All They Know About

Gerhard Lakemeyer
Institute of Computer Science III
University of Bonn
Römerstr. 164
5300 Bonn 1, Germany
gerhard@cs.uni-bonn.de

Abstract

We address the issue of agents reasoning about other agents' nonmonotonic reasoning ability in the framework of a multi-agent autoepistemic logic (AEL). In single-agent AEL, nonmonotonic inferences are drawn based on *all* the agent knows. In a multi-agent context such as Jill reasoning about Jack's nonmonotonic inferences, this assumption must be abandoned since it cannot be assumed that Jill knows everything Jack knows. Given a specific subject matter like Tweety the bird, it is more realistic and sufficient if Jill only assumes to know all Jack knows about Tweety in order to arrive at Jack's nonmonotonic inferences about Tweety. This paper provides a formalization of *all an agent knows about a certain subject matter* based on possible-world semantics in a multi-agent AEL. Besides discussing various properties of the new notion, we use it to characterize formulas that are about a subject matter in a very strong sense. While our main focus is on subject matters that consist of atomic propositions, we also address the case where agents are the subject matter.

Introduction

Most of the research on nonmonotonic reasoning has concentrated on the single-agent case. However, there is little doubt that agents who have been invested with a nonmonotonic reasoning mechanism should be able to reason about other agents and their ability to reason nonmonotonically as well. For example, if we assume the common default that birds normally fly and if Jill tells Jack that she has just bought a bird, then Jill should be able to infer that Jack thinks that her bird flies. Other examples from areas like planning and temporal projection can be found in [Mor90].

One of the main formalisms of nonmonotonic reasoning is *autoepistemic logic* (AEL) (e.g. [Moo85]). The basic idea is that the beliefs of agents are closed under *perfect introspection*, that is, they know[1] what they know and do not know. Nonmonotonic reasoning comes about in this framework in that agents can draw inferences on the basis of their own ignorance. For example, Jack's flying-bird default can be phrased as the belief that birds fly unless known otherwise. If Jill tells Jack that she has a bird called Tweety, Jack will conclude that Tweety flies since he does not know of any reason why Tweety should not be able to fly.

Note that in standard AEL agents determine their beliefs and especially their non-beliefs with respect to *everything* they believe. Thus, if we think of Jack's beliefs being represented by a knowledge base (KB), then Jack's belief that Tweety flies follows in AEL because the KB is *all* he believes or, as we also say, because Jack *only-believes* the formulas in the KB. (See [Lev90] for a formalization of AEL with an explicit notion of only-knowing.) If we want to extend autoepistemic logic to the multi-agent case, the strong assumption of having access to everything that is known is not warranted when applied to knowledge about other agents' knowledge. In other words, while an agent may have access to everything she herself knows, it is certainly not the case that she knows everything agents other than herself know. How then should Jill, for example, conclude that Jack autoepistemically infers that Tweety flies without pretending that she knows everything Jack knows? Intuitively, it seems sufficient for Jill to assume that she knows everything Jack knows *about Tweety*, which is quite plausible since Jack heard about Tweety through Jill. Thus if Jill believes that all Jack knows about Tweety is that he is a bird and, hence, that the flying-bird default applies to Tweety, then Jill is justified to conclude that Jack believes that Tweety flies.

In this paper, we present an account of multi-agent AEL which explicitly models the notion of *only-knowing about a subject matter*, which we also call *only-knowing-about*, for short. The work is based on possible-world semantics and takes ideas from a single-agent version of only-knowing-about [Lak92] and combines them with a multi-agent version of only-

[1] We use the terms knowledge and belief interchangeably in this paper for stylistic reasons. However, the formalism presented allows agents to have false beliefs.

knowing [Lak93].

None of the existing approaches to multi-agent AEL [Mor90, MG92, Hal93, Lak93][2] addresses the issue of only-knowing-about and are thus forced, one way or the other, to unrealistic assumptions when it comes to reasoning about the autoepistemic conclusions of other agents. Morgenstern and Guerreiro [Mor90, MG92], for example, run into the problem of what they call *arrogance*, where an agent i ascribes a non-belief to an agent j only on the basis of i herself not knowing whether j has this belief. Their solution is to heuristically limit the use of such arrogant behavior depending on specific applications. In [Lak93, Hal93], the only way to model the Tweety example is to assume that Jill knows *all* Jack knows.

The multi-agent AEL OL_n of [Lak93] is the starting point of this paper. OL_n's syntax and semantics are presented in the next section. We then extend OL_n by incorporating an explicit notion of only-knowing-about. Besides discussing various properties of the new notion, we use it to characterize formulas that are about a subject matter in a very strong sense. While our main focus is on subject matters that consist of atomic propositions, we also address the case where agents are the subject matter. We conclude with a summary of the results and future work.

The Logic OL_n

After introducing the syntax of the logic, we define the semantics in two stages. First we describe that part of the semantics that does not deal with only-knowing. In fact, this is just an ordinary possible-world semantics for n agents with perfect introspection. Then we introduce the necessary extensions that give us the semantics of only-knowing. The properties of only-knowing are discussed briefly. A detailed account is given in [Lak93].

Syntax

Definition 1 *The Language*
The primitives of the language consist of a countably infinite set of atomic propositions (or atoms), the connectives $\vee$, $\neg$, *and the modal operators* $\mathbf{L}_i$ *and* $\mathbf{O}_i$ *for* $1 \leq i \leq n$. *(Agents are referred to as* $1, 2, \ldots, n$.*) Formulas are formed in the usual way from these primitives.*[3] $\mathbf{L}_i\alpha$ *should be read as "the agent* i *believes* α*" and* $\mathbf{O}_i\alpha$ *as "*α *is all agent* i *believes." A formula* α *is called* **basic** *iff there are no occurrences of* $\mathbf{O}_i$ $(1 \leq i \leq n)$ *in* α.

[2] Some notes on multi-agent AEL appear also in [HM84]. There has also been work in applying nonmonotonic theories to special multi-agent settings such as speech act theory, e.g. [Per87, AK88]. Yet these approaches do not aim to provide general purpose multi-agent nonmonotonic formalisms.

[3] We will freely use other connectives like $\wedge$, $\supset$ and $\equiv$, which should be understood as syntactic abbreviations of the usual kind.

Definition 2 *A modal operator occurs at* **depth** n *of a formula* α *iff it occurs within the scope of exactly* n *modal operators.*

For example, given $\alpha = p \wedge \mathbf{L}_1\mathbf{L}_2(\mathbf{L}_3 q \vee \neg\mathbf{O}_2 r)$, $\mathbf{L}_1$ occurs at depth 0, $\mathbf{L}_2$ at depth 1, and $\mathbf{L}_3$ and $\mathbf{O}_2$ occur both at depth 2.

Definition 3 *A formula* α *is called* **i-objective** *(for* $i = 1, \ldots n$*) iff every modal operator at depth 0 is of the form* $\mathbf{O}_j$ *or* $\mathbf{L}_j$ *with* $i \neq j$.

In other words, i-objective formulas talk about the external world from agent i's point of view, which includes beliefs of other agents but not his own.
For example, $(p \vee \mathbf{L}_2 q) \wedge \neg\mathbf{O}_3\mathbf{L}_1 p$ is 1-objective, but $(p \vee \mathbf{L}_2 q) \wedge \neg\mathbf{L}_1\mathbf{O}_3 p$ is not.

The Semantics of Basic Formulas

Basic formulas are given a standard possible-world semantics [Kri63, Hin62, Hin71], which the reader is assumed to be familiar with.[4] Roughly, a *possible-world model* consists of *worlds*, which determine the truth of atomic propositions, and binary *accessibility* relations between worlds. An agent's beliefs at a given world w are determined by what is true in all those worlds that are accessible to the agent from w. Since we are concerned with agents whose beliefs are consistent and closed under *perfect introspection*, we restrict the accessibility relations in the usual way. The resulting logic is called $KD45_n$.[5]

Definition 4 *A* $KD45_n$*-Model*
$M = \langle W, \pi, R_1, \ldots, R_n\rangle$ *is called a* $KD45_n$**-model** *(or simply* **model***) iff*

1. W *is a set (of worlds).*
2. π *is a mapping from the set of atoms into* 2^W.
3. $R_i \subseteq W \times W$ *for* $1 \leq i \leq n$.
4. R_i *is serial, transitive, and Euclidean*[6] *for* $1 \leq i \leq n$.

Given a model $M = \langle W, \pi, R_1, \ldots, R_n\rangle$ and a world $w \in W$, the meaning of basic formulas is defined as follows: Let p be an atom and α and β arbitrary basic formulas.

$$
\begin{array}{lcl}
w \models p & \Longleftrightarrow & w \in \pi(p) \\
w \models \neg\alpha & \Longleftrightarrow & w \not\models \alpha \\
w \models \alpha \vee \beta & \Longleftrightarrow & w \models \alpha \text{ or } w \models \beta \\
w \models \mathbf{L}_i\alpha & \Longleftrightarrow & \text{for all } w', \text{ if } wR_iw' \text{ then } w' \models \alpha
\end{array}
$$

The Canonical Model

It is well known that, as far as basic formulas are concerned, it suffices to consider just one, the so-called *canonical model* [HC84, HM92]. This canonical model will be used later on to define the semantics of only-knowing.

[4] See [HC84, HM92] for an introduction.

[5] We use the subscript n to indicate that we are concerned with the n-agent case.

[6] R_i is Euclidean iff $\forall w, w', w''$, if wR_iw' and wR_iw'', then $w'R_iw''$.

Definition 5 *Maximally Consistent Sets*
Given any proof theory of $KD45_n$ and the usual notion of theoremhood and consistency, a set of basic formulas Γ is called **maximally consistent** *iff Γ is consistent and for every basic α, either α or $\neg\alpha$ is contained in Γ.*

The canonical $KD45_n$-model M_c has as worlds precisely all the maximally consistent sets and a world w' is R_i-accessible from w just in case all of i's beliefs at w are included in w'.

Definition 6 *The Canonical $KD45_n$-Model M_c*
The canonical model is a Kripke structure $M_c = \langle W_c, \pi, R_1, \ldots, R_n\rangle$ such that

1. *$W_c = \{w \mid w$ is a maximally consistent set$\}$.*
2. *For all atoms p and $w \in W_c$, $w \in \pi(p)$ iff $p \in w$.*
3. *wR_iw' iff for all formulas $\mathbf{L}_i\alpha$, if $\mathbf{L}_i\alpha \in w$ then $\alpha \in w'$.*

The following (well known) theorem tells us that nothing is lost from a logical point of view if we confine our attention to the canonical model.

Theorem 1 *M_c is a $KD45_n$-model and for every set of basic formulas Γ, Γ is satisfiable[7] iff it is satisfiable in M_c.*

The Semantics of Only-Knowing

Given this classical possible-world framework, what does it mean for an agent i to only-know, say, an atom p at some world w in a model M? Certainly, i should believe p, that is, all worlds that are i-accessible from w should make p true. Furthermore, i should believe as little else as possible apart from p. For example, i should neither believe q nor believe that j believes p etc. Minimizing knowledge using possible worlds simply means *maximizing* the number of accessible worlds. Thus, in our example, there should be an accessible world where q is false and another one where j does not believe p and so on. It should be clear that in order for w to satisfy only-knowing α this way, the model M must have a huge supply of worlds that are accessible from w. While not essential for the definition of only-knowing, it turns out to be very convenient to simply restrict our attention to models that are guaranteed to contain a sufficient supply of worlds. In fact, we will consider just one, namely the *canonical model* of $KD45_n$. Let us call the set of all formulas that are true at some world w in some model of $KD45_n$ a *world state*. The canonical model has the nice property that it contains precisely one world for every possible world state, since world states are just maximally consistent sets.

With that agent i is said to only-know a formula α at some world w (in the canonical model) just in case α is believed and any world w' which satisfies α and from which the same worlds are i-accessible as from w is itself i-accessible from w.

Definition 7 *Given a model $M = \langle W, \pi, R_1, \ldots, R_n\rangle$ and worlds w and w' in W, we say that w and w' are* **i-equivalent** *($w \approx_i w'$) iff for all worlds $w^* \in W$, wR_iw^* iff $w'R_iw^*$.*

Given an arbitrary formula α, a world w in a model M, let

$$w \models \mathbf{O}_i\alpha \iff \text{for all } w' \text{ s.t. } w \approx_i w', wR_iw' \text{ iff } w' \models \alpha.$$

A formula α is a logical consequence of a set of formulas Γ ($\Gamma \models \alpha$) iff for all worlds w in the canonical model M_c, if $w \models \gamma$ for all $\gamma \in \Gamma$, then $w \models \alpha$. As usual, we say that α is valid ($\models \alpha$) iff $\{\} \models \alpha$. A formula α is satisfiable iff $\neg\alpha$ is not valid.

Some Properties of the Logic Here we can only sketch some of the properties of OL_n. The logic is discussed in more detail in [Lak93].[8] See also Halpern's logic of only-knowing [Hal93], which is closely related to OL_n.

Given Theorem 1, it is clear that the properties of basic formulas (no $\mathbf{O}_i$'s) are precisely those of $KD45_n$. Concerning only-knowing we restrict ourselves to the connection between only-knowing and a natural multi-agent version of the *stable expansions* of autoepistemic logic [Moo85].

Definition 8 *i-Epistemic State*
A set of basic formulas Γ is called an i-epistemic state iff there is a world w in M_c such that for all basic γ, $w \models \mathbf{L}_i\gamma$ iff $\gamma \in \Gamma$.

Given a set of basic formulas Γ, let $\overline{\Gamma} = \{\text{basic } \gamma \mid \gamma \notin \Gamma\}$, $\mathbf{L}_i\Gamma = \{\mathbf{L}_i\gamma \mid \gamma \in \Gamma\}$, and $\neg\mathbf{L}_i\overline{\Gamma} = \{\neg\mathbf{L}_i\gamma \mid \gamma \in \overline{\Gamma}\}$.

Definition 9 *i-Stable Expansion*
Let A be a set of basic formulas and let $\models_{\mathrm{KD45}}$ denote logical consequence in $KD45_n$. Γ is called an i-stable expansion of A iff
$\Gamma = \{\textit{basic } \gamma \mid A \cup \mathbf{L}_i\Gamma \cup \neg\mathbf{L}_i\overline{\Gamma} \models_{\mathrm{KD45}} \gamma\}$.

Note that the use of $\models_{\mathrm{KD45}}$ instead of logical consequence in propositional logic is essentially the only difference between Moore's original definition and this one. The following theorem establishes that the i-stable expansions of a formula α correspond precisely to the different i-epistemic states of agent i who only-knows α.

Theorem 2 *Only-Knowing and i-Stable Expansions*
Let α be a basic formula, $w \in W_c$, and let $\Gamma = \{$basic $\alpha \mid w \models \mathbf{L}_i\alpha\}$. Then $w \models \mathbf{O}_i\alpha$ iff Γ is an i-stable expansion of $\{\alpha\}$.

(See also [Hal93] for an analogous result.)

[7] A set of basic formulas Γ is *satisfiable* iff there is a model $M = \langle W, \pi, R_1, \ldots, R_n\rangle$ and $w \in W$ such that $w \models \gamma$ for all $\gamma \in \Gamma$.

[8] As a minor difference, [Lak93] uses $K45_n$ instead of $KD45_n$ as the base logic, that is, beliefs are not required to be consistent.

E-Clauses We end our discussion of OL_n by extending the notion of a clause of propositional logic to the modal case (e-clauses) and show that i-epistemic states are uniquely determined by the i-objective e-clauses they contain. This will be useful in developing the semantics of only-knowing-about in the next section.

Definition 10 *E-Clauses*
An **e-clause** *is a disjunct of the form*

$$(\bigvee_{i=1}^{u} l_i \vee \bigvee_{i=u+1}^{v} \mathbf{L}_{j_i} c_i \vee \bigvee_{i=v+1}^{w} \neg \mathbf{L}_{j_i} d_i),$$

where the l_i are literals and the c_i and d_i are themselves e-clauses.

Definition 11 *Extended Conjunctive Normal Form*
A basic formula α is in extended conjunctive normal form (ECNF) iff $\alpha = \bigwedge \alpha_i$ and every α_i is an e-clause.

Lemma 1 *For every basic α there is an α^* in ECNF such that $\models \alpha \equiv \alpha^*$.*

Every i-epistemic state is uniquely determined by the i-objective e-clauses it contains.

Theorem 3 *Let Γ and Γ' be two i-epistemic states. If Γ and Γ' agree on all their i-objective clauses, then $\Gamma = \Gamma'$.*

The Logic OL_n^a

We now extend OL_n to a logic that allows us to express things like "x is all agent i knows about subject y." A subject matter is defined as any finite subset π of atoms.[9] For every agent i and subject matter π, let $\mathbf{O}_i\langle\pi\rangle$ be a new modal operator. $\mathbf{O}_i\langle\pi\rangle\alpha$ should be read as "α is all agent i knows about π." If we refer to a subject matter extensionally, we sometimes leave out the curly brackets. For example, we write $\mathbf{O}_i\langle p, q\rangle\alpha$ instead of $\mathbf{O}_i\langle\{p, q\}\rangle\alpha$. Formulas in this extended language are formed in the usual way except for the following restriction: $\mathbf{O}_i\langle\pi\rangle$ may only be applied to **basic** formulas, that is, for any given $\mathbf{O}_i\langle\pi\rangle\alpha$, the only modal operators allowed in α are $\mathbf{L}_1, \ldots, \mathbf{L}_n$. Given a formula α and a subject matter π, we say that α **mentions** π iff at least one of the atoms of π occur in α.

To define the semantics of $\mathbf{O}_i\langle\pi\rangle\alpha$, we follow an approach similar to [Lak92], where only-believing-about is reduced to only-believing after "forgetting" everything that is irrelevant to π. Given a world w and a subject matter π, forgetting irrelevant beliefs is achieved by mapping w into a world $w|_{\pi,i}$ that is just like w except that only those beliefs of i are preserved that are *relevant* to π. Assuming we have such a $w|_{\pi,i}$, then i only-believes α about π at w just in case i believes it at w and i only-believes α at $w|_{\pi,i}$.

[9]The results of the paper do not hinge on π being finite. What matters is that we have a way to refer to each subject matter in our language. If π is finite, this can always be done.

The crucial part of the semantics then is the construction of $w|_{\pi,i}$. It is obtained in two steps. First we collect all the beliefs of i about π in a set $\Gamma^w_{\pi,i}$. Theorem 3 allows us to restrict ourselves to i-objective e-clauses only. What does it mean for i to believe an e-clause c that is relevant to π? A reasonable answer seems to be the following: any formula that is believed by i and that implies c must mention the subject matter. For example, if $\pi = \{p\}$ and all i believes at w is $(p \vee q) \wedge r$, then the clauses selected to be relevant beliefs about π are $(p \vee q)$ and weaker ones like $(p \vee q \vee s)$. However, neither $(p \vee r)$ nor any clause not mentioning p are selected. $(p \vee r)$ is disqualified because it is contingent on r, which is also believed, thus not conveying any information about p.

Given $\Gamma^w_{\pi,i}$ it is then easy to define $w|_{\pi,i}$ in such a way that it believes only the formulas contained in $\Gamma^w_{\pi,i}$.

Definition 12 *Given a world w in the canonical model M_c, let*

$$\Gamma^w_{\pi,i} = \{c \mid \begin{array}{l} c \textit{ is an i-objective e-clause}, w \models \mathbf{L}_i c \text{ and} \\ \textit{for all i-objective basic } \alpha, \textit{ if } w \models \mathbf{L}_i\alpha \\ \textit{and } \models \alpha \supset c \textit{ then } \alpha \textit{ mentions } \pi\}. \end{array}$$

Lemma 2 *Given a world w of M_c and a subject matter π, let*

$$\begin{aligned}\Lambda_1 =& \{\mathbf{L}_i\gamma \mid \gamma \textit{ is i-objective and } \mathbf{L}_i\Gamma^w_{\pi,i} \models \mathbf{L}_i\gamma\} \cup \\ & \{\neg\mathbf{L}_i\gamma \mid \gamma \textit{ is i-objective and } \mathbf{L}_i\Gamma^w_{\pi,i} \not\models \mathbf{L}_i\gamma\} \\ \Lambda_2 =& \{l \mid l \textit{ is a literal in } w\} \cup \textstyle\bigcup_{j \neq i}\{\mathbf{L}_j\gamma \mid \mathbf{L}_j\gamma \in w\} \cup \\ & \textstyle\bigcup_{j \neq i}\{\neg\mathbf{L}_j\gamma \mid \neg\mathbf{L}_j\gamma \in w\}.\end{aligned}$$

Then $\Lambda_1 \cup \Lambda_2$ is consistent.

Definition 13 *Given w, π, and Λ_1 and Λ_2 of the previous lemma, let $w|_{\pi,i}$ be a world (= maximally consistent set) that contains Λ_1 and Λ_2.*

Note that by containing Λ_2, $w|_{\pi,i}$ is exactly like w except for the beliefs of agent i. Furthermore, Λ_1 makes sure that agent i at $w|_{\pi,i}$ believes no more than what follows from $\Gamma^w_{\pi,i}$, that is, the agent believes only what is relevant to the subject matter π.

Lemma 3 *$w|_{\pi,i}$ is unique.*

With this machinery we are finally ready to formally define the semantics of $\mathbf{O}_i\langle\pi\rangle\alpha$ for any basic formulas α and subject matter π:

$$w \models \mathbf{O}_i\langle\pi\rangle\alpha \iff w|_{\pi,i} \models \mathbf{O}_i\alpha \text{ and } w \models \mathbf{L}_i\alpha.$$

Satisfiability, logical consequence, and validity in OL_n^a are defined as for OL_n.

Some Properties of Only-Knowing-About

So far we do not have a complete axiomatization of only-knowing-about. The following properties, which are natural generalizations of the single-agent case [Lak92], suggest that our definitions are reasonable.

Definition 14 *Given a formula α,*
let $\pi_\alpha = \{p \mid p \textit{ is an atom that occurs in } \alpha\}$.

1) $\models \mathbf{O}_i\langle\pi\rangle\alpha \supset \mathbf{L}_i\alpha$.
Follows immediately from the definition.

2) $\models \neg\mathbf{O}_i\langle\pi\rangle\alpha$ if $\not\models\alpha$ and $\pi \cap \pi_\alpha = \{\}$.
In other words, an agent cannot only-know something about π that is totally irrelevant to π.

3) $\models \mathbf{O}_i\langle\pi\rangle\alpha \equiv \mathbf{O}_i\langle\pi\rangle\beta$ if $\models\alpha \equiv \beta$.
In other words, the syntactic form of what is only-known about π does not matter.

4) $\models \mathbf{O}_i\langle p\rangle(p \vee q) \supset (\neg\mathbf{L}_i p \wedge \neg\mathbf{L}_i q \wedge \neg\mathbf{L}_i\neg p \wedge \neg\mathbf{L}_i\neg q)$.
Here, assuming any of the beliefs of the right-hand-side implies that i must know more about p than just $p \vee q$.

5) $\models \mathbf{O}_i\alpha \supset \mathbf{O}_i\langle\pi_\alpha\rangle\alpha$.
This says that, if all you know is α and if the subject matter spans everything you know (see Definition 14), then surely α is all you know about this subject matter.

The following theorem characterizes cases where reasoning from only-knowing-about is the same as reasoning from only-knowing. This is interesting for at least two reasons. For one, only-knowing has a much simpler definition than only-knowing-about. For another, the theorem tells us that even though one usually does not know all another agent knows, there are cases where one may pretend to know all the other agent knows without drawing false conclusions. This will be very useful in the following section, where we show that Jill is indeed able to infer that Jack believes that Tweety flies.

Theorem 4 *Let α and β be basic formulas such that $\pi_\beta \subseteq \pi_\alpha$. Then $\models \mathbf{O}_i\langle\pi_\alpha\rangle\alpha \supset \mathbf{L}_i\beta$ iff $\models \mathbf{O}_i\alpha \supset \mathbf{L}_i\beta$.*

Jack, Jill, and Tweety Revisited

We now demonstrate that our formalism is able to model our initial Tweety example correctly. Let us assume that all Jack believes about Tweety is the flying-bird default for Tweety (α) and the fact that Tweety is a bird (β).
$\alpha = [\mathtt{Bird}(Tweety) \wedge \neg\mathbf{L}_{jack}\neg\mathtt{Fly}(Tweety)] \supset \mathtt{Fly}(T.)$
$\beta = \mathtt{Bird}(Tweety)$
The subject matter Tweety can be characterized as the set of relevant predicates that mention Tweety, i.e. $\pi = \{\mathtt{Bird}(Tweety), \mathtt{Fly}(Tweety)\}$. We then obtain

$$\models \mathbf{O}_{jack}\langle\pi\rangle(\alpha \wedge \beta) \supset \mathbf{L}_{jack}\mathtt{Fly}(Tweety) \qquad (1)$$

Proof : To prove this fact, note that $\pi_{\alpha\wedge\beta} = \pi$. Thus Theorem 4 applies and it suffices to show that $\models \mathbf{O}_{jack}(\alpha \wedge \beta) \supset \mathbf{L}_{jack}\mathtt{Fly}(Tweety)$. It suffices to show that $\models \mathbf{O}_{jack}(\alpha \wedge \beta) \equiv \mathbf{O}_{jack}(\mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety))$, since $\models \mathbf{O}_{jack}(\mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety)) \supset \mathbf{L}_{jack}\mathtt{Fly}(Tweety)$ follows immediately from the semantics of $\mathbf{O}_{jack}$.

Let $w \models \mathbf{O}_{jack}(\mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety))$, that is, for all w' such that $w \approx_{jack} w'$, $wR_{jack}w'$ iff $w' \models (\mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety))$. For all w' s.t. $w \approx_{jack} w'$ we obtain that $w' \models (\mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety))$ iff $w' \models \alpha \wedge \beta$ since $w' \models \neg\mathbf{L}_{jack}\neg\mathtt{Fly}(Tweety)$. Thus $w \models \mathbf{O}_{jack}(\alpha \wedge \beta)$.

Conversely, let $w \models \mathbf{O}_{jack}(\alpha \wedge \beta)$, that is, for all w' such that $w \approx_{jack} w'$, $wR_{jack}w'$ iff $w' \models (\alpha \wedge \beta)$. Let w' be any world such that $w \approx_{jack} w'$. Let $w' \models \mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety)$. Then $w \models \alpha \wedge \beta$ and, hence $wR_{jack}w'$. Conversely, let $wR_{jack}w'$. Obviously, $w' \models \mathtt{Bird}(Tweety)$. Assume $w' \not\models \mathtt{Fly}(Tweety)$. Then $w' \models \mathbf{L}_{jack}\neg\mathtt{Fly}(Tweety)$ by assumption. Then there is a world w'' such that $w \approx_{jack} w''$ and $w'' \models \mathbf{L}_{jack}\neg\mathtt{Fly}(Tweety) \wedge \mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety)$. Hence $w'' \models (\alpha \wedge \beta)$ and, therefore, $wR_{jack}w''$, contradicting the assumption that $w' \models \mathbf{L}_{jack}\neg\mathtt{Fly}(Tweety)$. Thus $w \models \mathbf{O}_{jack}(\mathtt{Bird}(Tweety) \wedge \mathtt{Fly}(Tweety))$. ■

Given (1) and $\models \mathbf{L}_i(\alpha \supset \beta) \supset (\mathbf{L}_i\alpha \supset \mathbf{L}_i\beta)$ for all α, β and i, we immediately obtain the desired result

$$\models \mathbf{L}_{jill}\mathbf{O}_{jack}\langle\pi\rangle(\alpha \wedge \beta) \supset \mathbf{L}_{jill}\mathbf{L}_{jack}\mathtt{Fly}(Tweety),$$

that is, if Jill knows what Jack knows about Tweety, then she also knows what default inferences Jack makes about Tweety.

Strictly π-Relevant Formulas

It seems very hard to define what it means for a formula to be about a subject matter on purely syntactic grounds. For example, while $(p \supset q)$ is intuitively about p, $(p \supset q) \wedge (\neg p \supset q)$ (which is equivalent to q) is not. Also note that, while $(q \vee r)$ by itself is clearly not about p, $(q \vee r)$ becomes nontrivial information about p as part of the formula $(q \vee r) \wedge (p \equiv q)$. Our notion of only-knowing-about captures these subtleties of aboutness and yields a useful definition of formulas that are about a subject matter in a very strong sense.

Definition 15 *Let α be a basic formula such that $\not\models\alpha$ and let π be a subject matter. α is called* **strictly π-relevant** *iff $\mathbf{O}_i\langle\pi\rangle\alpha$ is satisfiable for some i.*

Intuitively, every piece of information conveyed by a strictly π-relevant formula tells us something nontrivial about π.

p, $\neg\mathbf{L}_i p \vee (q \wedge r)$, and $(q \vee r) \wedge (p \equiv q)$ are all strictly p-relevant, $(p \supset q) \wedge (\neg p \supset q)$, $p \wedge (p \supset q)$, and $(p \vee \neg p)$, on the other hand, are not. At first glance, one may want to include $(p \vee \neg p)$ among the p-relevant formulas. After all, $\mathbf{O}_i\langle p\rangle(p \vee \neg p)$ is satisfiable. However, the mention of p in $p \vee \neg p$ seems merely accidental since $\models (p \vee \neg p) \equiv \alpha$ for every valid α. Thus we feel justified in assuming that $p \vee \neg p$ does not convey any relevant information about p.[10] The following lemma identifies a simple case where formulas are strictly about a subject matter.

Lemma 4 *Let α be a basic i-objective formula such that $\not\models\alpha$ and $\not\models\neg\alpha$. Then α is strictly π_α-relevant.*

Finally, the following theorem allows us to characterize strictly π-relevant basic formulas without appealing to belief or only-believing-about.

[10]In this light, $\mathbf{O}_i\langle p\rangle(p \vee \neg p)$ is best understood as saying that *nothing* is known about p.

Theorem 5 *Let α be a basic i-objective formula such that $\not\models\alpha$ and let π be a subject matter. Let*

$$\Gamma = \{c \mid \text{ c is a basic i-objective e-clause such that } \models\alpha \supset c \text{ and for all basic i-objective } \beta, \text{ if } \models\alpha \supset \beta \text{ and } \beta \supset c \text{ then } \beta \text{ mentions } \pi\}$$

Then α is strictly π-relevant iff there are $c_1, \ldots, c_k \in \Gamma$ such that $\models \bigwedge_{i=1}^{k} c_i \supset \alpha$.

What Does Jill Know about Jack?

So far, we have assumed that the subject matter is a set of atomic propositions, just as in [Lak92]. In a multi-agent context, there is at least one other possible subject matter, namely other agents. In other words, we may want to ask the question what Jill believes about Jack's beliefs. For example, if all Jill believes is $p \vee (\neg \mathbf{L}_{jack} q \wedge r)$, then all Jill believes about Jack seems to be $p \vee \neg \mathbf{L}_{jack} q$. It turns out to be quite easy to formalize these ideas, requiring only minor changes to the definitions we already have.

For simplicity, we only consider the subject matter of one agent. Thus for all agents i and j with $i \neq j$, let $\mathbf{O}_i\langle j\rangle$ be a new modal operator. As before, we require that $\mathbf{O}_i\langle j\rangle$ only be applied to basic formulas.

Definition 16 *Given a world w in the canonical model M_c, let*

$$\Gamma^{w}_{j,i} = \{c \mid \text{ c is an i-objective e-clause}, w\models\mathbf{L}_i c \text{ and for all i-objective basic } \alpha, \text{ if } w\models\mathbf{L}_i\alpha \text{ and } \models\alpha \supset c \text{ then } \alpha \text{ mentions } \mathbf{L}_j\}.$$

Given a world w, let $w|_{j,i}$ be defined just as $w|_{\pi,i}$ was defined with $\Gamma^{w}_{j,i}$ now taking the role of $\Gamma^{w}_{\pi,i}$. The semantics of only-knowing about another agent is then simply:

$$w\models\mathbf{O}_i\langle j\rangle\alpha \iff w|_{j,i}\models\mathbf{O}_i\alpha \text{ and } w\models\mathbf{L}_i\alpha.$$

For example, given i's knowledge base $\mathrm{KB} = p\wedge(\neg\mathbf{L}_j q\vee q) \wedge \mathbf{L}_k\mathbf{L}_j p$, where i, j, and k denote three different agents, we obtain

$$\models\mathbf{O}_i\mathrm{KB} \supset \mathbf{O}_i\langle j\rangle[(\neg\mathbf{L}_j q \vee q) \wedge \mathbf{L}_k\mathbf{L}_j p].$$

Finally, if we were interested only in what i herself knows about j's beliefs, that is, if we want to exclude $\mathbf{L}_k\mathbf{L}_j p$ in the last example, we can do so by modifying the definition of $\Gamma^{w}_{j,i}$ in that we require that α mentions $\mathbf{L}_j$ at *depth 0*.

Conclusion

While an agent i in general does not know all another agent j knows, i may well know all j knows about a specific subject matter π, which suffices for i to infer j's nonmonotonic inferences regarding π. In this paper we formalized such a notion of only-knowing-about for two kinds of subject matters and discussed some of its properties. In addition, we were able to use our new notion of only-knowing-about to specify what it means for a formula to be strictly about a given subject matter.

As for future work, it would be desirable to obtain a complete axiomatization of only-knowing-about, simply because proof theories provide very concise characterizations. The first-order case, of course, needs to be addressed as well. We believe that our work, apart from the specific context of multi-agent autoepistemic reasoning, sheds some light on the concept of *aboutness*. However, much more remains to be done before this intriguing and difficult issue is fully understood.

References

[AK88] Appelt, D. and Konolige, K., A Practical Nonmonotonic Theory of Reasoning about Speech Acts, in *Proc. of the 26th Conf. of the ACL*, 1988.

[Hal93] Halpern, J. Y., Reasoning about only knowing with many agents, these proceedings.

[HM84] Halpern, J. Y. and Moses, Y. O., Towards a Theory of Knowledge and Ignorance: Preliminary Report, in Proceedings of The Non-Monotonic Workshop, New Paltz, NY, 1984, pp. 125–143.

[HM92] Halpern, J. Y. and Moses, Y. O., A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief, *Artificial Intelligence* **54**, 1992, pp. 319–379.

[HC84] Hughes, G. E. and Cresswell, M. J., *A Companion to Modal Logic*, Methuen & Co., London, 1984.

[Hin62] Hintikka, J., *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, 1962.

[Hin71] Hintikka, J., Semantics for Propositional Attitudes, in L. Linsky (ed.), *Reference and Modality*, Oxford University Press, Oxford, 1971.

[Kri63] Kripke, S. A., Semantical Considerations on Modal Logic, *Acta Philosophica Fennica* **16**, 1963, pp. 83–94.

[Lak92] All You Ever Wanted to Know about Tweety, in *Proc. of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, 1992, pp. 639–648.

[Lak93] All They Know: A Study in Multi-Agent Nonmonotonic Reasoning, in *Proc. of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, Morgan Kaufmann, 1993.

[Lev90] Levesque, H. J., All I Know: A Study in Autoepistemic Logic, *Artificial Intelligence*, North Holland, **42**, 1990, pp. 263–309.

[Moo85] Moore, R., Semantical Considerations on Nonmonotonic Logic, *Artificial Intelligence* **25**, 1985, pp. 75–94.

[Mor90] Morgenstern, L., A Theory of Multiple Agent Nonmonotonic Reasoning, in *Proc. of AAAI-90*, 1990, pp. 538–544.

[MG92] Morgenstern, L. and Guerreiro, R., Epistemic Logics for Multiple Agent Nonmonotonic Reasoning I, *Symposium on Formal Reasoning about Beliefs, Intentions, and Actions*, Austin, TX, 1992.

[Per87] Perrault, R., An Application of Default Logic to Speech Act Theory, in *Proc. of the Symposium on Intentions and Plans in Communication and Discourse*, Monterey, 1987.