.

[11] W. van der Hoek and E. Thijsse. A general approach to multi-agent minimal knowledge: with tools and samples. *Studia Logica*, 72(1):61–84, 2002.

[12] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, New York, 1996.

[13] A. Lomuscio. *Knowledge Sharing among Ideal Agents*. PhD thesis, School of Computer Science, University of Birmingham, Birmingham, UK, June 1999.

[14] J.-J. C. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1995.

[15] W. Penczek and A. Lomuscio. Verifying Epistemic Properties of multi-agent systems via model checking. Fundamenta Informaticae, volume 55(2), 2002.

[16] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484. Morgan Kaufmann Publishers, Apr. 1991.

[17] A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–343, June 1998.

[18] M. Wooldridge. Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1):9—31, February 200.