In Praise of Belief Bases: 
Doing Epistemic Logic without Possible Worlds 

Emiliano Lorini 
CNRS-IRIT, Toulouse University, France

Abstract
We introduce a new semantics for a logic of explicit and implicit beliefs based on the concept of multi-agent belief base. Differently from existing Kripke-style semantics for epistemic logic in which the notions of possible world and doxastic/epistemic alternative are primitive, in our semantics they are non-primitive but are defined from the concept of belief base. We provide a complete axiomatization and a decidability result for our logic.

Introduction
Epistemic logic and, more generally, formal epistemology are the areas at the intersection between philosophy (Hintikka 1962), artificial intelligence (AI) (Fagin et al. 1995; Meyer and van der Hoek 1995) and economics (Lismont and Mongin 1994) devoted to the formal representation of epistemic attitudes of agents including belief and knowledge. An important distinction in epistemic logic is between explicit belief and implicit belief. According to (Levesque 1984), "...a sentence is explicitly believed when it is actively held to be true by an agent and implicitly believed when it follows from what is believed" (p. 198). This distinction is particularly relevant for the design of resource-bounded agents who spend time to make inferences and do not believe all facts that are deducible from their actual beliefs.

The concept of explicit belief is tightly connected with the concept of belief base (Nebel 1992; Makinson 1985; Hansson 1993; Rott 1998). In particular, an agent’s belief base, which is not necessarily closed under deduction, includes all facts that are explicitly believed by the agent. Nonetheless, existing logical formalizations of explicit and implicit beliefs (Levesque 1984; Fagin and Halpern 1987) do not clearly account for this connection.

The aim of this paper is to fill this gap by providing a multi-agent logic that precisely articulates the distinction between explicit belief, as a fact in an agent’s belief base, and implicit belief, as a fact that is deducible from the agent’s explicit beliefs, given the agents’ common ground. The concept of common ground (Stalnaker 2002) corresponds to the body of information that the agents commonly believe to be the case and that has has to be in the deductive closure of their belief bases. The multi-agent aspect of the logic lies in the fact that it supports reasoning about agents’ high-order beliefs, i.e., an agent’s explicit (or implicit) belief about the explicit (or implicit) belief of another agent.

Differently from existing Kripke-style semantics for epistemic logic in which the notions of possible world and doxastic/epistemic alternative are primitive, in the semantics of our logic the notion of doxastic alternative is defined from — and more generally grounded on — the concept of belief base.

We believe that an explicit representation of agents’ belief bases is crucial in order to facilitate the task of designing intelligent systems such as robotic agents or conversational agents. The problem of extensional semantics for epistemic logic, whose most representative example is the Kripkean semantics, is their being too abstract and too far from the agent specification. More generally, the main limitation of the Kripkean semantics is that it does not say from where doxastic alternatives come from thereby being ungrounded.

The paper is organized as follows. We first present the language of our logic of explicit and implicit beliefs. Then, we introduce a semantics for this language based on the notion of multi-agent belief base. We also consider two additional Kripke-style semantics in which the notion of doxastic alternative is primitive. These additional semantics will be useful for proving completeness and decidability of our logic. We show that the three semantics are all equivalent with respect to the formal language under consideration. Then, we provide an axiomatization for our logic of explicit and implicit belief and prove that its satisfiability problem is decidable. After having discussed related work, we conclude.

A Language for Explicit and Implicit Beliefs

LDA (Logic of Doxastic Attitudes) is a logic for reasoning about explicit beliefs and implicit beliefs of multiple agents. Assume a countably infinite set of atomic propositions \( \text{Atm} = \{p, q, \ldots\} \) and a finite set of agents \( \text{Agt} = \{1, \ldots, n\} \).

\footnote{The need for a grounded semantics for doxastic/epistemic logics has been pointed out by other authors including (Lomuscio, Qu, and Raimondi 2015).}
We define the language of the logic LDA in two steps. We first define the language $\mathcal{L}_0(\text{Atm})$ by the following grammar in Backus-Naur Form (BNF):

$$\alpha ::= p \mid \neg \alpha \mid \alpha_1 \land \alpha_2 \mid \mathcal{E}_i \alpha$$

where $p$ ranges over $\text{Atm}$ and $i$ ranges over $\text{Agt}$. $\mathcal{L}_0(\text{Atm})$ is the language for representing explicit beliefs of multiple agents. The formula $\mathcal{E}_i \alpha$ is read “agent $i$ explicitly (or actually) believes that $\alpha$ is true”. In this language, we can represent high-order explicit beliefs, i.e., an agent’s explicit belief about another agent’s explicit beliefs.

The language $\mathcal{L}(\text{Atm})$, extends the language $\mathcal{L}_0(\text{Atm})$ by modal operators of implicit belief and is defined by the following grammar:

$$\varphi ::= \alpha \mid \neg \varphi \mid \varphi_1 \land \varphi_2 \mid I_i \varphi$$

where $\alpha$ ranges over $\mathcal{L}_0(\text{Atm})$. For notational convenience we write $\mathcal{L}_0$ instead of $\mathcal{L}_0(\text{Atm})$ and $\mathcal{L}$ instead of $\mathcal{L}(\text{Atm})$, when the context is unambiguous.

The other Boolean constructions $\top$, $\perp$, $\lor$, $\rightarrow$ and $\leftrightarrow$ are defined from the primitive concept of belief.

For every formula $\varphi \in \mathcal{L}$, we write $\text{Atm}(\varphi)$ to denote the set of atomic propositions of type $p$ occurring in $\varphi$. Moreover, for every set of formulas $X \subseteq \mathcal{L}$, we define $\text{Atm}(X) = \bigcup_{\varphi \in X} \text{Atm}(\varphi)$.

The formula $I_i \varphi$ has to be read “agent $i$ implicitly (or potentially) believes that $\varphi$ is true”. We define the dual operator $\bar{I}_i$ as follows:

$$\bar{I}_i \varphi \overset{\text{def}}{=} \neg I_i \neg \varphi.$$  

$I_i \varphi$ has to be read “$\varphi$ is compatible with agent $i$’s implicit beliefs”.

**Formal Semantics**

In this section, we present three formal semantics for the language of explicit and implicit beliefs defined above. In the first semantics, the notion of doxastic alternative is not primitive but it is defined from the primitive concept of belief base. The second semantics is a Kripke-style semantics, based on the concept of notional doxastic model, in which an agent’s set of doxastic alternatives coincides with the set of possible worlds in which the agent’s explicit beliefs are true. The third semantics is a weaker semantics, based on the concept of quasi-notional doxastic model. It only requires that an agent’s set of doxastic alternatives has to be included in the set of possible worlds in which the agent’s explicit beliefs are true. At a later stage in the paper, we will show that three semantics are equivalent with respect to the formal language under consideration.

**Multi-agent belief base semantics**

We first consider the semantics based on the concept of multi-agent belief base that is defined as follows.

**Definition 1 (Multi-agent belief base)** A multi-agent belief base is a tuple $B = (B_1, \ldots, B_n, V)$ where:

- for every $i \in \text{Agt}$, $B_i \subseteq \mathcal{L}_0$ is agent $i$’s belief base,
- $V \subseteq \text{Atm}$ is the actual state.

A similar concept is used in belief merging (Koieczny and Pérez 2002) in which each agent is identified with her belief base. Our concept of multi-agent belief base also includes the concept of actual state, as the set of true atomics facts.

The sublanguage $\mathcal{L}_0(\text{Atm})$ is interpreted with respect to multi-agent belief bases, as follows.

**Definition 2 (Satisfaction relation)** Let $B = (B_1, \ldots, B_n, V)$ be a multi-agent belief base. Then:

$$B \models p \iff p \in V$$
$$B \models \neg \alpha \iff B \not\models \alpha$$
$$B \models \alpha_1 \land \alpha_2 \iff B \models \alpha_1 \text{ and } B \models \alpha_2$$
$$B \models \mathcal{E}_i \alpha \iff \alpha \in B_i$$

The following definition introduces the concept of doxastic alternative.

**Definition 3 (Doxastic alternatives)** Let $B = (B_1, \ldots, B_n, V)$ and $B' = (B'_1, \ldots, B'_n, V')$ be two multi-agent belief bases. Then, $B \mathcal{R}_i B'$ if and only if, for every $\alpha \in B_i$, $B' \models \alpha$.

$B \mathcal{R}_i B'$ means that $B'$ is a doxastic alternative for agent $i$ at $B$ (i.e., at $B$ agent $i$ considers $B'$ possible). The idea of the previous definition is that $B'$ is a doxastic alternative for agent $i$ at $B$ if and only if, $B'$ satisfies all facts that agent $i$ explicitly believes at $B$.

A multi-agent belief model (MAB) is defined to be a multi-agent belief base supplemented with a set of multi-agent belief bases, called context. The latter includes all multi-agent belief bases that are compatible with the agents’ common ground (Stalnaker 2002), i.e., the body of information that the agents commonly believe to be the case.

**Definition 4 (Multi-agent belief model)** A multi-agent belief model (MAB) is a pair $(B, \text{Cxt})$, where $B$ is a multi-agent belief base and $\text{Cxt}$ is a set of multi-agent belief bases. The class of MABs is denoted by MAB.

Note that in the previous definition we do not require $B \in \text{Cxt}$. Let us illustrate the concept of MAB with an example.

**Example 1** Let $\text{Agt} = \{1, 2\}$ and $\text{Atm} = \{p, q\}$. Moreover, let $(B_1, B_2, V)$ such that:

- $B_1 = \{p, E_2 p\}$,
- $B_2 = \{p\}$,
- $V = \{p, q\}$.

Suppose that the agents have in their common ground the fact $p \rightarrow q$. In other words, they commonly believe that $p$ implies $q$. This means that:

$$\text{Cxt} = \{B' : B' \models p \rightarrow q\}.$$  

The following definition generalizes Definition 2 to the full language $\mathcal{L}(\text{Atm})$. Its formulas are interpreted with respect to MABs. (Boolean cases are omitted, as they are defined in the usual way.)

**Definition 5 (Satisfaction relation (cont.))** Let $(B, \text{Cxt}) \in \text{MAB}$. Then:

$$B, \text{Cxt} \models \alpha \iff B \models \alpha$$
$$B, \text{Cxt} \models I_i \varphi \iff \forall B' \in \text{Cxt} : \text{if } B \mathcal{R}_i B' \text{ then } (B', \text{Cxt}) \models \varphi$$
Let us go back to the example.

**Example 2** It is to check that the following holds:

\[(B, Cxt) \models I_1(p \land q) \land I_2(p \land q) \land I_1I_2(p \land q).\]

Indeed, we have:

\[
\begin{align*}
R_1(B) \cap Cxt &= \{B' : B' \models p \land \exists 2p \land (p \to q)\}, \\
R_2(B) \cap Cxt &= \{B' : B' \models p \land (p \to q)\},
\end{align*}
\]

and, consequently,

\[
\left( R_1 \circ R_2(B) \right) \cap Cxt = \{B' : B' \models p \land (p \to q)\},
\]

where \( \circ \) is the composition operation between binary relations and \( R_i(B) = \{B' : BR_iB'\} \).

Here, we consider consistent MABs that guarantee consistency of the agents’ belief bases. Specifically:

**Definition 6** (Consistent MAB) \((B, Cxt)\) is a consistent MAB (CMAB) if and only if, for every \(B' \in Cxt \cup \{B\}\), there exists \(B'' \in Cxt\) such that \(B'R_1B''\). The class of CMABs is denoted by CMAB.

Let \(\varphi \in \mathcal{L}\), we say that \(\varphi\) is valid for the class of CMABs if and only if, for every \((B, Cxt) \in \text{CMAB}\), we have \((B, Cxt) \models \varphi\). We say that \(\varphi\) is satisfiable for the the class of CMABs if and only if \(\neg \varphi\) is not valid for the the class of CMABs.

**Notional doxastic model semantics**

Let us now consider the semantics for LDA based on the concept of notional doxastic model (NDM). It is defined in the next Definition 7, together with the satisfaction relation for the formulas of the language \(\mathcal{L}(Atm)\).

**Definition 7** (Doxastic model) A notional doxastic model (NDM) is a tuple \(M = (W, D, \mathcal{N}, \mathcal{V})\) where:

- \(W\) is a set of worlds,
- \(D : \text{Agt} \times W \to 2^W\) is a doxastic function,
- \(\mathcal{N} : \text{Agt} \times W \to 2^W\) is a notional function, and
- \(\mathcal{V} : \text{Atm} \to 2^W\) is a valuation function,

and that satisfies the following conditions for all \(i \in \text{Agt}\) and \(w \in W\):

1. \(\mathcal{N}(i, w) = \bigcap_{\alpha \in D(i, w)} ||\alpha||_M\), and
2. there exists \(v \in W\) such that \(v \in \mathcal{N}(i, w)\), with:

\[
\begin{align*}
(M, w) &\models p \iff w \in \mathcal{V}(p) \\
(M, w) &\models \neg \varphi \iff (M, w) \not\models \varphi \\
(M, w) &\models \varphi \land \psi \iff (M, w) \models \varphi \text{ and } (M, w) \models \psi \\
(M, w) &\models \exists_i \alpha \iff \alpha \in D(i, w) \\
(M, w) &\models \mathcal{I}_i \varphi \iff \forall v \in \mathcal{N}(i, w) : (M, v) \models \varphi
\end{align*}
\]

and

\[
||\alpha||_M = \{v \in W : (M, v) \models \alpha\}.
\]

The class of notional doxastic models is denoted by NDM.

We say that a NDM \(M = (W, A, D, N, V)\) is finite if and only if \(W, D(i, w)\) and \(V^{-1}(w)\) are finite sets for every \(i \in \text{Agt}\) and for every \(w \in W\), where \(V^{-1}\) is the inverse function of \(V\).

For every agent \(i\) and world \(w\), \(D(i, w)\) denotes agent \(i\)’s set of explicit beliefs at \(w\).

The set \(\mathcal{N}(i, w)\), used in the interpretation of the implicit belief operator \(\mathcal{I}_i\), is called agent \(i\)’s set of notional worlds at world \(w\). The term ‘notional’ is taken from (Dennett 1996; 1988) (see, also, (Konolige 1986)): an agent’s notional world is a world in which all the agent’s explicit beliefs are true. This idea is clearly expressed by the Condition C1. According to the Condition C2, an agent’s set of notional worlds must be non-empty. This guarantees consistency of the agent’s implicit beliefs.

Let \(\varphi \in \mathcal{L}\), we say that \(\varphi\) is valid for the the class of NDMs if and only if, for every \(M = (W, A, D, N, V) \in \text{NDM}\) and for every \(w \in W\), we have \((M, w) \models \varphi\). We say that \(\varphi\) is satisfiable for the the class of NDMs if and only if \(\neg \varphi\) is not valid for the the class of NDMs.

**Quasi-model semantics**

In this section we provide an alternative semantics for the logic LDA based on a more general class of models, called quasi-notional doxastic models (quasi-NDMs). This semantics will be fundamental for proving completeness of LDA.

**Definition 8** (Quasi-notional doxastic model) A quasi-notional doxastic model (quasi-NDM) is a tuple \(M = (W, D, N, V)\) where \(W, D, N, \mathcal{V}\) and \(V\) are as in Definition 7 except that Condition C1 is replaced by the following weaker condition, for all \(i \in \text{Agt}\) and \(w \in W\):

\((C1^*)\) \(\mathcal{N}(i, w) \subseteq \bigcap_{\alpha \in D(i, w)} ||\alpha||_M\).

The class of quasi-notional doxastic models is denoted by QNDM. Truth conditions of formulas in \(\mathcal{L}\) relative to this class are the same as truth conditions of formulas in \(\mathcal{L}\) relative to the class NDM. Validity and satisfiability of a LDA formula \(\varphi\) for the the class of quasi-NDMs are defined in the usual way.

As for NDMs, we say that a quasi-NDM \(M = (W, D, N, V)\) is finite if and only if \(W, D(i, w)\) and \(V^{-1}(w)\) are finite sets for every \(i \in \text{Agt}\) and for every \(w \in W\).

**Equivalences between semantics**

The present section is devoted to present equivalences between the different semantics for the language \(\mathcal{L}(Atm)\). The results of the section are summarized in Figure 1.

The figure highlights that the five semantics for the language \(\mathcal{L}(Atm)\) defined in the previous section are all equivalent, as from every node in the graph we can reach all other nodes.

**Equivalence between quasi-NDMs and finite quasi-NDMs** We use a filtration argument to show that if a formula \(\varphi\) of the language \(\mathcal{L}\) is true in a (possibly infinite) quasi-NDM then it is true in a finite quasi-NDM.
Figure 1: Relations between semantics. An arrow means that satisfiability relative to the first class of structures implies satisfiability relative to the second class of structures. Dotted arrows denote relations that follow straightforwardly given the inclusion between classes of structures.

Let $M = (W, \mathcal{D}, \mathcal{N}, \mathcal{V})$ be a (possibly infinite) quasi-NDM and let $\Sigma \subseteq \mathcal{L}$ be an arbitrary finite set of formulas which is closed under subformulas. (Cf. Definition 2.35 in Blackburn, de Rijke, and Venema 2001) for a definition of subformulas closed set of formulas.) Let the equivalence relation $\equiv_\Sigma$ on $W$ be defined as follows. For all $w, v \in W$:

$$w \equiv_\Sigma v \text{ iff } \forall \varphi \in \Sigma : (M, w) \models \varphi \iff (M, v) \models \varphi.$$  

Let $|w|_\Sigma$ be the equivalence class of the world $w$ with respect to the equivalence relation $\equiv_\Sigma$.

We define $W_\Sigma$ to be the filtrated set of worlds with respect to $\Sigma$:

$$W_\Sigma = \{|w|_\Sigma : w \in W\}.$$  

Clearly, $W_\Sigma$ is a finite set.

Let us define the filtrated valuation function $\nu_\Sigma$. For every $p \in \text{Atm}(\Sigma)$, we define:

$$\nu_\Sigma(p) = \{|w|_\Sigma : (M, w) \models p\} \quad \text{if } p \in \text{Atm}(\Sigma) \quad \text{and otherwise}.$$  

The next step in the construction consists in defining the filtrated doxastic function. For every $i \in \text{Agt}$ and for every $|w|_\Sigma \in W_\Sigma$, we define:

$$\mathcal{D}_\Sigma(i, |w|_\Sigma) = \mathcal{D}(i, w) \cap \Sigma.$$  

Finally, for every $i \in \text{Agt}$ and for every $|w|_\Sigma \in W_\Sigma$, we define agent $i$’s set of notional worlds at $|w|_\Sigma$ as follows:

$$\mathcal{N}_\Sigma(i, |w|_\Sigma) = \{|v|_\Sigma : v \in \mathcal{N}(i, w)\}.$$  

We call the model $M_\Sigma = (W_\Sigma, \mathcal{A}_\Sigma, \mathcal{D}_\Sigma, \mathcal{N}_\Sigma, \nu_\Sigma)$ the filtrated of $M$ under $\Sigma$.

We can state the following filtration lemma.

**Lemma 1** Let $\varphi \in \Sigma$ and let $w \in W$. Then, $(M, w) \models \varphi$ if and only if $(M_\Sigma, |w|_\Sigma) \models \varphi$.

**Proof**. The proof is by induction on the structure of $\varphi$. For the ease of exposition, we prove our result for the language $\mathcal{L}$ in which the “diamond” operator $\Diamond_i$ is taken as primitive and the “box” operator $\square_i$ is defined from it. Since the two operators are inter-definable, this does not affect the validity of our result.

The case $\varphi = p$ is immediate from the definition of $\nu_\Sigma$. The boolean cases $\varphi = \neg \psi$ and $\varphi = \psi \land \psi_2$ follow straightforwardly from the fact that $\Sigma$ is closed under subformulas. This allows us to apply the induction hypothesis.

Let us prove the case $\varphi = E_i I$.  

(\Rightarrow) Suppose $(M, w) \models E_i I$ with $E_i I \in \Sigma$. Thus, $\alpha \in \mathcal{D}(i, w)$. Hence, by definition of $\mathcal{D}_\Sigma(i, |w|_\Sigma)$ and the fact that $\Sigma$ is closed under subformulas, we have $\alpha \in \mathcal{D}_\Sigma(i, |w|_\Sigma)$. It follows that $(M_\Sigma, |w|_\Sigma) \models E_i I$.

(\Leftarrow) For the other direction, suppose $(M_\Sigma, |w|_\Sigma) \models E_i I$ with $E_i I \in \Sigma$. Thus, $\alpha \in \mathcal{D}_\Sigma(i, |w|_\Sigma)$. Hence, by definition of $\mathcal{D}_\Sigma(i, |w|_\Sigma)$, $\alpha \in \mathcal{D}(i, w)$.

Let us conclude the proof for the case $\varphi = I_i \psi$. It is easy to check that $\mathcal{N}_\Sigma$ gives rise to the smallest filtration and that the following two properties hold for all $w, v \in W$ and for all $i \in \text{Agt}$:

(i) if $v \in \mathcal{N}(i, w)$ then $|v|_\Sigma \in \mathcal{N}_\Sigma(i, |w|_\Sigma)$, and

(ii) if $|v|_\Sigma \in \mathcal{N}_\Sigma(i, |w|_\Sigma)$ then $I_i \psi \in \Sigma$, if $M, v \models \varphi$ then $(M, w) \models \varphi$.

(\Rightarrow) Suppose $(M, w) \models I_i \psi$ with $I_i \psi \in \Sigma$. Thus, there exists $v \in \mathcal{N}(i, w)$ such that $(M, v) \models \psi$. By the previous item (i), $|v|_\Sigma \in \mathcal{N}_\Sigma(i, |w|_\Sigma)$. Since $\Sigma$ is closed under subformulas, we have $\psi \in \Sigma$. Thus, by the induction hypothesis, $(M_\Sigma, |w|_\Sigma) \models \psi$. It follows that $(M_\Sigma, |w|_\Sigma) \models I_i \psi$.

(\Leftarrow) For the other direction, suppose $(M_\Sigma, |w|_\Sigma) \models I_i \psi$ with $I_i \psi \in \Sigma$. Thus, there exists $|v|_\Sigma \in \mathcal{N}_\Sigma(i, |w|_\Sigma)$, such that $(M_\Sigma, |v|_\Sigma) \models \psi$. Since $\Sigma$ is closed under subformulas, by the induction hypothesis, we have $(M, v) \models \psi$. By the item (ii) above, it follows that $(M, w) \models I_i \psi$.

The next step consists in proving that $M_\Sigma$ is the right model construction.

**Proposition 1** The tuple $M_\Sigma = (W_\Sigma, \mathcal{A}_\Sigma, \mathcal{D}_\Sigma, \mathcal{N}_\Sigma, \nu_\Sigma)$ is a finite quasi-NDM.

**Proof**. Clearly, $M_\Sigma$ is finite. Moreover, it is easy to verify that it satisfies the Condition C2 in Definition 7. We are going to prove that it satisfies the Condition C1* in Definition 8.

By Lemma 1, if $\alpha \in \mathcal{D}(i, w) \cap \Sigma$ then $|\alpha|_{M_\Sigma} = \{|v|_\Sigma : v \in |\alpha|_M\}$. Moreover, as $M$ is a quasi-NDM, we have

$$\mathcal{N}(i, w) \subseteq \bigcap_{\alpha \in \mathcal{D}(i, w) \cap \Sigma} |\alpha|_{M_\Sigma}.$$  

Hence, by definition of $\mathcal{N}_\Sigma(i, |w|_\Sigma)$ and $\mathcal{D}_\Sigma$,

$$\mathcal{N}_\Sigma(i, |w|_\Sigma) \subseteq \bigcap_{\alpha \in \mathcal{D}_\Sigma(i, |w|_\Sigma)} |\alpha|_{M_\Sigma}.$$  

The following is our first result about equivalence between the semantics in terms of quasi-NDMs and the semantics in terms of finite quasi-NDMs.

**Theorem 1** Let $\varphi \in \mathcal{L}$. Then, if $\varphi$ is satisfiable for the class of quasi-NDMs, if and only if it is satisfiable for the class of finite quasi-NDMs.
Equivalence between finite NDMs and finite quasi-NDMs

As the following theorem highlights, the LDA semantics in terms of finite NDMs and the LDA semantics in terms of finite quasi-NDMs are equivalent.

Theorem 2 Let \( \varphi \in \mathcal{L} \). Then, \( \varphi \) is satisfiable for the class of finite NDMs if and only if \( \varphi \) is satisfiable for the class of finite quasi-NDMs.

Proof. The left-to-right direction is obvious. Since \( \varphi \) is satisfiable for the class of finite quasi-NDMs, it is also satisfiable for the class of finite NDMs.

PROOF. The right-to-left direction is obvious. As for the left-to-right direction, let \( M \) be a possibly infinite quasi-NDM and let \( w \) be a world in \( M \) such that \( (M, w) \models \varphi \). Moreover, let \( \text{sub}(\varphi) \) be the set of subformulas of \( \varphi \). Then, by Lemma 1 and Proposition 1, \( (M_{\text{sub}(\varphi)},|w|_{\text{sub}(\varphi)}) \models \varphi \) and \( M_{\text{sub}(\varphi)} \) is a finite quasi-NDM.

Equivalence between CMABs and NDMs

Our third equivalence result is between CMABs and NDMs.

Theorem 3 Let \( \varphi \in \mathcal{L} \). Then, \( \varphi \) is satisfiable for the class of CMABs if and only if \( \varphi \) is satisfiable for the class of NDMs.

Proof. We first prove the left-to-right direction. Let \( (B, \text{Cxt}) \) be a CMAB with \( B = (B_1, \ldots, B_n, \mathcal{V}) \) and such that \((B, \text{Cxt}) \models \varphi \). We define the structure \( (M, \mathcal{D}, \mathcal{N}, \mathcal{V}) \) as follows:

- \( \mathcal{W} \cap \mathcal{V} \subseteq \mathcal{D} \cap \mathcal{A} \)
- For every \( e \in \mathcal{E} \), we define \( \mathcal{N}(e, w) = \mathcal{N}(e, w) \cap \mathcal{W} \cap \mathcal{V} \)
- For every \( e \in \mathcal{E} \), we define \( \mathcal{A}(e, w) = \mathcal{A}(e, w) \cap \mathcal{W} \cap \mathcal{V} \)
- For every \( e \in \mathcal{E} \), we define \( \mathcal{D}(e, w) = \mathcal{D}(e, w) \cap \mathcal{W} \cap \mathcal{V} \)
- For every \( e \in \mathcal{E} \), we define \( \mathcal{L}(e, w) = \mathcal{L}(e, w) \cap \mathcal{W} \cap \mathcal{V} \)

One can show that \( M \) so defined is a NDM. Moreover, by induction on the structure of \( \varphi \), one can prove that, for all \( w_B \in \mathcal{W} \), \( M, w_B \models \varphi \) iff \((B, \text{Cxt}) \models \varphi \). Thus, \((M, w_B) \models \varphi \).

We now prove the right-to-left direction. Let \( (M, \mathcal{D}, \mathcal{N}, \mathcal{V}) \) be a NDM and let \( w \) be a world in \( M \) such that \( (M, w) \models \varphi \). We let \( \mathcal{W} \subseteq \mathcal{V} \subseteq \mathcal{D} \subseteq \mathcal{A} \) and, for all \( v \in \mathcal{W} \), define \( \mathcal{N}(v, w) = \mathcal{N}(v, w) \cap \mathcal{W} \cap \mathcal{V} \) and \( \mathcal{A}(v, w) = \mathcal{A}(v, w) \cap \mathcal{W} \cap \mathcal{V} \). It is straightforward to show that if \( \varphi \) is satisfiable for the class of NDMs then \( \varphi \) is satisfiable for the class of non-redundant NDMs. Thus, from the initial model \( M \), we can find a non-redundant NDM \( M' = (\mathcal{W}', \mathcal{D}', \mathcal{N}', \mathcal{V}') \) and \( v \in \mathcal{W}' \) such that \( (M', v) \models \varphi \). For every \( u \in \mathcal{W}' \), we define \( B_u = (B_1^u, \ldots, B_n^u, \mathcal{V}^u) \) such that \( B_u \models (M, w) \) for every \( e \in \mathcal{E} \) and \( \mathcal{V}^u = \mathcal{V}^u \cap \mathcal{W}' \). Moreover, we define the context \( \text{Cxt}' = \{ B_u : u \in \mathcal{W}' \} \). One can show that, for every \( B_u \in \text{Cxt}' \), \( (B_u, \text{Cxt}') \) is a CMAB. The fact that \( M' \) is non-redundant is essential to guarantee that there is a one-to-one correspondence between \( \mathcal{W}' \) and \( \mathcal{W} \). By induction on the structure of \( \varphi \), one can prove that, for all \( B_u \in \text{Cxt}' \), \((B_u, \text{Cxt}') \models \varphi \) iff \( M', u \models \varphi \). Thus, \( B_u \models \varphi \).

Axiomatics and decidability

This section is devoted to provide an axiomatic definition and a decidability result for LDA. To this aim, we first provide a formal definition of this logic.
Definition 9. We define LDA to be the extension of classical propositional logic given by the following axioms and rule of inference:

\[(I_1 \varphi \land I_1 (\varphi \rightarrow \psi)) \rightarrow I_1 \psi \quad (K_{I_1})\]
\[\neg (I_1 \varphi \land I_1 \neg \varphi) \quad (D_{I_1})\]
\[E_\alpha \rightarrow I_1 \alpha \quad (Int_{E_1 I_1})\]
\[\varphi \quad I_1 \varphi \quad (Nec_{I_1})\]

We denote that \(\varphi\) is derivable in LDA by \(\vdash_{LDA} \varphi\). We say that \(\varphi\) is LDA-consistent if \(\nvdash_{LDA} \neg \varphi\).

The logic LDA includes the principles of system KD for the implicit belief operator \(I_i\) as well as an axiom \(Int_{E_1 I_1}\) relating explicit belief with implicit belief. Note that there is no consensus in the literature about introspection for implicit belief. For instance, in his seminal work on the logics of knowledge and belief (Hintikka 1962), Hintikka only assumed positive introspection for belief (Axiom 4) and rejected negative introspection (Axiom 5). Other logicians such as (Jones 2015) have argued against the use of both positive and negative introspection axioms for belief. Nonetheless, all approaches unanimously assume that a reasonable notion of implicit belief should satisfy Axioms K and D. In this sense, system KD can be conceived as the minimal logic of implicit belief. On this point, see (Banerjee and Dubois 2014).

To prove our main completeness result, we first prove a theorem about soundness and completeness of LDA for the class of quasi-NDMs.

Soundness and completeness for quasi-NDMs. To prove completeness of LDA for the class of quasi-NDMs, we use a canonical model argument.

We consider maximally LDA-consistent sets of formulas in \(L\) (MCSs). The following proposition specifies some usual properties of MCSs.

Proposition 2. Let \(\Gamma\) be a MCS and let \(\varphi, \psi \in L\). Then:

- \(\varphi, \psi \rightarrow \psi \in \Gamma\) then \(\psi \in \Gamma\);
- \(\varphi \in \Gamma\) or \(\neg \varphi \in \Gamma\);
- \(\varphi \land \psi \in \Gamma\) iff \(\varphi \in \Gamma\) or \(\psi \in \Gamma\).

The following is the Lindenbaum’s lemma for our logic. Its proof is standard (cf. Lemma 4.17 in (Blackburn, de Rijke, and Venema 2001)) and we omit it.

Lemma 2. Let \(\Delta\) be a LDA-consistent set of formulas. Then, there exists a MCS \(\Gamma\) such that \(\Delta \subseteq \Gamma\).

The next step in the proof consists in stating the following existence lemma. The proof is again standard (cf. Lemma 4.20 in (Blackburn, de Rijke, and Venema 2001)) and we omit it.

Lemma 3. Let \(\varphi \in L\) and let \(w \in W^c\). Then, if \(I_i \varphi \in w\) then there exists \(v \in N^c(i, w)\) such that \(\varphi \in v\).

Then, we prove the following truth lemma.

Lemma 4. Let \(\varphi \in L\) and let \(w \in W^c\). Then, \(M^c, w \models \varphi\) iff \(\varphi \in w\).

The proof is by induction on the structure of the formula. The cases with \(\varphi\) atomic, Boolean, and of the form \(I_1 \psi\) are provable in the standard way by means of Proposition 2 and Lemma 3 (cf. Lemma 4.21 in (Blackburn, de Rijke, and Venema 2001)). The proof for the case \(\varphi = E_\alpha\) goes as follows: \(E_\alpha \models w\) iff \(\alpha \in D^c(i, w)\) iff \(M^c, w \models E_\alpha\).

The last step consists in proving that the canonical model belongs to the class QNDM.

Proposition 3. \(M^c\) is a quasi-NDM.

Proof. Thanks to Axiom \((D_{I_1})\), it is easy to prove that \(M^c\) satisfies Condition C2 in Definition 7.

Let us prove that it satisfies Condition C1* in Definition 8. To this aim, we just need to prove that if \(\alpha \in D^c(i, w)\) then \(N^c(i, w) \subseteq ||\alpha||_{M^c}\). Suppose \(\alpha \in D^c(i, w)\). Thus, \(E_\alpha \models w\). Hence, by Axiom \((Int_{E_1 I_1})\) and Proposition 2, \(E_\alpha \models \alpha\). By the definition of \(M^c\), it follows that, for all \(v \in N^c(i, w)\), \(\alpha \in v\). Thus, by Lemma 4, for all \(v \in N^c(i, w)\), \((M^c, v) \models \alpha\). The latter means that \(N^c(i, w) \subseteq ||\alpha||_{M^c}\).

The following is our first intermediate result.

Theorem 4. The logic LDA is sound and complete for the class of quasi-NDMs.

Proof. As for soundness, it is routine to check that the axioms of LDA are all valid for the class of quasi-NDMs and that the rule of inference \((Nec_{I_1})\) preserves validity.

As for completeness, suppose that \(\varphi\) is a LDA-consistent formula in \(L\). By Lemma 2, there exists \(w \in W^c\) such that \(\varphi \in w\). Hence, by Lemma 4, there exists \(w \in W^c\) such that \(M^c, w \models \varphi\). Since, by Proposition 3, \(M^c\) is a quasi-NDM, we can conclude that \(\varphi\) is satisfiable for the class of quasi-NDM.

Soundness and completeness for NDMS and CMAbs.

We can state the two main results of this section. The first is about soundness and completeness for the class of NDMs.

Theorem 5. The logic LDA is sound and complete for the class of NDMs.

Proof. It is routine exercise to verify that LDA is sound for the class of NDMs. Now, suppose that formula \(\varphi\) is LDA-consistent. Then, by Theorems 4 and 1, it is satisfiable for the class of finite quasi-NDMs. Hence, by Theorem 2, it is satisfiable for the class of finite NDMs. Thus, more generally, \(\varphi\) is satisfiable for the class of NDMs.
The second is about soundness and completeness for the class of CMABs.

**Theorem 6** The logic LDA is sound and complete for the class of CMABs.

**Sketch of Proof.** The theorem is provable by means of Theorem 5 and Theorem 3.

**Decidability.** The second main result of this section is decidability of LDA.

**Theorem 7** The satisfiability problem of LDA is decidable.

**Proof.** Suppose \( \varphi \) is satisfiable for the class of NDMs. Thus, by Theorem 5, it is LDA-consistent. Hence, by Theorem 4, it is satisfiable for the class of quasi-NDMs. From the proof of Theorem 1, we can observe that if \( \varphi \) is satisfiable for the class of quasi-NDMs then there exists a quasi-NDM satisfying \( \varphi \) such that (i) its set of worlds contains at most \( 2^n \) elements, (ii) the atomic propositions outside \( \text{Atm}(\text{sub}(\varphi)) \) are false everywhere in the model, and (iii) the belief base of an agent at a world contains only formulas from \( \text{sub}(\varphi) \), where \( n \) is the size of \( \text{sub}(\varphi) \). The construction in the proof of Theorem 2 ensures that from this finite quasi-NDM, we can build a finite NDM satisfying \( \varphi \) for which (i) holds and such that (iv) the atomic propositions outside \( \text{Atm}(\text{sub}(\varphi)) \cup X \) are false everywhere in the model, and (v) the belief base of an agent at a world contains only formulas from \( \text{sub}(\varphi) \cup X \), where \( X \) is an arbitrary set of atoms from \( \text{Atm}(\varphi) \) of size at most \( 2^n \times |\text{Agt}| \). Thus, in order to verify whether \( \varphi \) is satisfiable, we fix a \( X \) and check satisfiability of \( \varphi \) for all NDMs satisfying (i), (iv) and (v). There are finitely many NDMs of this kind.

**Related work**

The present work lies in the area of logics for non-omniscient agents. Purely syntactic approaches to the logical omniscience problem have been proposed in which an agent’s beliefs are described either by a set of formulas which is not necessarily closed under deduction (Eberle 1974; Moore and Hendrix 2011) or by a set of formulas obtained by the application of an incomplete set of deduction rules (Konolige 1986). Logics of time-bounded reasoning have also been studied (Alechina, Logan, and Whitney 2004; Grant, Kraus, and Perlis 2000), in which reasoning is a represented as a process that requires time due to the time-consuming application of inference rules. Finally, logics of (un)awareness have been studied both in AI (Fagin and Halpern 1987; van Ditmarsch and French 2014; Ágotnes and Alechina 2014) and economics (Modica and Rustichini 1994; Heifetz, Meyer, and Schipper 2006; Halpern and Rêgo 2009).

The closest system to our logic LDA is the logic of local reasoning by (Fagin and Halpern 1987) in which the distinction between explicit and implicit beliefs is also captured. Fagin & Halpern (F&H) use a neighborhood semantics for explicit belief: every agent is associated with a set of sets of worlds, called frames of mind. They define an agent’s set of doxastic alternatives as the intersection of the agent’s frames of mind. According to F&H’s semantics, an agent explicitly believes that \( \varphi \) if and only if she has a frame of mind in which \( \varphi \) is globally true. Moreover, an agent implicitly believes that \( \varphi \) if and only if, \( \varphi \) is true at all her doxastic alternatives. In their semantics, there is no representation of an agent’s belief base, corresponding to the set of formulas explicitly believed by the agent. Moreover, differently from our notion of explicit belief, their notion does not completely solve the logical omniscience problem. For instance, while their notion of explicit belief is closed under logical equivalence, our notion is not. Specifically, the following rule of equivalence preserves validity in F&H’s logic but not in our logic:

\[
\frac{\alpha \leftrightarrow \alpha'}{
E_i \alpha \leftrightarrow E_i \alpha'}
\]

This is a consequence of their use of an extensional semantics for explicit belief. Levesque too provides an extensional semantics for explicit belief with no connection with the notion of belief base (Levesque 1984). In his logic, explicit beliefs are closed under conjunction, while they are not in our logic LDA.

**Conclusion**

We have presented a logic of explicit and implicit beliefs with a semantics based on belief bases. In the future, we plan to study a variant of this logic in which explicit and implicit beliefs are replaced by *truthful* explicit and implicit knowledge. At the semantic level, we will move from multi-agent belief bases to multi-agent knowledge bases in which the epistemic accessibility relation \( R_i \) is assumed to be reflexive. The logic will include the following extra-axiom:

\[
I_i \varphi \rightarrow \varphi \quad (T_1)
\]

We also expect to study a variant of our logic with the following extra-axioms of positive and negative introspection for implicit beliefs:

\[
\begin{align*}
I_i \varphi \rightarrow I_i I_i \varphi & \quad (4_1) \\
\neg I_i \varphi \rightarrow I_i \neg I_i \varphi & \quad (5_1)
\end{align*}
\]

Moreover, we plan to extend the logic LDA and its epistemic variant by concepts of distributed belief and distribute knowledge.

We will also investigate dynamic extensions of LDA and its epistemic variant by public announcements (Plaza 1989). The core idea is that a public announcement directly affects explicit beliefs. Given the connection between an agent’s belief base and her implicit beliefs, it should indirectly affect the latter.

Last but not least, we plan to study the model checking problem for our logic by using the compact representation offered by multi-agent belief models, as defined in Definition 4. As shown by (van Bentham et al. 2015), the possibility of using compact models could be beneficial for model checking.
References


