Multi-Agent Only Knowing on Planet Kripke

Guillaume Aucher, Vaishak Belle

Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Multi-Agent Only Knowing on Planet Kripke

Guillaume Aucher
University of Rennes 1
INRIA
Rennes, France
guillaume.aucher@irisa.fr

Vaishak Belle*
Dept. of Computer Science
KU Leuven
Leuven, Belgium
vaishak@cs.kuleuven.be

Abstract

The idea of only knowing is a natural and intuitive notion to precisely capture the beliefs of a knowledge base. However, an extension to the many agent case, as would be needed in many applications, has been shown to be far from straightforward. For example, previous Kripke frame-based accounts appeal to proof-theoretic constructions like canonical models, while more recent works in the area abandoned Kripke semantics entirely. We propose a new account based on Moss' characteristic formulas, formulated for the usual Kripke semantics. This is shown to come with other benefits: the logic admits a group version of only knowing, and an operator for assessing the epistemic entrenchment of what an agent or a group only knows is definable. Finally, the multi-agent only knowing operator is shown to be expressible with the cover modality of classical modal logic, which then allows us to obtain a completeness result for a fragment of the logic.

1 Introduction

Suppose a modeler were to provide a collection of logical sentences Σ as a knowledge base to characterize an agent. One would expect that the beliefs of the agent would be exactly those that follow from Σ.[1] However, in classical epistemic logic [Fagin *et al.*, 1995], $K\alpha$ does not preclude $K(\alpha \wedge \beta)$ from holding in general. So unless Σ carefully includes both the beliefs and non-beliefs of the agent, it is not the case the $K\Sigma$ can succinctly characterize all the beliefs of the agent.

In this regard, the idea of only knowing is a natural and intuitive notion to precisely capture the beliefs of a knowledge base. First introduced by Levesque [1990], it enriches classical epistemic logic with a new operator O, the idea being Op holds precisely when the worlds considered possible by the agent are all and only those where p holds. So, $Op \supset Kp$ and more interestingly, $Op \supset \neg Kq$. For knowledge-based agents, in particular, this seems like the right kind of modality to include in our language. And indeed, it is very closely related to important concepts such as *minimal knowledge* [Halpern and Moses, 1984]. The main difference being that the logic of only knowing includes the O modality in the language, whereas in [Halpern and Moses, 1984], knowledge minimization is purely a meta-theoretic concept; so, only knowing has some advantages.

Somewhat surprisingly, extending only knowing's simple semantics to the many agent case has been far from straightforward. Independently, Halpern [1993] and Lakemeyer [1993] attempted extensions, but these were shown to exhibit unintuitive properties [Halpern and Lakemeyer, 2001]. In later work, Halpern and Lakemeyer [2001] do manage to capture only knowing, but by appealing to proof-theoretic constructs such as canonical models in the semantics. (Canonical models, moreover, which are based on sets of maximally consistent sets of formulas, are also perhaps not realizable in practice.) The approach of Waaler and Solhaug [2005] was based on model-theoretic constraints, which as the authors admit, "is complex and hard to penetrate". Finally, in more recent work, Belle and Lakemeyer [2010] show how the proof-theoretic construction of [Halpern and Lakemeyer, 2001] can be avoided to naturally capture multi-agent only knowing, but at the cost of introducing a semantics that significantly deviates from the classical (that is, Kripke) account.

The purpose of this paper is to revisit the notion of multi-agent only knowing, but to phrase the truth conditions in terms of the usual Kripke framework in a natural way, that is, by avoiding canonical models and other proof-theoretic machinery. There are several reasons why this is being attempted. Formulating the account in a more familiar truth theory has the advantage that deep results known in other areas of modal logic can finally be put in the context of only knowing. Indeed, as one would observe, the basis for our reconstruction is Moss' investigation into normal forms [Moss, 2007], which itself is inspired by and builds on the seminal work of Fine [1975]. Second, we show how the multi-agent only knowing modality developed here can be expressed in terms of the *cover modality* [D'Agostino and Lenzi, 2005], which is intimately connected to co-algebras and their role in central results on modal logic, such as interpolation and Beth definability. In the longer term, a multi-agent only know-

*Partially funded by the FWO project on Data Cleaning and KU Leuven's GOA on Declarative Modeling for Mining and Learning.

[1] In this work, we do not differentiate between "knowledge" and "belief", and use these terms interchangeably.

ing framework in classical modal logic would be more accessible for dynamic epistemic logic based knowledge representation languages [Demolombe, 2003; Herzig *et al.*, 2000; van Ditmarsch *et al.*, 2011]. In sum, only knowing is arguably an essential companion to the classical knowledge operator in AI applications, and the work considered here would allow modal logicians to use it more readily.

The paper is organized as follows. In Section 2, we recall the essentials of epistemic logic and characteristic formulas, as introduced by Moss, together with some related results used in the sequel. Then, in Section 3, we define our logic with our specific truth conditions for the multi-agent only knowing connectives which use the characteristic formulas. In Section 4, we compare and relate formally our logic and our definition of multi-agent only knowing with the recent approach proposed by Belle and Lakemeyer [2010]. In Section 5, we present the cover modality and establish formally its connection with our multi-agent only knowing modality. This allows us in Section 6 to axiomatize the validities of our logic once it is extended with the cover modality. We discuss other related efforts in Section 7 before concluding.

2 Preliminaries

In this section, we first recall the basics of epistemic logic. Then, we introduce the characteristic formulas for modal logic as defined by Moss and recall a number of results about these formulas that will be used in the rest of the article.

2.1 Epistemic Logic

In the rest of the paper, $\mathcal{P}$ is a set of propositional letters called *atomic facts* and $\mathcal{A}$ is a finite set whose elements are called *agents*.

Definition 1 (Language $\mathcal{L}$). We define the language $\mathcal{L}$ inductively as follows.

$$\varphi \ ::= \ p \ \mid \ \neg\varphi \ \mid \ (\varphi \wedge \varphi) \ \mid \ K_j\varphi \ \mid \ C_G\varphi$$

where $G \subseteq \mathcal{A}$, $j \in \mathcal{A}$ and $p \in \mathcal{P}$. In the sequel, we use the following abbreviations:

$$E_G\varphi \doteq \bigwedge_{j\in G} K_j\varphi; \qquad \widehat{K_j}\varphi \doteq \neg K_j\neg\varphi; \qquad \widehat{C_G}\varphi \doteq \neg C_G\neg\varphi$$

Let $\varphi \in \mathcal{L}$. The *modal depth of* φ, denoted $d(\varphi)$, is defined inductively as follows: $d(p) = 0, d(\varphi \wedge \psi) = max\{d(\varphi), d(\psi)\}, d(\neg\varphi) = d(\varphi), d(K_j\varphi) = d(\varphi) + 1$ and $d(C_G\varphi) = d(\varphi) + 1$. The set of propositional letters appearing in φ is denoted $P(\varphi)$.

The formula $K_j\varphi$ reads as "agent j *K*nows φ". Dually, the formula $\widehat{K_j}\varphi$ reads as "agent j considers it possible that φ holds". The formula $E_G\varphi$ reads as "everybody in group G knows that φ holds". Common knowledge of φ means that everybody knows that φ but also that everybody knows that everybody knows φ, that everybody knows that everybody knows that everybody knows φ, and so on *ad infinitum*. Formally, this corresponds to the following formula $C_G\varphi := E_G\varphi \wedge E_GE_G\varphi \wedge E_GE_GE_G\varphi \wedge \ldots$, and is infinitary in nature. But as we do not allow infinite conjunction, the common knowledge operator is introduced as a primitive connective in our language.

PL	Axioms and Inference Rules of Prop. Logic
Dist	$K_j(\varphi \to \psi) \to (K_j\varphi \to K_j\psi)$
E	$E_A\varphi \leftrightarrow \bigwedge_{j\in A} K_j\varphi$
Mix	$C_A\varphi \to E_A(\varphi \wedge C_A\varphi)$
Ind	if $\varphi \to E_A(\psi \wedge \varphi)$ then $\varphi \to C_A\psi$

Figure 1: Proof System L for $\mathcal{L}$

A (pointed) epistemic model (M, w) represents how the actual world represented by w is perceived by the agents. Atomic facts are used to state properties of this actual world.

Definition 2 (Epistemic model). An *epistemic model* is a tuple $M = (W, R, V)$ where:

- W is a non-empty set of *possible worlds*,
- $R : G \to 2^{W\times W}$ is a function assigning to each agent $j \in G$ an *accessibility relation* on W,
- $V : \Phi \to 2^W$ is a function assigning to each propositional letter of Φ a subset of W. V is called a *valuation*.

We write $w \in M$ for $w \in W$, and (M, w) is called a pointed epistemic model (w often represents the actual world). If $w, v \in W$, we write wR_jv for $R(j)(w, v)$. Finally, we write

$$R_j(w) := \left\{ v \in W \;\middle|\; wR_jv \right\}$$

$$R_G(w) := \left\{ v \in W \;\middle|\; v \in \left(\bigcup_{j\in G} R_j \right)^+ (w) \right\}.$$

Intuitively, in the definition above, $v \in R_j(w)$ means that at w, the agent j believes that v might be the real world.

The truth conditions of $K_j\varphi$ are defined in such a way that $K_j\varphi$ holds in a possible world when φ holds in *all* the worlds agent j considers possible.

Definition 3 (Truth conditions). Let M be an epistemic model, $w \in M$ and $\varphi \in \mathcal{L}$. Then, $M, w \models \varphi$ is defined inductively as follows:

$M, w \models p$	iff	$w \in V(p)$
$M, w \models \neg\varphi$	iff	$M, w \not\models \varphi$
$M, w \models \varphi \wedge \psi$	iff	$M, w \models \varphi$ and $M, w \models \psi$
$M, w \models K_j\varphi$	iff	for all $v \in R_j(w)$, we have $M, v \models \varphi$
$M, w \models C_G\varphi$	iff	for all $v \in R_G(w)$, we have $M, v \models \varphi$

We write $M \models \varphi$ when $M, w \models \varphi$ for all $w \in M$, and $\models \varphi$ when for all epistemic models M, $M \models \varphi$. In that latter case, we say that φ is L-*valid*.

The following theorem shows that the set of validities of $\mathcal{L}$ can be axiomatized by the proof system $\mathcal{L}$.

Theorem 1. *The proof system* L *for* $\mathcal{L}$ *defined in Figure 1 is sound and strongly complete for* $\mathcal{L}$ *w.r.t. the class of epistemic models.*

2.2 Characteristic Formulas

We will resort in our definitions and proofs to particular kinds of modal formulas which capture the structure of epistemic models (modulo bisimulation) up to a given modal depth. These formulas were defined by Moss [2007] and are very similar to the normal forms for modal logic as introduced by Fine [1975].[2]

In what follows, we use S and E, possibly decorated with superscripts and subscripts, to denote sets of formulas. The subscripts refer to constructions using inductive definitions. When the superscript is an agent index j or the group G, it means that the formulas in the set are in the context of K_j and C_G respectively.

Definition 4. [Moss, 2007] Let $P \subseteq \mathcal{P}$ be finite. We inductively define the sets E^P_n as follows:

$$E^P_0 = \left\{ \bigwedge_{p \in S_0} p \wedge \bigwedge_{p \in P - S_0} \neg p \;\middle|\; S_0 \subseteq P \right\}$$

$$E^P_{n+1} = \left\{ \delta_0 \wedge \bigwedge_{j \in \mathcal{A}} \left(\bigwedge_{\delta \in S^j_n} \widehat{K_j}\delta \wedge K_j \left(\bigvee_{\delta \in S^j_n} \delta \right) \right) \wedge \bigwedge_{G \subseteq \mathcal{A}} \left(\bigwedge_{\delta \in S^G_n} \widehat{C_G}\delta \wedge C_G \left(\bigvee_{\delta \in S^G_n} \delta \right) \right) \;\middle|\; \delta_0 \in E^P_0 \text{ and } S^j_n, S^G_n \subseteq E^P_n \right\}.$$

An element δ of E^P_n with $n > 0$ will often be written as follows (note that the S^j_n are replaced by $R_j(\delta)$ and the S^G_n are replaced by $R_G(\delta)$):

$$\delta = \delta_0 \wedge \bigwedge_{j \in G} \left(\bigwedge_{\delta_j \in R_j(\delta)} \widehat{K_j}\delta_j \wedge K_j \bigvee_{\delta_j \in R_j(\delta)} \delta_j \right) \wedge \bigwedge_{G \subseteq Agt} \left(\bigwedge_{\delta_G \in R_G(\delta)} \widehat{C_G}\delta_G \wedge C_G \bigvee_{\delta_G \in R_G(\delta)} \delta_G \right).$$

Basically, a characteristic formula δ_{n+1} provides a complete syntactic representation of a pointed epistemic model up to modal depth $n+1$. So, intuitively, if we view a characteristic formula δ_{n+1} of E^P_{n+1} as the syntactic representation up to modal depth $n+1$ of a possible world w where it holds, a formula δ_n of S^j_n can also be viewed as a syntactic representation up to modal depth n of a possible world accessible by R_j from w.

The following proposition not only tells us that a formula δ_n completely characterizes the structure up to modal depth n of any pointed epistemic model where it holds (first item), but also that the structure of *any* epistemic model up to modal depth n can be characterized by such a δ_n (second item).

Proposition 1. *[Moss, 2007] Let $P \subseteq \mathcal{P}$ be finite, let $n \in \mathbb{N}$ and let $\delta \in E^P_n$. Let $\varphi \in \mathcal{L}$ such that $d(\varphi) \leq n$ and $P(\varphi) \subseteq P$. Then, the following hold:*

Either $\models \delta \rightarrow \varphi$ or $\models \delta \rightarrow \neg\varphi$.

$$\models \bigvee_{\delta \in E^P_n} \delta.$$

The following corollary is also used in the sequel. It states that any formula can be reduced to a disjunction of δs. (Thus, they are referred to as *normal forms.*) The decomposition of a formula φ into δs syntactically (and precisely) captures the relevant structure of the set of pointed epistemic models which make φ true. Put differently, each δ can be seen as a syntactic description of the modal structure (up to depth n and modulo bisimulation) of a pointed epistemic model which makes φ true.

Corollary 1. *Let $\varphi \in \mathcal{L}$ and let $k \in \mathbb{N}$. Let $P := P(\varphi)$ and let $n = d(\varphi) + k$. Then, there is $S^\varphi_n = \{\delta^1_n, \ldots, \delta^m_n\} \subseteq E^P_n$ such that*

$$\models \varphi \leftrightarrow \bigvee_{\delta \in S^\varphi_n} \delta.$$

Moreover, for each n, this set S^φ_n is unique.

3 Defining Multi-Agent Only Knowing

We define our logic for multi-agent only knowing in two steps. First, we define a language where we cannot nest the multi-agent only knowing connectives. Then, we generalize it to allow for an arbitrary nesting of these connectives.

3.1 A Constrained Logic

The class of models for our first logic is the class of epistemic models. Only the syntax of the logical language and the truth conditions change, because the language is extended with only knowing operators. Our first language is in fact an extension of the language $\mathcal{ONL}^-_n$ of Belle and Lakemeyer [2010], in the sense that no connector O^k_j or O^k_G may occur in the scope of a K_j, a O^k_j or a O^k_G.

Definition 5 (Language $\mathcal{L}^O_{G-}$). We define the languages $\mathcal{L}^O_{G-}$ inductively as follows.

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_j\varphi \mid O^k_j\psi \mid C_G\varphi \mid O^k_G\psi$$

where $\psi \in \mathcal{L}$, $k \in \mathbb{N}$, $G \subseteq Agt$, $j \in \mathcal{A}$, and $p \in \mathcal{P}$. The formulas $O_j\varphi$ and $O_G\varphi$ are abbreviations for $O^0_j\varphi$ and $O^0_G\varphi$ respectively.

The formula $O^k_j\psi$ reads "the agent j only believes (knows) ψ up to degree k" and the formula $O^k_G\psi$ reads "the group of agents G only believes (knows) ψ up to degree k".

Definition 6 (Truth conditions for $\mathcal{L}^O_{G-}$). The satisfaction relation $\models$ between pointed epistemic models and formulas of $\mathcal{L}^O_{G-}$ is defined inductively as follows. The basic case as well as the cases for the connectives $\neg, \wedge, K_j$ and C_G are defined like in Definition 3. We only define the cases for O^k_j and O^k_G. Let M be an epistemic model, $w \in M$ and let $\psi \in \mathcal{L}$. Let

[2] Halpern and Lakemeyer [2001] discuss a similar normal form in the context of only knowing.

$n = k + d(\psi)$ and let $X \in \mathcal{A} \cup 2^{\mathcal{A}}$. Then,

$$M, w \models O_X^k \psi \quad \text{iff} \quad \begin{array}{l} \text{for all } v \in R_X(w) \text{ there is } \delta \in S_n^\psi \\ \text{such that } M, v \models \delta, \\ \text{and for all } \delta \in S_n^\psi \text{ there is } v \in R_X(w) \\ \text{such that } M, v \models \delta \end{array}$$

We recall that the finite set $S_n^\psi = \{\delta_1, \ldots, \delta_m\} \subseteq E_n^P$ (where $P = P(\psi)$) is such that $\models \psi \leftrightarrow \bigvee_{\delta \in S_n^\psi} \delta$.

Let $\varphi \in \mathcal{L}_{G-}^O$. We write $M \models \varphi$ when $M, w \models \varphi$ for all $w \in M$, and $\models \varphi$ when for all epistemic models M, $M \models \varphi$.

The intuition underlying our definition of the truth condition for our multi-agent only knowing operator can be explained as follows. As noted in the previous section, each δ of S_n^ψ can be seen as a syntactic description of the modal structure (up to depth n and modulo bisimulation) of a pointed epistemic model which makes ψ true. So, the agent (or the group of agents) only knows ψ if she considers possible *all and only* the possible worlds that make ψ true, disregarding the structure of these worlds after a certain depth k: the bigger k is, the more entrenched this knowledge will be. Note that if we restrict our setting to the propositional case with a single agent, then we recover the original definition of only knowing as defined by Levesque [1990].

Importantly, note that we do not need to resort to an operator $N_j\varphi$ like in the usual definitions of multi-agent only knowing. Instead, we refer in the semantics to characteristic formulas and our only knowing operator is introduced here as a primitive connective.

Example 1. One would think that this constrained language is little more than the single agent version, that is, formulas such as $O_i p \supset \neg K_i q$ are the only interesting valid sentences in this logic. In fact, this language can be used to capture *multi-agent autoepistemic defaults* [Lakemeyer, 1993].[3] Suppose p denotes a secret, and consider the default assumption that if i has no knowledge of j knowing the secret, then it is the case that j does not know it. Let $\delta = \neg K_i K_j p \supset \neg K_j p$.

It can now be shown that if δ is the only formula that i only knows, then, $O_i \delta \supset K_i \neg K_j p$ is a valid sentence. That is, i actually believes that j does not know p. Note, for example, adding objective knowledge to i's knowledge does not change his conclusion, that is, $\models O_i(p \wedge \delta) \supset K_i \neg K_j p$. Of course, if i where to believe that j indeed knows p, this conclusion is now retracted: $\models O_i(p \wedge \delta \wedge K_j p) \supset K_i K_j p$.

3.2 A Logic of Multi-Agent Only Knowing

As in the logic of the previous section, the class of models is again the class of (pointed) epistemic models. Only the syntax of the logical language changes, the satisfaction relation and the class of models remains the same as before. The syntax of this full language is, in fact, an extension of the language $ON\mathcal{L}_n$ of Belle and Lakemeyer [2010]. Like for $ON\mathcal{L}_n$, we also allow an arbitrary nesting of the connectors K_j, O_j^k and O_G^k.

[3] Levesque [1990] had previously shown that only knowing a subjective formula can capture such defaults in the single agent case.

Definition 7 (Languages $\mathcal{L}_G^O$ and $\mathcal{L}^O$). The language $\mathcal{L}_G^O$ is defined inductively as follows.

$$\varphi \;::=\; p \;\mid\; \neg\varphi \;\mid\; (\varphi \wedge \varphi) \;\mid\; K_j\varphi \;\mid\; O_j^k\varphi \;\mid\; C_G\varphi \;\mid\; O_G^k\varphi$$

where $k \in \mathbb{N}$, $G \subseteq Agt$, $j \in \mathcal{A}$ and $p \in \mathcal{P}$. The formulas $O_j\varphi$ and $O_G\varphi$ are abbreviations for $O_j^0\varphi$ and $O_G^0\varphi$ respectively. We define the language $\mathcal{L}^O$ as the language $\mathcal{L}_G^O$ without the group connectives C_G and O_G^k.

Proposition 2. *Let $k \in \mathbb{N}$ and $\psi \in \mathcal{L}$. Let $P := P(\psi)$ and $n = d(\psi)+k$. Then by Corollary 1, there is $S_n = \{\delta_n^1, \ldots, \delta_n^m\} \subseteq E_n^P$ such that $\models \psi \leftrightarrow \bigvee_{\delta_n \in S_n} \delta_n$. Then, we have that the following holds:*

$$\models O_j^k\psi \leftrightarrow \bigwedge_{\delta \in S_n} \widehat{K_j}\delta \wedge K_j\left(\bigvee_{\delta \in S_n} \delta\right) \tag{1}$$

and

$$\models O_G^k\psi \leftrightarrow \bigwedge_{\delta \in S_n} \widehat{C_G}\delta \wedge C_G\left(\bigvee_{\delta \in S_n} \delta\right). \tag{2}$$

Proof. It follows straightforwardly from the truth conditions for O_j^k and O_G^k. QED

Corollary 2. *Let $k \in \mathbb{N}$ and $\varphi \in \mathcal{L}_G^O$. Let $P := P(\varphi)$ and $n = d(\varphi) + k$. Then, there is $S_n^\varphi = \{\delta_n^1, \ldots, \delta_n^m\} \subseteq E_n^P$ such that*

$$\models \varphi \leftrightarrow \bigvee_{\delta \in S_n} \delta.$$

Moreover, for each n, this set S_n^φ is unique.

Proof. The proof is by induction on the nesting depth n of operators O_j^k and O_G^k. The base case $n = 0$ holds trivially by definition. The induction step is proved by applying Proposition 2: the right-hand sides of Expressions (1) and (2) do not contain operators O_j^k and O_G^k. QED

Definition 8 (Truth conditions for $\mathcal{L}_G^O$). The truth conditions are exactly the same as in Definition 6. When $\models \varphi$ holds for some $\varphi \in \mathcal{L}_G^O$, we say that φ is $\mathcal{L}^O$*-valid*.

Example 2. With the enriched language, we can handle entailments of the sort:

$$\models K_i(p \wedge O_j q) \supset K_i K_j q \wedge K_i \neg K_j p$$

That is, if i believes p and also believes j to only know q, he not only knows that j knows q (as in classical epistemic logic) but also knows that j does not know p (a property of only knowing).

As an extension to our previous example, suppose ξ says that if i does not believe p is the only thing that j knows then it is the case that j knows more. That is,

$$\xi = \neg K_i O_j p \supset \neg O_j p$$

We can then show that $O_i \xi \supset K_i \neg O_j p$ is a valid sentence, that is, i believes that j knows more than p.

Given the reading of our graded multi-agent only knowing operator, we have the following expected result, which is the counterpart of [Belle and Lakemeyer, 2010, Theorem 5]: if an agent (or a group of agents) only knows that φ holds up to a certain degree, then she also only knows it up to a lower degree.

Proposition 3. *Let $m, n \in \mathbb{N}$ such that $n \geq m$ and let $\varphi \in \mathcal{L}_G^O$. Let $j \in \mathcal{A}$ and $G \subseteq \mathcal{A}$. Then, the following holds:*

$$O_j^n\varphi \rightarrow O_j^m\varphi \qquad\qquad O_G^n\varphi \rightarrow O_G^m\varphi$$

Proof. The proof follows the same lines as the proof of [Belle and Lakemeyer, 2010, Theorem 5]. Let (M, v) be a pointed epistemic model. First, one should observe that for all $\delta_m \in E_m^P$, there is $\delta_n \in S_n^{\delta_m}$ such that $M, v \models \delta_n$ if, and only if, $M, v \models \delta_m$ (this δ_n depends on the structure of (M, v) beyond modal depth m). Vice versa, for all $\delta_n \in E_n^P$, there is $\delta_m \in E_m^P$ such that $M, v \models \delta_n$ if, and only if, $M, v \models \delta_m$ (this δ_n does *not* depend on the structure of (M, v) beyond modal depth m). From this observation and by examining the truth conditions of Definition 6, we conclude easily that our two implications hold. QED

However, the reverse implications in Proposition 3 do not hold necessarily, because of the dependence on the structure of (M, v) in the first case. We now explore the relation to the Belle and Lakemeyer scheme in more detail in the next section.

4 Relation to Belle and Lakemeyer

First, we recall the essentials of the approach of Belle and Lakemeyer [2010]. They do not consider only knowing operators dealing with groups of agents and do not consider degrees of only knowing as we do. Moreover, and without loss of generality, we assume that there are only two agents: $\mathcal{A} := \{a, b\}$, as in [Belle and Lakemeyer, 2010]. Then, we consider the fragment $ON\mathcal{L}_n^O$ of the language $\mathcal{L}_G^O$ defined as follows:

Definition 9 (Languages $ON\mathcal{L}_n$ and $ON\mathcal{L}_n^O$). The language $ON\mathcal{L}_n$ is defined inductively as follows:

$$\varphi \;::=\; p \;\mid\; \neg\varphi \;\mid\; (\varphi \wedge \varphi) \;\mid\; K_j\varphi \;\mid\; N_j\varphi$$

where $p \in \mathcal{P}$ and $j \in \mathcal{A}$. The formula $O_j\varphi$ is an abbreviation for $K_j\varphi \wedge N_j\neg\varphi$. The language with the connectives $\neg, \wedge, K_j$ and O_j (instead of N_j) is denoted $ON\mathcal{L}_n^O$.

The semantics of Belle and Lakemeyer [2010] is based on the notion of *k-structures*. We consider a non-empty set of possible worlds $\mathcal{W}$ which simply consists of all the propositional valuations for the propositional letters in $\mathcal{P}$. A *k-structure* ($k \geq 1$), say e^k, for an agent is defined inductively as follows:

- $e^1 \subseteq \mathcal{W} \times \{\{\}\}$,
- $e^k \subseteq \mathcal{W} \times \mathbb{E}^{k-1}$, where $\mathbb{E}^m$ is the set of all m-structures.

A e^1 for a, denoted as e_a^1, is intended to represent a set of worlds $\{(w, \{\}), \ldots\}$. A e^2 is of the form $\{(w, e_b^1), (w', e_b'^1), \ldots\}$, and it is to be read as "at w, a believes b considers worlds from e_b^1 possible but at w', a believes b to consider worlds from $e_b'^1$ possible". This conveys the idea that a has only partial information about b, and so at different worlds, a believes different things about b.

We define a e^k for a, a e^j for b and a world $w \in \mathcal{W}$ as a (k, j)-model (e_a^k, e_b^j, w). Only formulas of a maximal a-depth of k, and a maximal b-depth of j are interpreted w.r.t. a (k, j)-model. (See [Belle and Lakemeyer, 2010] for more details on these notions; definitions are not reproduced here.) The truth conditions are defined as follows:

$e_a^k, e_b^j, w \models p$	iff	$p \in w$
$e_a^k, e_b^j, w \models \neg\varphi$	iff	it is not the case that $e_a^k, e_b^j, w \models \varphi$
$e_a^k, e_b^j, w \models \varphi \wedge \psi$	iff	$e_a^k, e_b^j, w \models \varphi$ and $e_a^k, e_b^j, w \models \psi$
$e_a^k, e_b^j, w \models K_a\varphi$	iff	for all $(w', e_b^{k-1}) \in e_a^k$, we have that $e_a^k, e_b^{k-1}, w' \models \varphi$
$e_a^k, e_b^j, w \models N_a\varphi$	iff	for all $(w', e_b^{k-1}) \notin e_a^k$, we have that $e_a^k, e_b^{k-1}, w' \models \varphi$

With these definitions and our abbreviations, we have that

$$e_a^k, e_b^j, w \models O_a\varphi \quad \text{iff} \quad \text{for all worlds } w', \text{ for all } e^{k-1} \text{ for } b, (w', e_b^{k-1}) \in e_a^k \text{ iff } e_a^k, e_b^{k-1}, w' \models \varphi$$

We say that a formula $\varphi \in ON\mathcal{L}_n^O$ is BL*-valid* when it is valid in the sense of Belle and Lakemeyer [2010], that is when it is true in all (k, j)-models, if φ is of a-depth k and b-depth j.

We now show that the set of BL-validities for the language $ON\mathcal{L}_n^O$ is the same as the set of validities in our logic L^O.

Lemma 1. *Let $\varphi \in ON\mathcal{L}_n^O$ be a formula of a-depth k and of b-depth j. Then, for all (k, j)-models (e_a^k, e_b^j, w), there is a pointed epistemic model (M, w) such that $e_a^k, e_b^j, w \models \varphi$ if, and only if, $M, w \models \varphi$. Vice versa, for all pointed epistemic models (M, w), there is a (k, j)-model (e_a^k, e_b^j, w) such that $e_a^k, e_b^j, w \models \varphi$ if, and only if, $M, w \models \varphi$.*

Proof. For the first part, the worlds of M are w and all the k'–structures and the j'–structures present in e_a^k and e_b^j. The valuations and the accessibility relations for these worlds are defined canonically. For the second part, we *unravel* the pointed epistemic model (M, w) up to modal depth k for the worlds accessible from w by a and up to modal depth j for the worlds accessible from w by b (see [Blackburn *et al.*, 2001]). QED

Theorem 2. *A formula $\varphi \in ON\mathcal{L}_n^O$ is L^O-valid if, and only if, it is BL-valid.*

Proof. It follows from the previous Lemma 1. QED

Note that, unlike the approach of Belle and Lakemeyer [2010], we can simultaneously satisfy an infinite set of sentences of unbounded modal depth.

5 The Cover Modality

The cover modality ∇ has been used as a syntactic primitive in modal logics [D'Agostino and Lenzi, 2005]. It has recently been axiomatized [Bílková *et al.*, 2008]. Here, we define a multi-modal version of this cover logic.

Definition 10 (Language $\mathcal{L}_\nabla$). The language $\mathcal{L}_\nabla$ is defined inductively as follows.

$$\varphi \ ::= \ p \ \mid \ \neg\varphi \ \mid \ (\varphi \wedge \varphi) \ \mid \ \nabla_j\{\varphi, \ldots, \varphi\}$$

where $j \in \mathcal{A}$ and $p \in \mathcal{P}$. Moreover, we use the following abbreviations: $K_j\varphi := \nabla_j\emptyset \vee \nabla_j\{\varphi\}$.

Definition 11 (Truth conditions for $\mathcal{L}_\nabla$). The satisfaction relation $\models$ is defined like in Definition 6 for the base case and for the connectives $\neg$ and $\wedge$. For the cover modalities, the truth conditions are defined as follows:

$$M, w \models \nabla_j\{\varphi_1, \ldots, \varphi_m\} \quad \text{iff}$$

for all $v \in R_j(w)$ there is $i \in \{1, \ldots, m\}$ such that $M, v \models \varphi_i$, and for all $i \in \{1, \ldots, m\}$ there is $v \in R_j(w)$ such that $M, v \models \varphi_i$.

Note that $\models \widehat{K_j}\varphi \leftrightarrow \nabla_j\{\top, \varphi\}$. Likewise, we can define our multi-agent only knowing modalities O_j^k with the cover modality as follows.

Proposition 4 (Expressiveness of the cover modality). *Let $\varphi \in \mathcal{L}_G^O$ and $k \in \mathbb{N}$. Let $P := P(\varphi)$ and $n = d(\varphi) + k$. Then for all pointed epistemic model (M, w), we have that the following holds: for any $j \in \mathcal{A}$,*

$$M, w \models O_j^k\varphi \quad \textit{iff} \quad M, w \models \nabla_j\{\delta_1, \ldots, \delta_m\}$$

where $\{\delta_1, \ldots, \delta_m\} \subseteq E_n^P$ is such that $\models \varphi \leftrightarrow \delta_1 \vee \ldots \vee \delta_m$.

Proof. It follows straightforwardly from the truth conditions of the multi-agent only knowing operators O_j^k and O_G^k given in Definition 6. QED

The above proposition shows that $\mathcal{L}_\nabla$ is at least as expressive as $\mathcal{L}^O$ on the class of epistemic models. Therefore:

Theorem 3 (Decidability of $\mathcal{L}^O$). *Let $\varphi \in \mathcal{L}^O$. The problem of determining whether φ is L^O-valid is decidable.*

Proof. The proof follows from the fact that $\mathcal{L}_\nabla$ is at least as expressive as $\mathcal{L}^O$ on the class of epistemic models by Proposition 4 and the fact that the validity problem of $\mathcal{L}_\nabla$ is decidable, as shown by Bılková *et al.* [2008]. QED

6 Proof System

To define the proof system for $\mathcal{L}_\nabla$, we need to introduce some further notations. Typically, formulas of $\mathcal{L}_\nabla$ are denoted $\varphi, \psi, \ldots$, finite sets of formulas are denoted $\alpha, \beta, \ldots$ and sets of sets of formulas are denoted $A, B, \ldots$

We define the (*power set*) *lifting* of the relation $\in \subseteq \mathcal{L}_\nabla \times 2^{\mathcal{L}_\nabla}$ as the relation $\overline{\in} \subseteq 2^{\mathcal{L}_\nabla} \times 2^{2^{\mathcal{L}_\nabla}}$ given by $\alpha\overline{\in}A$ if, and only if, for all $\varphi \in \alpha$, there is $\beta \in A$ such that $\varphi \in \beta$, and for all $\beta \in A$, there is $\varphi \in \alpha$ such that $\varphi \in \beta$.

Let E be a non-empty set. An object $A \in 2^{2^E}$ is a *redistribution* of a set $B \in 2^{2^E}$ if $\alpha\overline{\in}A$ for all $\alpha \in B$ (hence in particular $\bigcup B \subseteq \bigcup A$). We call such a redistribution *slim* if moreover $\bigcup B = \bigcup A$. The set of all slim redistributions of A is denoted by $SRD(A)$.

(∇_0) Axioms and Inference Rules of Prop. Logic

(∇_1) If $\forall\varphi \in \alpha, \exists\psi \in \beta$ such that $\varphi \rightarrow \psi$, and $\forall\psi \in \beta, \exists\varphi \in \alpha$ such that $\varphi \rightarrow \psi$ then $\nabla_j\alpha \rightarrow \nabla_j\beta$

(∇_2) $$\bigwedge\left\{\nabla_j\alpha \,\middle|\, \alpha \in A\right\} \rightarrow \bigvee\left\{\nabla_j\left\{\bigwedge\varphi \,\middle|\, \varphi \in B\right\} \,\middle|\, B \in SRD(A)\right\}$$

(∇_3) $$\nabla_j\left\{\bigvee\varphi \,\middle|\, \varphi \in A\right\} \rightarrow \bigvee\left\{\nabla_j\beta \,\middle|\, \beta\overline{\in}A\right\}$$

Figure 2: Proof System L_∇ for $\mathcal{L}_\nabla$

Theorem 4. *[Bılková* et al.*, 2008] The proof system L_∇ defined in Figure 2 is sound and strongly complete for $\mathcal{L}_\nabla$ w.r.t. the class of epistemic models.*

We are going to extend the language of $\mathcal{L}_\nabla$ in order to include explicitly the only knowing operator that we introduced in Section 3. Since this modality is definable in terms of the cover modality by Proposition 4, we straightforwardly obtain an axiomatization of this extended language.

Definition 12 (Language $\mathcal{L}_\nabla^O$). The language $\mathcal{L}_\nabla^O$ is defined inductively as follows.

$$\varphi \ ::= \ p \ \mid \ \neg\varphi \ \mid \ (\varphi \wedge \varphi) \ \mid \ \nabla_j\{\varphi, \ldots, \varphi\} \ \mid \ O_j^k\psi$$

where $\psi \in \mathcal{L}^O$, $k \in \mathbb{N}$, $j \in \mathcal{A}$ and $p \in \mathcal{P}$. Moreover, we use the following abbreviation: $K_j\varphi := \nabla_j\emptyset \vee \nabla_j\{\varphi\}$.

The truth conditions are defined like in Definition 3 for the base case and for the connectives $\neg$ and $\wedge$. For the cover modalities, the truth conditions are defined like in Definition 11, and for the multi-agent only knowing modality the truth conditions are defined like in Definition 6.

Theorem 5. *The proof system L_∇^O is the proof system L_∇ of Figure 2 to which we add the following inference rule: for all $j \in \mathcal{A}$, for all $\varphi \in \mathcal{L}^O$, for all $k \in \mathbb{N}$,*

(O_j^k) *If $\varphi \leftrightarrow \delta_1 \vee \ldots \vee \delta_m$ with $\{\delta_1, \ldots, \delta_m\} \subseteq E_n^P$ then $O_j^k(\varphi) \leftrightarrow \nabla_j\{\delta_1, \ldots, \delta_m\}$*

where $n := d(\varphi) + k$ and $P := P(\varphi)$.

Then, the proof system L_∇^O is sound and strongly complete for $\mathcal{L}_\nabla^O$ w.r.t. the class of epistemic models.

Proof. The proof of soundness is standard. Completeness is obtained by observing that Theorem 4 gives us completeness for the language $\mathcal{L}_\nabla$ without the multi-agent only knowing modality. Then, we obtain completeness for the full language $\mathcal{L}_\nabla^O$ because the multi-agent only knowing modality is definable in terms of the cover modality by Proposition 4 and this definition corresponds in fact to our inference rule (O_j^k). QED

The language $\mathcal{L}_\nabla^O$ is an extension of the language $\mathcal{L}^O$ of Section 3 with the cover modality. From Theorem 5, we obtain a soundness and completeness result for this restricted language $\mathcal{L}^O$:

Corollary 3. *Let $\varphi \in \mathcal{L}^O$. Then, φ is derivable in the proof system L_∇^O if, and only if, φ is L^O-valid.*

7 Other Related Work

Besides the immediately relevant work already discussed above, let us note the following related efforts.

Minimal knowledge approaches have also enjoyed multi-agent extensions [Hoek and Thijsse, 2002]. While minimal knowledge is related to only knowing, it differs in the sort of conclusions one can draw. We refer readers to [Levesque and Lakemeyer, 2001] for discussions. A related proposal is that of *total knowledge* [Pratt-Hartmann, 2000], where knowledge is required to be true. So defaults, for example, would lead to an inconsistency.

Although defaults are not the focus of this paper, there are also proposals that study the interaction between knowledge and defaults in a multi-agent setting, such as [Morgenstern, 1990]. In terms of proof theory, Bonatti and Olivetti [2002], among others, have developed proof systems for such defaults. The relation between concepts like only knowing and this work, however, remains to be explored.

Finally, there are numerous modal logics for multi-agent systems for concepts such as beliefs, dynamics, and desires, which we do not review here. See [van der Hoek and Wooldridge, 2012; Fagin *et al.*, 1995] and references therein. Only knowing is not considered by these, however.

8 Conclusion

We investigated a new version of multi-agent only knowing in the classical Kripke setting, while putting it in context of existing proposals on this topic. This comes with a number of benefits – for example, the definability with the cover modality, a group version of multi-agent only knowing, and an operator for assessing the epistemic entrenchment of what an agent or a group only knows – while avoiding proof-theoretic constructions.

At this point, the development of this paper could lead to dynamic logic based representation formalisms such as [van Ditmarsch *et al.*, 2011] finally embracing the only knowing modality in a multi-agent and dynamic setting. These formalisms are also based on a Kripke semantics.

References

[Belle and Lakemeyer, 2010] V. Belle and G. Lakemeyer. Multi-agent only-knowing revisited. In Fangzhen Lin, Ulrike Sattler, and Miroslaw Truszczynski, editors, *KR*. AAAI Press, 2010.

[Bılková *et al.*, 2008] M. Bılková, A. Palmigiano, and Y. Venema. Proof systems for the coalgebraic cover modality. *Advances in Modal Logic*, 7:1–21, 2008.

[Blackburn *et al.*, 2001] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Computer Science*. Cambridge University Press, 2001.

[D'Agostino and Lenzi, 2005] G. D'Agostino and G. Lenzi. An axiomatization of bisimulation quantifiers via the μ-calculus. *Theoretical Computer Science*, 338(1):64–95, 2005.

[Fine, 1975] K. Fine. Normal forms in modal logic. *Notre Dame Journal of Formal Logic*, 16:229–237, 1975.

[Levesque, 1990] H. J Levesque. All I know: a study in autoepistemic logic. *Artificial intelligence*, 42(2):263–309, 1990.

[Moss, 2007] L. S. Moss. Finite models constructed from canonical formulas. *Journal of Philosophical Logic*, 36(6):605–640, 2007.

[Belle and Lakemeyer, 2010] V. Belle and G. Lakemeyer. Multi-agent only-knowing revisited. In *Proc. KR*, pages 49–60, 2010.

[Bonatti and Olivetti, 2002] P. A. Bonatti and N. Olivetti. Sequent calculi for propositional nonmonotonic logics. *ACM Trans. Comput. Log.*, 3(2):226–278, 2002.

[Demolombe, 2003] R. Demolombe. Belief change: from situation calculus to modal logic. In *Proc. Nonmonotonic Reasoning, Action, and Change (NRAC)*, 2003.

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

[Halpern and Lakemeyer, 2001] J.Y. Halpern and G. Lakemeyer. Multi-agent only knowing. *Journal of Logic and Computation*, 11(1):251–265, 2001.

[Halpern and Moses, 1984] J. Y. Halpern and Y. Moses. Towards a theory of knowledge and ignorance: Preliminary report. In *Proc. NMR*, pages 125–143, 1984.

[Halpern, 1993] J. Y. Halpern. Reasoning about only knowing with many agents. In *Proc. AAAI*, pages 655–661, 1993.

[Herzig *et al.*, 2000] A. Herzig, J. Lang, D. Longin, and T. Polacsek. A logic for planning under partial observability. In *Proc. AAAI / IAAI*, pages 768–773, 2000.

[Hoek and Thijsse, 2002] 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.

[Lakemeyer, 1993] G. Lakemeyer. All they know: A study in multi-agent autoepistemic reasoning. In *Proc. IJCAI*, pages 376–381, 1993.

[Levesque and Lakemeyer, 2001] H.J. Levesque and G. Lakemeyer. *The logic of knowledge bases*. The MIT Press, 2001.

[Morgenstern, 1990] L. Morgenstern. A formal theory of multiple agent nonmonotonic reasoning. In *Proc. AAAI*, pages 538–544, 1990.

[Pratt-Hartmann, 2000] I. Pratt-Hartmann. Total knowledge. In *Proc. AAAI*, pages 423–428, 2000.

[van der Hoek and Wooldridge, 2012] W. van der Hoek and M. Wooldridge. Logics for multiagent systems. *AI Magazine*, 33(3):92–105, 2012.

[van Ditmarsch *et al.*, 2011] H. P. van Ditmarsch, A. Herzig, and T. De Lima. From situation calculus to dynamic epistemic logic. *J. Log. Comput.*, 21(2):179–204, 2011.

[Waaler and Solhaug, 2005] A. Waaler and B. Solhaug. Semantics for multi-agent only knowing: extended abstract. In *Proc. TARK*, pages 109–125, 2005.