An Epistemic Theory of Deductive Arguments

Emiliano Lorini
IRIT, CNRS, Toulouse University, France
Emiliano.Lorini@irit.fr

Abstract

Epistemic logic and the theory of argumentation have only very recently started to interact, despite the central role that the epistemic view of argument plays in contemporary epistemology. In this paper, we present a novel epistemic language for reasoning about three types of beliefs of agents: explicit belief, plain implicit belief, and focused implicit belief. We use it to represent the concept of deductive argument and to elucidate its connection with the concept of belief. Our language is interpreted through a formal semantics that relies on belief bases. This semantics allows us to naturally represent the reasons an agent has for believing something, which we show to be closely related to the notion of argument. We provide results on expressiveness, axiomatization and decidability for the language.

1 Introduction

Epistemic logic and formal argumentation are two prominent areas of research in artificial intelligence that have traditionally developed in relative isolation from each other. Epistemic logic focuses on the formal representation of agents' epistemic states, particularly knowledge and belief (Hintikka 1962; Meyer and van der Hoek 1995; Fagin et al. 1995), while argumentation theory investigates the structure and dynamics of arguments, both in terms of logical inference and dialogical interaction (Dung 1995; Walton and Krabbe 1995; Bench-Capon and Dunne 2007). Despite their separate trajectories, a growing body of work in philosophy and cognitive science has emphasized a fundamental link between beliefs and arguments. For instance, Mercier & Sperber (Mercier and Sperber 2017; Mercier and Sperber 2011) propose that the formation of arguments originates at the cognitive level, grounded in the agent's pre-existing beliefs and inferential capabilities. In informal logic (Groarke 2017), the notion of argument has been shown to be intimately connected with the notions of evidence and reason that have an explicit epistemic connotation: an argument can be seen as a kind of reason in support of a conclusion or a claim.

In recent times, there have been several attempts to integrate epistemic logic with formal argumentation. Two traditions can be distinguished. The first includes works that focus on the formal representation of agents' beliefs and uncertainty about argumentation frameworks, e.g., belief and uncertainty about the attack relations between arguments. The second tradition aims to use concepts from argumentation theory to develop natural notions of justified belief, in line with research in formal epistemology on justification logic (Artemov 2008) and the logic of evidence-based beliefs (van Benthem and Pacuit 2011; van Benthem, Duque, and Pacuit 2014). We will discuss related works in the two traditions in more detail in Section 2.

In this paper, we follow the second tradition by proposing an integration of epistemic logic with Besnard & Hunter (B&H)'s deductive theory of arguments (Besnard and Hunter 2001; Besnard and Hunter 2008). This theory offers a key advantage over abstract argumentation by making the internal logical structure of arguments explicit, allowing for more fine-grained reasoning about how conclusions follow from premises, whereas abstract argumentation treats arguments as atomic entities without analyzing their content. It is also well-suited for integration with epistemic logic, as it defines arguments from propositional belief bases. This aligns naturally with epistemic logic, which—being a branch of modal logic—extends propositional logic. At the conceptual level, our integration consists in "internalizing" the notion of deductive argument at the agent's epistemic level. We assume that each agent has their own private belief base, and that an argument is simply some information in the agent's belief base that supports a certain conclusion.

To achieve this integration formally, we rely on the belief base semantics for epistemic logic presented in (Lorini 2018; Lorini 2020). In this semantics, the states (or worlds) in a model are not treated as primitive entities, as in standard epistemic logic semantics, but are instead decomposed into two components: the belief bases of the agents and a propositional valuation. Moreover, the agents' doxastic accessibility relations between possible states are not given as primitives but are computed from their belief bases. In line with (Konolige 1986), this semantics distinguishes an agent's explicit beliefs (i.e., the information in the agent's belief base) from implicit ones (i.e., the information that is deducible from the agent's belief base). The possibility of using a semantics for epistemic logic in which agents' belief bases are represented is key to capturing B&H's notion of deductive argument, which requires modeling the information in a belief base that supports a given conclusion.

We will leverage this semantics to interpret a novel epistemic language for reasoning about three types of belief: explicit belief, plain implicit belief, and focused implicit belief. While plain implicit belief captures what an agent can deduce from their belief base, focused implicit belief captures what an agent can deduce when focusing on a part of their belief base. We will provide results on expressiveness, axiomatization, and decidability for the language. We will express the concept of deductive argument in the epistemic language and study its connection with the concept of belief. Furthermore, we will use the semantics to establish the formal connection between the notion of deductive argument and a notion of substantive reason coming from the logical theory of evidence-based beliefs.

Our contribution can also be seen as a generalization of B&H's theory to both higher-order information and the multi-agent setting. In fact, our semantics allows for the representation of arguments from different agents, as well as arguments that involve, in both their premises and supported conclusions, information about the beliefs of other agents. The following example clarifies this point.

Example 1. *An argument for Ann to conclude that her husband Bob is going to fix her bike tire—so that she can use the bike to go to work—may consist in Ann explicitly believing that Bob explicitly believes the tire is flat, and Ann explicitly believing that if Bob explicitly believes the tire is flat, he will fix it. This argument could be counterbalanced by an argument supporting the opposite conclusion: that Bob will not fix the tire. This counterargument may consist in Ann explicitly believing that Bob explicitly believes she does not intend to use the bike to go to work, and Ann explicitly believing that if Bob explicitly believes this, he will not fix the tire.*

The paper is organized as follows. In Section 2, we discuss related work. In Section 3, we present the belief base semantics. Then, in Section 4, we introduce the language for explicit belief, plain implicit belief, and focused implicit belief. In Section 5, we use the language to formalize the concept of deductive argument. In Section 6, we provide a semantic analysis of the relationship between deductive argument and substantive reason. Section 7 is devoted to the proof-theoretic aspects and decidability of the language. In Section 8, we discuss some perspectives for future work.

2 Related Work

As pointed out in the introduction, two traditions in the integration of epistemic logic and formal argumentation can be identified: the representation of agents' knowledge and uncertainty about argumentation frameworks, and logical theories of the connection between beliefs and arguments. One of the first works in the first tradition is Schwarzentruber et al. (Schwarzentruber, Vesic, and Rienstra 2012). They extend possible worlds models of multi-agent epistemic logic with argumentation frameworks—associating one argumentation framework with each world—and interpret a variety of epistemic languages over this class of models. Sakama & Son (Sakama and Son 2019; Sakama and Son 2020) introduce the notion of an epistemic argumentation framework (EAF), which extends a basic argumentation framework with the reasoner's view, represented by an epistemic formula. Herzig & Yuste-Ginel (Herzig and Yuste-Ginel 2021a) formalize the notion of a higher-order view on an argumentation framework, that is, what an agent knows about what another agent knows about an argumentation framework. In another work (Herzig and Yuste-Ginel 2021b), they show that incomplete argumentation frameworks (IAFs), as defined in (Baumeister et al. 2018), can be naturally represented within the epistemic logic of visibility (Herzig, Lorini, and Maffre 2018).[1]

A notable work in the second tradition is Grossi & van der Hoek's (Grossi and van der Hoek 2014) on a two-dimensional semantics for argument-based beliefs. They introduce a new class of structures, called doxo-argumentative structures, defined as the modal product of doxastic structures and argumentative structures, and use it to define a variety of notions of argument-based belief. Shi et al. (Shi, Smets, and Velázquez-Quesada 2017; Shi, Smets, and Velázquez-Quesada 2023) extend topological models for epistemic logic (Baltag et al. 2016) with an argumentation component, namely, an attack relation between the elements of the topological space, capturing the evidences that support beliefs. They use the resulting structures, called topological argumentation models, to define argument-based notions of belief, including grounded beliefs and fully grounded beliefs. For instance, grounded belief is the belief supported by at least one argument in the grounded extension computed from the topological argumentation model. Finally, the work by Wáng & Li (Wáng and Li 2021) is in line with the so-called justified true belief (JTB) theory of knowledge. They extend relational doxastic models of epistemic logic with an argumentation function that associates a set of arguments to each world in the model, where an argument is seen as a set of sets of worlds. They use their semantics to define a modal operator for "true belief supported by an argument". The last work to be mentioned is Amgoud & Demolombe's (Amgoud and Demolombe 2014). They combine epistemic logic and abstract argumentation to formalize how different arguments interact in assessing trustworthiness of information sources.

Our work falls within the second tradition by proposing an integration of epistemic logic and the theory of deductive arguments, one of the main theories in formal argumentation. As far as we know, such an integration has not been proposed before.

3 Belief Base Semantics

Following (Lorini 2020), in this section we present a semantics for epistemic logic exploiting the notion of belief base. Unlike the standard Kripke semantics for epistemic logic in which possible states and epistemic alternatives are primitive, in the belief base semantics they are defined from the primitive concept of belief base.

Assume a countably infinite set of atomic propositions $Atm = \{p, q, \ldots\}$ and a finite set of agents $Agt =$

[1] The relation between IAFs and the possible world semantics of epistemic logic is also explored in (Proietti and Yuste-Ginel 2021).

$\{1,\ldots,n\}$. We define the language $\mathcal{L}_0$ for talking agents' explicit beliefs by the following grammar:

$$\mathcal{L}_0 \stackrel{\texttt{def}}{=} \quad \alpha \quad ::= \quad p \mid \neg\alpha \mid \alpha\wedge\alpha \mid \triangle_i\alpha,$$

where p ranges over Atm and i ranges over Agt. The formula $\triangle_i\alpha$ is read "agent i has the explicit belief that α".

The following notion of state is needed to provide a semantic interpretation of the formulas in the language $\mathcal{L}_0$. A state has two components: a propositional valuation representing the atomic facts that are true (resp. false) in the environment and one belief base for every agent in Agt made of formulas from $\mathcal{L}_0$. Thus, an agent's belief base can contain not only propositional information (i.e., information about the environment) but also higher-order information regarding the explicit beliefs of other agents.

Definition 1 (State)**.** *A state is a tuple $S = \big((B_i)_{i\in Agt}, V\big)$ where:*

- *B_i is a finite set of formulas from $\mathcal{L}_0$ representing agent i's belief base,*
- *V is a set of atomic propositions from Atm representing the actual environment.*

The set of all states is denoted by $\mathbf{S}$.

The following definition specifies the truth conditions for the formulas in the base language $\mathcal{L}_0$ relative to a state.

Definition 2 (Satisfaction relation)**.** *Let $S = \big((B_i)_{i\in Agt}, V\big)\in\mathbf{S}$. Then,*

$$\begin{aligned}
S\models p &\iff p\in V,\\
S\models\neg\alpha &\iff S\not\models\alpha,\\
S\models\alpha_1\wedge\alpha_2 &\iff S\models\alpha_1 \text{ and } S\models\alpha_2,\\
S\models\triangle_i\alpha &\iff \alpha\in B_i.
\end{aligned}$$

According to the previous semantic interpretation, p is actually true (i.e., $S\models p$) if p is a property of the actual environment (i.e., $p\in V$). Pay special attention to the interpretation of the explicit belief operators in the previous definition: agent i has the explicit belief that α (i.e., $S\models\triangle_i\alpha$) if α is included in its belief base (i.e., $\alpha\in B_i$). The following definition introduces the notion of doxastic alternative.

Definition 3 (Doxastic alternatives)**.** *Let $i\in Agt$. Then, $\mathcal{R}_i$ is the binary relation on the set $\mathbf{S}$ such that, for all $S = \big((B_i)_{i\in Agt}, V\big), S' = \big((B'_i)_{i\in Agt}, V'\big)\in\mathbf{S}$:*

$$S\mathcal{R}_iS' \text{ if and only if } \forall\alpha\in B_i, S'\models\alpha.$$

$S\mathcal{R}_iS'$ means that S' is a doxastic alternative for agent i at S, that is to say, S' is a state that at S agent i considers possible. The idea of the previous definition is that S' is a doxastic alternative for agent i at S if and only if, S' satisfies all facts that agent i explicitly believes at S.

Let us introduce the the notion of focus-based doxastic alternative into the following definition: an alternative that an agent considers possible when focusing its attention on a restricted set of formulas from $\mathcal{L}_0$.

Definition 4 (Focus-based doxastic alternatives)**.** *Let $i\in Agt$ and $X\subseteq\mathcal{L}_0$ Then, $\mathcal{R}_i^X$ is the binary relation on the set $\mathbf{S}$ such that, for all $S = \big((B_i)_{i\in Agt}, V\big), S' = \big((B'_i)_{i\in Agt}, V'\big)\in\mathbf{S}$:*

$$S\mathcal{R}_i^XS' \text{ if and only if } \forall\alpha\in\big(B_i\cap X\big), S'\models\alpha.$$

$S\mathcal{R}_i^XS'$ means that S' is a doxastic alternative for agent i at S under focus X, that is to say, S' is a state that at S i considers possible, when focusing on the formulas in X. According to Definition 4, the latter means that S' satisfies all facts in the restriction of i's belief base at S to X.

The last notion we consider is that of model. A model is a state supplemented with a set of states, called *context* or *universe*, that captures the agents' common ground, namely, the body of information of which the agents have common knowledge (Stalnaker 2002). The context is not necessarily equal to the set of all states $\mathbf{S}$, since there could be states in $\mathbf{S}$ which are incompatible with the agents' common knowledge of the "laws of the domain". For example, we might want to exclude from the context all states in which the propositions "Ann is in Paris" and "Ann is in Rome" are true at the same time, under the assumption that the agents have common knowledge of the basic principles of physics, according to which a person or object cannot be in two different places simultaneously.

Definition 5 (Model)**.** *A multi-agent belief base model (or simply model) is a pair (S,U) where $S\in U\subseteq\mathbf{S}$. The class of models is denoted by* $\mathbf{M}$.

Definition 5 requires $S\in U$ since we conceive the agents' common ground as their common knowledge and not as their common belief. By definition, the agents' common knowledge is correct (i.e., the actual state is included in it). Note that the model $(S,\mathbf{S})$ is a model with *maximal uncertainty* since it includes all possible states. This means that in this model the agents have no shared information in their common ground. For notational convenience, we write $S\models\varphi$ instead of $(S,\mathbf{S})\models\varphi$. Let us go back to the Example 1 given in Section 1 to illustrate the semantics.

Example 2. *We consider the model (S,U) where the state $S = \big((B_{Ann}, B_{Bob}), V\big)$ represents the explicit beliefs of Ann and Bob as well as the actual environment:*

$$\begin{aligned}
B_{Ann} =& \{\mathit{fl}, \triangle_{Bob}\mathit{fl}, \triangle_{Bob}\mathit{fl}\rightarrow\mathit{fi},\\
&\triangle_{Bob}\neg ab, \triangle_{Bob}\neg ab\rightarrow\neg\mathit{fi}\},\\
B_{Bob} =& \{\mathit{fl}, \mathit{fi}, \triangle_{Ann}\triangle_{Bob}\mathit{fl}, \triangle_{Ann}(\triangle_{Bob}\mathit{fl}\rightarrow\mathit{fi})\},\\
V =& \{\mathit{fl}, \mathit{fi}, ab\}.
\end{aligned}$$

The atomic propositions fl, fi, and ab denote, respectively, "the bike tire is flat", "Bob fixes the bike tire", and "Ann wants to use the bike to go to work". Thus, Ann explicitly believes i) that the bike tire is flat (i.e., $\mathit{fl}\in B_{Ann}$); ii) that Bob explicitly believes that the bike tire is flat (i.e., $\triangle_{Bob}\mathit{fl}\in B_{Ann}$); iii) that if Bob explicitly believes that the bike tire is flat then he will fix it (i.e., $\triangle_{Bob}\mathit{fl}\rightarrow\mathit{fi}\in B_{Ann}$); iv) that Bob explicitly believes that Ann does not wish to use the bike to go to work (i.e., $\triangle_{Bob}\neg ab\in B_{Ann}$); and v) that if Bob explicitly believes that Ann does not wish to use the bike to go to work, then he will not fix the bike tire (i.e.,

$\triangle_{Bob}\neg ab \rightarrow \neg fi \in B_{Ann}$). Moreover, Bob explicitly believes i) that the bike tire is flat (i.e., $fl \in B_{Bob}$), ii) that he is going to fix it (i.e., $fi \in B_{Bob}$), iii) that Ann explicitly believes that Bob explicitly believes that the bike tire is flat (i.e., $\triangle_{Ann}\triangle_{Bob}fl \in B_{Bob}$); iv) that Ann explicitly believes that if Bob explicitly believes that the bike tire is flat, then he will fix it (i.e., $\triangle_{Ann}(\triangle_{Bob}fl \rightarrow fi) \in B_{Bob}$). Finally, the bike tire is flat, Bob is going to fix it, and Ann wishes to use the bike to go to work (i.e., $fl, fi, ab \in V$).

The context U consists of all states in which Bob will fix the bike tire only if it is flat. Indeed, by definition, an object is 'fixed' only when it is damaged, broken, or not functioning properly, and this is common knowledge between Ann and Bob. That is, $U = \{S' \in \mathbf{S} : S' \models fi \rightarrow fl\}$.

4 Epistemic Language

We extend the language $\mathcal{L}_0$ of explicit beliefs defined in the previous section with modal operators for plain implicit belief and focused implicit belief. We call $\mathcal{L}$ the resulting language and define it by the following grammar:

$$\mathcal{L} \stackrel{\text{def}}{=} \quad \varphi \quad ::= \quad \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid \Box_i^X\varphi,$$

where α ranges over $\mathcal{L}_0$, i ranges over Agt and X is a finite subset of $\mathcal{L}_0$ which we note $X \subseteq^{\mathsf{fin}} \mathcal{L}_0$.

The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from α, $\neg$ and $\wedge$ in the standard way. The formula $\Box_i\varphi$ has to be read "agent i implicitly believes that φ" or "agent i can deduce that φ from their explicit beliefs", while the formula $\Box_i^X\varphi$ has to be read "agent i implicitly believes that φ, when focusing on the formulas in X" or "agent i can deduce that φ from their explicit beliefs, when focusing on the formulas in X".

The duals of the modal operators $\Box_i$ and $\Box_i^X$ are defined in the usual way: $\Diamond_i\varphi \stackrel{\text{def}}{=} \neg\Box_i\neg\varphi$ and $\Diamond_i^X\varphi \stackrel{\text{def}}{=} \neg\Box_i^X\neg\varphi$.

The following definition extends the definition of the satisfaction relation (Definition 2) to the full language $\mathcal{L}$. In particular, formulas of the language $\mathcal{L}$ are interpreted with respect to a model (S,U), as defined in Definition 5. (Boolean cases are omitted since they are defined as usual.)

Definition 6 (Satisfaction relation (cont.)). *Let $(S,U) \in \mathbf{M}$. Then:*

$$\begin{aligned}
(S,U) \models \alpha &\iff S \models \alpha,\\
(S,U) \models \Box_i\varphi &\iff \forall S' \in U, \text{ if } S\mathcal{R}_iS' \text{ then } (S',U) \models \varphi,\\
(S,U) \models \Box_i^X\varphi &\iff \forall S' \in U, \text{ if } S\mathcal{R}_i^XS' \text{ then } (S',U) \models \varphi.
\end{aligned}$$

According to the previous definition, agent i implicitly believes that φ if φ is true at all states in the actual context that i considers possible. Moreover, agent i implicitly believes that φ when focusing on X if φ is true at all states in the actual context that i considers possible when focusing on X.

Let $\varphi \in \mathcal{L}$. We say that φ is valid for the class $\mathbf{M}$, denoted by $\models \varphi$, if for every $(S,U) \in \mathbf{M}$ we have $(S,U) \models \varphi$. We say that φ is satisfiable for the class $\mathbf{M}$ if $\neg\varphi$ is not valid for the class $\mathbf{M}$.

We conclude this section with an expressiveness result for the language $\mathcal{L}$: plain implicit belief and focused implicit belief are not reducible to each other.

Theorem 1. *The operators $\Box_i\varphi$ and $\Box_i^X\varphi$ are not interdefinable.*

Proof. We first prove that $\Box_i^X\varphi$ is not expressible in the $\Box_i$-fragment.

Toward a contradiction, we suppose that there is a formula φ of the $\Box_i$-fragment which is equivalent to $\Box_i^{\{q\}}q$. We consider a proposition p which does not occur in φ. We have:

$$\begin{aligned}
i)\ &\big((\{p, p \rightarrow q, q\}, \emptyset)\big) \models \Box_i^{\{q\}}q,\\
ii)\ &\big((\{p, p \rightarrow q\}, \emptyset)\big) \not\models \Box_i^{\{q\}}q,\\
iii)\ &\big((\{p, p \rightarrow q, q\}, \emptyset)\big) \models \varphi,\\
iv)\ &\big((\{p, p \rightarrow q\}, \emptyset)\big) \not\models \varphi.
\end{aligned}$$

By induction on the structure of φ, it is routine to show that if p does not occur in φ then:

$$\big((\{p, p \rightarrow q, q\}, \emptyset)\big) \models \varphi \text{ iff } \big((\{p, p \rightarrow q\}, \emptyset)\big) \models \varphi.$$

Thus, iii) implies $\big((\{p, p \rightarrow q\}, \emptyset)\big) \models \varphi$ which contradicts iv).

Now we are going to prove that $\Box_i\varphi$ is not expressible in the $\Box_i^X$-fragment. Toward a contradiction, we suppose that there is a formula φ of the $\Box_i^X$-fragment which is equivalent to $\Box_iq$. We consider a proposition p which does not occur in φ. We have:

$$\begin{aligned}
i)\ &\big((\{p, p \rightarrow q\}, \emptyset)\big) \models \Box_iq,\\
ii)\ &\big((\{p \rightarrow q\}, \emptyset)\big) \not\models \Box_iq,\\
iii)\ &\big((\{p, p \rightarrow q\}, \emptyset)\big) \models \varphi,\\
iv)\ &\big((\{p \rightarrow q\}, \emptyset)\big) \not\models \varphi.
\end{aligned}$$

By induction on the structure of φ, it is routine to show that if p does not occur in φ then:

$$\big((\{p, p \rightarrow q\}, \emptyset)\big) \models \varphi \text{ iff } \big((\{p \rightarrow q\}, \emptyset)\big) \models \varphi.$$

Thus, iii) implies $\big((\{p \rightarrow q\}, \emptyset)\big) \models \varphi$. It contradicts iv). $\Box$

5 The Concept of Deductive Argument

Having introduced the belief base semantics and the epistemic language in the previous two sections, we can now define the notion of a deductive argument, according to Besnard and Hunter's deductive theory (Besnard and Hunter 2001; Besnard and Hunter 2008). We define it as an abbreviation within the epistemic language. Conceptually, it can be viewed as a ternary predicate involving three components: an agent $i \in Agt$, the holder of the argument (i.e., the agent who forms the argument); a non-empty, finite set of formulas $X \subseteq^{\mathsf{fin}} \mathcal{L}_0$ representing the support or content of the argument; and a formula $\varphi \in \mathcal{L}$ representing the supported conclusion:

$$\mathsf{Arg}_i(X,\varphi) \stackrel{\text{def}}{=} \bigwedge_{\alpha \in X} \triangle_i\alpha \wedge \Box_i^X\varphi \wedge \neg\Box_i^X\bot \wedge \bigwedge_{X' \subset X} \neg\Box_i^{X'}\varphi.$$

The abbreviation $\mathsf{Arg}_i(X,\varphi)$ stands for "the information in X is, for agent i, an argument supporting the conclusion that φ." This means that: i) every fact in X is explicitly believed to be true by agent i (i.e., $\bigwedge_{\alpha \in X} \triangle_i \alpha$); ii) agent i implicitly believes that φ when focusing on the information in X (i.e., $\Box_i^X \varphi$); iii) agent i does not believe a contradiction when focusing on X (i.e., $\neg\Box_i^X \bot$); and iv) agent i can no longer implicitly believe that φ when neglecting some information in X (i.e., $\bigwedge_{X' \subset X : |X'| = |X|-1} \neg\Box_i^{X'} \varphi$).

Let us go back to Ann and Bob's example to illustrate the previous definition of deductive argument. The example shows that our approach generalizes B&H's theory to arguments whose premises may contain information about agents' explicit beliefs and whose conclusions may be about agents' implicit beliefs. Indeed, in B&H's theory, both premises and conclusion of an argument are propositional.

Example 3. *It is straightforward to verify that the following holds, where (S,U) is the model given in Example 2:*

$$(S,U) \models \mathsf{Arg}_{Ann}(\{\triangle_{Bob} \mathit{fl}, \triangle_{Bob} \mathit{fl} \to \mathit{fi}\},\mathit{fi}) \wedge$$
$$\mathsf{Arg}_{Ann}(\{\triangle_{Bob} \neg ab, \triangle_{Bob} \neg ab \to \neg \mathit{fi}\},\neg \mathit{fi}) \wedge$$
$$\mathsf{Arg}_{Bob}(\{\triangle_{Ann} \triangle_{Bob} \mathit{fl}, \triangle_{Ann}(\triangle_{Bob} \mathit{fl} \to \mathit{fi})\},\Box_{Ann} \mathit{fi}).$$

This means that at the model (S,U) Ann holds both an argument supporting the conclusion that Bob is going to fix the bike tire and a counter-argument supporting the conclusion that Bob is not going to fix the bike tire. Moreover, Bob holds an argument supporting the conclusion that Ann implicitly believes that Bob is going to fix the bike tire.

We are now going to show the correspondence between the previous definition of deductive argument and that of B&H. To illustrate this correspondence, some preliminaries are necessary. First, we treat the language $\mathcal{L}_0$ as a propositional language built from the set of atomic formulas $Atm^+ = Atm \cup \{\triangle_i \alpha : i \in Agt \text{ and } \alpha \in \mathcal{L}_0\}$. Second, we consider the classical deductive closure operator Cn on sets of propositional-like formulas from $\mathcal{L}_0$, where explicit belief formulas "$\triangle_i \alpha$" are regarded as atomic formulas.

As the following proposition highlights, in a model with maximal uncertainty and when the supported formula belongs to $\mathcal{L}_0$, the previous notion of deductive argument has a deductive characterization that corresponds to B&H's definition (Besnard and Hunter 2001, Definition 3.1).

Proposition 1. *Let $\alpha \in \mathcal{L}_0$, $S = \big((B_i)_{i \in Agt}, V\big) \in \mathbf{S}$ and $X \subseteq^{\mathsf{fin}} \mathcal{L}_0$. Then,*

$$\begin{aligned} S \models \mathsf{Arg}_i(X,\alpha) \textit{ iff } & i)\ X \subseteq B_i, \\ & ii)\ \alpha \in Cn(B_i \cap X), \\ & iii)\ \bot \notin Cn(B_i \cap X), \\ & iv)\ \forall X' \subset X, \alpha \notin Cn(B_i \cap X'). \end{aligned}$$

Proof. The proposition is a direct consequence of the following property:

$$i)\ S \models \Box_i^X \alpha \text{ iff } \alpha \in Cn(B_i \cap X),$$

together with the semantic interpretation of the formulas of the language $\mathcal{L}$ (Definition 6) and the definition of $\mathsf{Arg}_i(X,\alpha)$. We are goint to prove property i). Let Ω be the set of propositional interpretations for the language $\mathcal{L}_0$, when formulas "$\triangle_i \alpha$" are regarded as atomic formulas It is straightforward to verify that there exists a bijection $f : \Omega \longrightarrow \mathbf{S}$ such that $f(\omega), \mathbf{S} \models \alpha$ iff $\omega \models \alpha$, for all $\omega \in \Omega$ and $\alpha \in \mathcal{L}_0$.

We are going to prove the equivalent statement "$S \models \Diamond_i^X \neg\alpha$ iff $\alpha \notin Cn(B_i \cap X)$".

Suppose that $\alpha \notin Cn(B_i \cap X)$. This means that $(B_i \cap X) \cup \{\neg\alpha\}$ is propositionally consistent. Hence, by the strong completeness of propositional logic, there exists $\omega \in \Omega$ such that $\omega \models \beta$ for all $\beta \in (B_i \cap X) \cup \{\neg\alpha\}$. Thus, there exists $S' \in \mathbf{S}$ such that $f(\omega) = S'$ and $S' \models \beta$ for all $\beta \in (B_i \cap X) \cup \{\neg\alpha\}$. Consequently, by the semantic interpretation of the modality $\Diamond_i^X$, $(S,\mathbf{S}) \models \Diamond_i^X \neg\alpha$. The other direction is proved in an analogous way. □

According to the previous proposition, the information in X is, for agent i, an argument supporting the conclusion that α if i) all information in X is part of i's belief base, ii) i can deduce α from the information in X, iii) the information in X is not inconsistent, iv) i can no longer deduce α when neglecting some information in X. In this sense, X is a minimal set of consistent information in i's belief base that enables i to deduce α.

We conclude this section by listing some notable properties of the previous notion of argument, presented in the form of validities. Proving them is routine, using the semantic interpretation of formulas given in Definition 6 and the definition of $\mathsf{Arg}_i(X,\varphi)$, so we omit their proofs:

$$\models \neg\mathsf{Arg}_i(X,\top) \tag{1}$$
$$\models \neg\mathsf{Arg}_i(X,\bot) \tag{2}$$
$$\models \big(\mathsf{Arg}_i(X,\varphi) \wedge \mathsf{Arg}_i(X,\psi)\big) \to \mathsf{Arg}_i(X,\varphi \wedge \psi) \tag{3}$$
$$\models \mathsf{Arg}_i(X,\varphi) \to \mathsf{Arg}_i(X,\varphi \vee \psi) \tag{4}$$

According to the validity (1), an agent cannot have an argument in support of a tautology, while according to the validity (2), it cannot have an argument in support of a contradiction. According to the validity (3), arguments aggregate under conjunction: if X is an argument in support of φ and X is an argument in support of ψ, then X is an argument in support of $\varphi \wedge \psi$. Validity (4) is a form of weakening: if X is an argument in support of φ then it is also an argument in support of $\varphi \vee \psi$. Note that weakening for conjunction, i.e., $\mathsf{Arg}_i(X,\varphi \wedge \psi) \to \mathsf{Arg}_i(X,\varphi)$, does not hold due to the negative condition $\bigwedge_{X' \subset X} \neg\Box_i^{X'} \varphi$ in the definition of deductive argument. Moreover, we have:

$$\models \mathsf{Arg}_i(X,\varphi) \to \Box_i \varphi \tag{5}$$
$$\models \big(\mathsf{Arg}_i(X,\varphi) \wedge \Diamond_i \top\big) \to \neg\mathsf{Arg}_i(X',\neg\varphi) \tag{6}$$

According to the validity (5), if an agent has an argument supporting φ, then it implicitly believes that φ. Notice, however, that the agent might have an inconsistent belief base that leads it to believe everything. Indeed, the formula $\mathsf{Arg}_i(X,\varphi) \wedge \Box_i \bot$ is satisfiable. According to the validity (6), if an agent has an argument supporting φ and its belief base is consistent, then it cannot have another argument supporting $\neg\varphi$.

6 Argument and Substantive Reason

In this section, we are going to show that the notion of deductive argument defined in the previous section is intimately related to a notion of substantive reason that can be expressed using the following notion of plausibility ordering over states induced by a belief base.

Definition 7 (Plausibility ordering). *Let $i \in Agt$ and $S, S', S'' \in \mathbf{S}$. Then,*

$$S' \preceq_{i,S} S'' \text{ iff } Sat_i(S, S') \subseteq Sat_i(S, S''),$$

where for all $S = \big((B_i)_{i \in Agt}, V\big), S' = \big((B'_i)_{i \in Agt}, V'\big) \in \mathbf{S}$, $Sat_i(S, S') = \{\alpha \in B_i : S' \models \alpha\}$.

According to the previous definition, at state S agent i considers state S'' at least as plausible as state S' if all explicit beliefs of agent i at state S that are satisfied by state S' are also satisfied by state S''.

It is easy to verify that $\preceq_{i,S}$ is a partial preorder over the set of states $\mathbf{S}$. From this partial preorder we can define the corresponding strict ordering $\prec_{i,S}$, indifference relation $\sim_{i,S}$ and incomparability relation $\|_{i,S}$:

$$S' \prec_{i,S} S'' \text{ iff } S' \preceq_{i,S} S'' \text{ and } S'' \not\preceq_{i,S} S';$$
$$S' \sim_{i,S} S'' \text{ iff } S' \preceq_{i,S} S'' \text{ and } S'' \preceq_{i,S} S';$$
$$S' \|_{i,S} S'' \text{ iff } S' \not\preceq_{i,S} S'' \text{ and } S'' \not\preceq_{i,S} S'.$$

Some preliminary notions are needed to formally define the notion of substantive reason. First of all, we need to introduce the following notion of epistemic core.

Definition 8 (Epistemic core). *Let $i \in Agt$ and $(S, U) \in \mathbf{M}$. Then, agent i's epistemic core at model (S, U) is:*

$$Core(i, S, U) = \big\{S' \in U : \forall S'' \in U, S'' \preceq_{i,S} S' \text{ or } S'' \|_{i,S} S'\big\}.$$

$Core(i, S, U)$, includes all states in context U that, according to agent i, have no strictly more plausible state in U. The following proposition is a consequence of the fact that, according to Definition 1, belief bases are finite.

Proposition 2. *Let $(S, U) \in \mathbf{M}$. Then, $Core(i, S, U) \neq \emptyset$.*

Proof. Towards a contradiction suppose $Core(i, S, U) = \emptyset$. This means that there is no $S' \in U$ such that $\forall S'' \in U, S'' \preceq_{i,S} S'$ or $S'' \|_{i,S} S'$. The latter is equivalent to the fact that there is no $S' \in U$ such that $\forall S'' \in U, Sat_i(S, S') \not\subset Sat_i(S, S'')$. The latter is equivalent to the fact that for all $S' \in U$, $\exists S'' \in U$ such that $Sat_i(S, S') \subset Sat_i(S, S'')$. The latter is equivalent to the fact that for all $S' \in U, \exists S'' \in U$ such that $\{\alpha \in B_i : S' \models \alpha\} \subset \{\alpha \in B_i : S'' \models \alpha\}$. The latter contradicts the fact that, by Definition 1, B_i should be finite. □

From the notion of epistemic core, we define the notion of epistemic support set. We use the term 'epistemic support set' instead of 'evidence set', as used in (van Benthem and Pacuit 2011), because it is broader and better aligned with our semantics. While 'evidence set' typically implies empirical or observational data, 'epistemic support set' encompasses not only explicit beliefs based on observable evidence but also explicit beliefs formed through inference.[2]

Definition 9 (Epistemic support set). *Let $i \in Agt$ and $(S, U) \in \mathbf{M}$. Then, agent i's epistemic support set at state S relative to the context U, denoted by $Supp(i, S, U)$, is the partition of $Core(i, S, U)$ induced by the indifference relation $\sim_{i,S}$. Elements of $Supp(i, S, U)$ are denoted by $\Theta, \Theta', \ldots$*

Elements of $Supp(i, S, U)$ are agent i's epistemic supports at state S, relative to context U.

The following two propositions are direct consequences of Proposition 2 and the fact that $Supp(i, S, U)$ is a partition of $Core(i, S, U)$.

Proposition 3. *Let $(S, U) \in \mathbf{M}$. Then, $Supp(i, S, U) \neq \emptyset$.*

Proposition 4. *Let $(S, U) \in \mathbf{M}$ and $\Theta \in Supp(i, S, U)$. Then, $\Theta \neq \emptyset$.*

We are going to show next that an agent's epistemic supports have a syntactic counterpart in the form of maximally consistent subsets (MCS) of the agent's belief base, which are defined next.

Definition 10 (MCS). *Let $(S, U) \in \mathbf{M}$. We define $MCS(i, S, U) \subseteq 2^{\mathcal{L}_0}$ such that, for every $X \subseteq \mathcal{L}_0$, $X \in MCS(i, S, U)$ if and only if*

$$i)\ X \subseteq B_i,$$
$$ii)\ \mathcal{R}_i^X(S) \cap U \neq \emptyset,$$
$$iii)\ \forall X' \subseteq B_i, \text{ if } X \subset X' \text{ then } \mathcal{R}_i^{X'}(S) \cap U = \emptyset.$$

$MCS(i, S, U)$ is agent i's set of maximally consistent subsets of its belief base at state S, relative to context U. As the following theorem highlights, Θ is an epistemic support of agent i if and only if there exists an MCS X of i's belief base such that the set of states satisfying all formulas in X coincides with Θ.

Theorem 2. *Let $(S, U) \in \mathbf{M}$. Then,*

$$\Theta \in Supp(i, S, U) \text{ iff } \exists X \in MCS(i, S, U) \text{ such that } \mathcal{R}_i^X(S) \cap U = \Theta.$$

Proof. ($\Rightarrow$) Suppose $\Theta \in Supp(i, S, U)$. By Proposition 4, we have $\Theta \neq \emptyset$. Thanks to $\Theta \in Supp(i, S, U)$, we have that $\exists X \subseteq B_i$ such that $\Theta = \{S' \in U : Sat_i(S, S') = X\}$ and $\forall X' \subseteq B_i$, if $X \subset X'$ then $\{S' \in U : Sat_i(S, S') = X'\} = \emptyset$. The latter together with the fact that $\Theta \neq \emptyset$ imply that $\exists X \subseteq B_i$ such that $\mathcal{R}_i^X(S) \cap U = \Theta, \mathcal{R}_i^X(S) \cap U \neq \emptyset$ and $\forall X' \subseteq B_i$, if $X \subset X'$ then $\mathcal{R}_i^{X'}(S) \cap U = \emptyset$. The latter means that $\exists X \in MCS(i, S, U)$ such that $\mathcal{R}_i^X(S) \cap U = \Theta$.

($\Leftarrow$) Suppose $\exists X \in MCS(i, S, U)$ such that $\mathcal{R}_i^X(S) \cap U = \Theta$. The latter implies that $\exists X \subseteq B_i$ such that $\mathcal{R}_i^X(S) \cap U = \Theta$ and $\forall X' \subseteq B_i$, if $X \subset X'$ then $\mathcal{R}_i^{X'}(S) \cap U = \emptyset$. The latter implies that $\Theta \in Supp(i, S, U)$. □

[2]The representation of inference-based explicit belief within the belief base semantics is discussed in (Lorini 2020, Section 7.2).

The previous Theorem 2 highlights the connection between the notion of epistemic support and the notion of *maximal body of evidence*, as defined in van Benthem & Pacuit's semantics for evidence-based beliefs (van Benthem and Pacuit 2011), namely, a maximal collection of evidence pieces that have the finite intersection property (see also (Özgün 2017, Chapter 5) for a discussion of this notion). Nonetheless, there is a difference between our semantics and that of van Benthem & Pacuit that should be noted: while they model evidence through an extensional neighbourhood semantics, we treat evidence as explicit beliefs and model them as syntactic entities in an agent's belief base. However, it is straightforward to see that our belief base model induces an evidence model in Benthem & Pacuit's sense. A further difference is that we assume the explicit beliefs in an agent's belief base to be finite, whereas in their semantics an agent's set of evidence can be infinite. For this reason, they consider the 'finite' intersection property, which corresponds to condition ii) in Definition 10.

We have now all the necessary elements to define the notion of substantive reason and to establish its formal connection with the notion of deductive argument defined in Section 5. The definition of MCS and Theorem 2 will turn out to be crucial for this latter point.

We say that agent i has a substantive reason to believe φ at a model (S, U), abbreviated by $\mathrm{SReas}(i{,}\varphi{,}S{,}U)$, if and only if i) there exists a support Θ in the epistemic support set $Supp(i, S, U)$ such every state in Θ satisfies φ, and ii) φ is not universally true. The condition i) corresponds to the concept of reason, while condition ii) indicates that for a reason to be substantive it has to be meaningful and contribute real weight. Indeed, if condition ii) were not met, the agent would have no way of imagining the possibility that φ might be false. In that case, the reason supporting φ would not be substantive, as it would lack epistemic value. Note that, in virtue of Theorem 2, condition i) is analogous to the notion of a maximal body of evidence supporting a proposition in van Benthem & Pacuit's theory.

Definition 11 (Substantive reason)**.** *Let $(S, U) \in \mathbf{M}$. We say that agent i has a substantive (or non-trivial) reason to believe that φ at model (S, U), denoted by $\mathrm{SReas}(i{,}\varphi{,}S{,}U)$, if and only if*

i) $\exists \Theta \in Supp(i, S, U)$ *such that* $\forall S' \in \Theta, (S', U) \models \varphi,$ *and*

ii) $\exists S'' \in U$ *such that* $(S'', U) \models \neg\varphi$.

The following Theorem 3 is the central result of this section. It shows that the notion of substantive reason is reducible to the notion of argument. Specifically, an agent has a substantive reason to believe φ iff there exists an argument X for the agent that supports φ.

Theorem 3. *Let $(S, U) \in \mathbf{M}$. Then,*

$$\mathrm{SReas}(i{,}\varphi{,}S{,}U) \text{ iff } \exists \text{ non-empty } X \subseteq^{\mathsf{fin}} \mathcal{L}_0 \text{ such that } (S, U) \models \mathsf{Arg}_i(X{,}\varphi).$$

Proof. ($\Rightarrow$) We first prove the left-to-right direction of the theorem. We are going to prove two lemmas (Lemma 1 and Lemma 2) in succession, as they are useful for proving this direction of the proof.

Lemma 1. *Let $X \subseteq \mathcal{L}_0$ and $\Theta \subseteq \mathbf{S}$ such that $\Theta \neq \emptyset$. Then, if $\big(\Theta \cap \mathcal{R}_i^X(S)\big) = \emptyset$ then $\exists X' \subseteq X$ such that $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) = \emptyset$ and $\forall X'' \subset X'$, $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) \neq \emptyset$.*

Proof. We prove the lemma by induction on the size of X. It is easy to verify that $\mathcal{R}_i^{\emptyset}(S) = \mathbf{S}$. Thus, we only need to prove for $|X| > 0$. Suppose $\Theta \neq \emptyset$ and $\big(\Theta \cap \mathcal{R}_i^X(S)\big) = \emptyset$.

Base case: $|X| = 1$. By $\mathcal{R}_i^{\emptyset}(S) = \mathbf{S}$ and $\Theta \neq \emptyset$, we have $\big(\Theta \cap \mathcal{R}_i^{\emptyset}(S)\big) \neq \emptyset$. Thus, thanks to $|X| = 1$, $\Theta \neq \emptyset$ and $\big(\Theta \cap \mathcal{R}_i^{\emptyset}(S)\big) \neq \emptyset$, we have that $\exists X' \subseteq X$ (viz. $X' = X$) such that $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) = \emptyset$ and $\forall X'' \subset X'$, $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) \neq \emptyset$.

Inductive case: $|X| = k + 1$ with $k > 0$. We distinguish two cases.

Case 1: $\forall X' \subset X$, if $|X'| = |X| - 1$ then $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) \neq \emptyset$. It follows that $\forall X'' \subset X$, $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) \neq \emptyset$, since if $X'' \subseteq X'$ and $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) \neq \emptyset$ then $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) \neq \emptyset$. Thus, by $\big(\Theta \cap \mathcal{R}_i^{X}(S)\big) = \emptyset$, we have that $\exists X' \subseteq X$ (viz. $X' = X$) such that $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) = \emptyset$ and $\forall X'' \subset X'$, $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) \neq \emptyset$.

Case 2: $\exists X' \subset X$ such that $|X'| = |X| - 1$ and $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) = \emptyset$. By induction hypothesis, it follows that $\exists X' \subset X, \exists X'' \subseteq X'$ such that $|X'| = |X| - 1$, $\big(\Theta \cap \mathcal{R}_i^{X'}(S)\big) = \emptyset$, $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) = \emptyset$ and $\forall X''' \subset X''$, $\big(\Theta \cap \mathcal{R}_i^{X'''}(S)\big) \neq \emptyset$. Hence, $\exists X'' \subseteq X$ such that $\big(\Theta \cap \mathcal{R}_i^{X''}(S)\big) = \emptyset$ and $\forall X''' \subset X''$, $\big(\Theta \cap \mathcal{R}_i^{X'''}(S)\big) \neq \emptyset$. $\square$

Lemma 2. *Let $\Theta \in Supp(i, S, U)$ with $S = \big((B_i)_{i \in Agt}, V\big)$. Then, $\exists X \subseteq B_i$ such that $\mathcal{R}_i^X(S) \cap U = \Theta$.*

Proof. Suppose $\Theta \in Supp(i, S, U)$. Thus, $\exists X' \subseteq B_i$ such that $\Theta = \{S' \in U : Sat_i(S, S') = X'\}$. Hence, $\exists X \subseteq B_i$ such that $\mathcal{R}_i^X(S) \cap U = \Theta$. $\square$

Suppose $\mathrm{SReas}(i{,}\varphi{,}S{,}U)$ with $S = \big((B_i)_{i \in Agt}, V\big)$. Thus, i) $\exists \Theta \in Supp(i, S, U)$ such that $\forall S' \in \Theta, (S', U) \models \varphi$, and ii) $\exists S'' \in U$ such that $(S'', U) \models \neg\varphi$, that is, $||\neg\varphi||_U \neq \emptyset$ with $||\neg\varphi||_U = \{S' \in U : (S', U) \models \neg\varphi\}$.

By item i), Lemma 2 and Proposition 4, we have that iii) $\exists X \subseteq B_i$ such that $\mathcal{R}_i^X(S) \cap ||\neg\varphi||_U = \emptyset$ and $\mathcal{R}_i^X(S) \cap U \neq \emptyset$.

By item ii), item iii), Lemma 1 and the fact that if $X' \subseteq X$ and $\mathcal{R}_i^X(S) \cap U \neq \emptyset$ then $\mathcal{R}_i^{X'}(S) \cap U \neq \emptyset$, we have that iv) $\exists X' \subseteq B_i$ such that $||\neg\varphi||_U \cap \mathcal{R}_i^{X'}(S) = \emptyset$, $\mathcal{R}_i^{X'}(S) \cap U \neq \emptyset$ and $\forall X'' \subset X'$, $||\neg\varphi||_U \cap \mathcal{R}_i^{X''}(S) \neq \emptyset$.

By item iv) and the fact that, according to Definition 1, every belief base in a state S is finite, we have that $\exists X' \subseteq^{\mathsf{fin}} \mathcal{L}_0$ such that $(S, U) \models \bigwedge_{\alpha \in X'} \triangle_i \alpha \wedge \Box_i^{X'} \varphi \wedge \neg \Box_i^{X'} \bot \wedge \bigwedge_{X'' \subset X'} \neg \Box_i^{X''} \varphi$. The latter means that $\exists X' \subseteq^{\mathsf{fin}} \mathcal{L}_0$ such that $(S, U) \models \mathsf{Arg}_i(X'{,}\varphi)$.

($\Leftarrow$) Let us now prove the right-to-left direction of the theorem. The following fact is useful for our proof. It is straightforward to prove it from the definition of $MCS(i, S, U)$.

Fact 1. *If* $\big(\mathcal{R}_i^X(S) \cap U\big) \neq \emptyset$ *then* $\exists X' \in MCS(i,S,U)$ *such that* $X \subseteq X'$.

Suppose $(S,U) \models \mathsf{Arg}_i(X,\varphi)$ for some non-empty $X \subseteq^{\mathsf{fin}} \mathcal{L}_0$. The latter implies that i) $\exists X \subseteq^{\mathsf{fin}} \mathcal{L}_0$ such that $\big(\mathcal{R}_i^X(S) \cap U\big) \neq \emptyset$ and $\big(\mathcal{R}_i^X(S) \cap U\big) \subseteq ||\varphi||_U$, and ii) $\exists S'' \in U$ such that $(S'',U) \models \neg\varphi$ (since X is non-empty). By item i), Fact 1 and the fact that if $X \subseteq X'$ and $\big(\mathcal{R}_i^X(S) \cap U\big) \subseteq ||\varphi||_U$ then $\big(\mathcal{R}_i^{X'}(S) \cap U\big) \subseteq ||\varphi||_U$, the latter implies that iii) $\exists X' \in MCS(i,S,U)$ such that $\big(\mathcal{R}_i^{X'}(S) \cap U\big) \subseteq ||\varphi||_U$.

By item iii) and the right-to-left direction of Theorem 2, we have that iv) $\Theta \in Supp(i,S,U)$ such that $\Theta \subseteq ||\varphi||_U$. By item ii) and item iv), we have that $\mathrm{SReas}(i,\varphi,S,U)$. $\square$

In the next section, we will move from the conceptual analysis to the study of the proof-theoretic and decidability aspects of our framework.

7 Axiomatization and Decidability

In this section we focus on the axiomatization of the language $\mathcal{L}$ and prove decidability of the satisfiability checking problem. We first present an alternative semantics for the language $\mathcal{L}$, which serves as a technical device. Then, after having introduced the logic for the language $\mathcal{L}$, we state its soundness and completeness with respect to the model class $\mathbf{M}$.

7.1 Alternative Semantics

The alternative semantics for the language relies on Kripke structures with belief bases (KB) of the form $M = \big(W, \mathcal{B}, \mathcal{C}, \sim, \mathcal{V}\big)$ where:

- W is a set of worlds;
- $\mathcal{B} : Agt \times W \longrightarrow 2^{\mathcal{L}_0}$ is a belief base function such $\mathcal{B}(i,w)$ is finite for all $i \in Agt$ and for all $w \in W$;
- $\mathcal{C} : Agt \times W \longrightarrow 2^W$ is a base-compatibility function;
- $\sim$ is an equivalence relation;
- $\mathcal{V} : Atm \longrightarrow 2^W$ is a valuation function.

For every agent $i \in Agt$ and for every world $w \in W$, $\mathcal{B}(i,w)$ denotes agent i's set of explicit beliefs at world w, while $\mathcal{C}(i,w)$ denotes the set of worlds that are compatible with agent i's explicit beliefs at w. The equivalence relation $\sim$ capures context-equivalence: if $w \sim v$ then worlds w and v belong to the same context.

We use such structures to interpret formulas of the language $\mathcal{L}$. In particular, given $\varphi \in \mathcal{L}$, we interpret it relative to a pair (M,w) with M a KB and $w \in W$, as follows. (We omit the Boolean cases $\neg$ and $\wedge$, since they are defined as usual.)

$$
\begin{aligned}
(M,w) \models p &\iff w \in \mathcal{V}(p),\\
(M,w) \models \triangle_i \alpha &\iff \alpha \in \mathcal{B}(i,w),\\
(M,w) \models \Box_i \varphi &\iff \forall v \in \mathcal{C}(i,w), (M,v) \models \varphi,\\
(M,w) \models \Box_i^X \varphi &\iff \forall v \in \mathcal{F}(i,w,X), (M,v) \models \varphi,
\end{aligned}
$$

with

$$\mathcal{F}(i,w,X) = \bigcap_{\alpha \in \big(\mathcal{B}(i,w) \cap X\big)} ||\alpha||_{(M,w)}$$

and $||\alpha||_{(M,w)} = \big\{v \in W : w \sim v \text{ and } (M,v) \models \alpha\big\}$. The set $\mathcal{F}(i,w,X)$ denotes the set of worlds in the actual context that are compatible with agent i's explicit beliefs when the agent focuses on the information in X.

In a KB there is no connection between the function $\mathcal{B}$ and the function $\mathcal{C}$. The following definition introduces the notion of proper Kripke structure with belief bases (PKB) in which the connection is given.

Definition 12 (Proper Kripke structure with belief bases)**.** *A proper Kripke structure with belief bases (PKB) is a Kripke structure with belief bases* $M = \big(W, \mathcal{B}, \mathcal{C}, \sim, \mathcal{V}\big)$ *that satisfies the following condition, for all* $X \subseteq \mathcal{L}_0$*, for all* $i \in Agt$ *and for all* $w \in W$*:*

$$\textit{(C1)}\ \mathcal{C}(i,w) = \bigcap_{\alpha \in \mathcal{B}(i,w)} ||\alpha||_{(M,w)}.$$

The class of PKBs is denoted by $\mathbf{PKB}$.

Condition C1 in the previous definition simply states that, for a KB to be proper, the set of belief-compatible worlds for an agent must coincide with the set of worlds in the actual context that satisfy all the formulas in the agent's actual belief base. We are going to show that the epistemic language $\mathcal{L}$ is not expressive enough to capture the class of PKBs, but can only capture the strictly more general class in which condition C1 is weakened.

In particular, we define a *quasi* proper Kripke structure with belief bases (quasi-PKB) to be like a proper Kripke structure with belief bases, as in Definition 12, except that condition C1 is replaced by the following two weaker conditions, C1* and C1**:

$$
\begin{aligned}
&(\mathrm{C1}^*)\ \mathcal{C}(i,w) \subseteq \bigcap_{\alpha \in \mathcal{B}(i,w)} ||\alpha||_M,\\
&(\mathrm{C1}^{**})\ \mathcal{C}(i,w) \subseteq {\sim}(w),
\end{aligned}
$$

with ${\sim}(w) = \big\{v \in W : w \sim v\big\}$. The class of quasi-PKBs is denoted by $\mathbf{QPKB}$.

A PKB (resp. quasi-PKB) $M = \big(W, \mathcal{B}, \mathcal{C}, \sim, \mathcal{V}\big)$ is said to be *finite* if and only if W and $\mathcal{V}^{\leftarrow}(w) = \{p \in Atm : w \in \mathcal{V}(p)\}$ are finite sets. The class of finite PKBs (resp. finite quasi-PKBs) is denoted by finite-$\mathbf{PKB}$ (resp. finite-$\mathbf{QPKB}$). The following theorem highlights that the five semantics for the language $\mathcal{L}$, the four ones introduced in this section and belief base semantics relying on the model class $\mathbf{M}$ defined in Section 3, are all equivalent.

Theorem 4. *Let* $\varphi \in \mathcal{L}$*. Then, the following five statements are equivalent.*

1. φ *is satisfiable relative to class* $\mathbf{QPKB}$*,*
2. φ *is satisfiable relative to class finite-*$\mathbf{QPKB}$*,*
3. φ *is satisfiable relative to class finite-*$\mathbf{PKB}$*,*
4. φ *is satisfiable relative to class* $\mathbf{PKB}$*,*

5. φ is satisfiable relative to class $\mathbf{M}$.

SKETCH OF PROOF. We use a filtration-like method to prove that (1) implies (2). In particular, starting from a possibly infinite quasi-PKB, we can construct a finite quasi-PKB whose size is exponential in the size of φ. To show that (2) implies (3), we employ a technique that involves expanding an agent's belief base so that the set of worlds that are compatible with the agent's belief base is reduced to, and exactly matches, the set of worlds in which all formulas in the agent's belief base hold true, as required by Condition C1 in Definition 12. Finally, we prove that (4) and (5) are equivalent. The right-to-left direction is straightforward: from a multi-agent belief base model, we can easily construct the corresponding PKB. The left-to-right direction is less direct. A PKB can be redundant, meaning it may contain two distinct worlds with the same valuation of propositional atoms and identical belief bases for the agents. We therefore need to transform a possibly redundant PKB into a non-redundant one. From such a non-redundant PKB, we can construct the corresponding multi-agent belief base model that satisfies the same formulas. ■

The following complexity upper bound for satisfiability checking is a direct consequence of the previous theorem 4 whose proof shows that a formula φ is satisfiable for the class $\mathbf{M}$ iff it satisfied by a finite quasi-PKB of exponential size.

Theorem 5. *Checking satisfiability of formulas in $\mathcal{L}$ relative to the class $\mathbf{M}$ is in NExpTime.*

7.2 Logic

The following definition introduces the Logic of explicit, plain implicit and focused implicit Belief (LB) that we will show to be sound and complete for the model class $\mathbf{M}$.

Definition 13 (Logic LB)**.** *LB is the extension of classical propositional logic by the following axioms and inference rule:*

$$\big(\Box_i\varphi \wedge \Box_i(\varphi \rightarrow \psi)\big) \rightarrow \Box_i\psi \qquad (\mathbf{K}_{\Box_i})$$

$$\big(\Box_i^{\emptyset}\varphi \wedge \Box_i^{\emptyset}(\varphi \rightarrow \psi)\big) \rightarrow \Box_i^{\emptyset}\psi \qquad (\mathbf{K}_{\Box_i^{\emptyset}})$$

$$\Box_i^{\emptyset}\varphi \rightarrow \varphi \qquad (\mathbf{T}_{\Box_i^{\emptyset}})$$

$$\Box_i^{\emptyset}\varphi \rightarrow \Box_i^{\emptyset}\Box_i^{\emptyset}\varphi \qquad (\mathbf{4}_{\Box_i^{\emptyset}})$$

$$\neg\Box_i^{\emptyset}\varphi \rightarrow \Box_i^{\emptyset}\neg\Box_i^{\emptyset}\varphi \qquad (\mathbf{5}_{\Box_i^{\emptyset}})$$

$$\Box_i^{\emptyset}\varphi \leftrightarrow \Box_j^{\emptyset}\varphi \qquad (\mathbf{Equiv}_{\Box_i^{\emptyset},\Box_j^{\emptyset}})$$

$$\triangle_i\alpha \rightarrow \Box_i\alpha \qquad (\mathbf{Int}_{\triangle_i,\Box_i})$$

$$\Box_i^{\emptyset}\varphi \rightarrow \Box_i\varphi \qquad (\mathbf{Int}_{\Box_i^{\emptyset},\Box_i})$$

$$\Box_i^{X}\varphi \leftrightarrow \bigwedge_{X' \subseteq X}\Big(\mathsf{cnb}_{i,X,X'} \rightarrow \Box_i^{\emptyset}\big(\bigwedge_{\alpha \in X'}\alpha \rightarrow \varphi\big)\Big) \qquad (\mathbf{Red}_{\Box_i^{X}})$$

$$\frac{\varphi}{\Box_i^{\emptyset}\varphi} \qquad (\mathbf{Nec}_{\Box_i^{\emptyset}})$$

where for every $X, X' \subseteq^{\mathsf{fin}} \mathcal{L}_0$ such that $X' \subseteq X$:

$$\mathsf{cnb}_{i,X,X'} \stackrel{\mathtt{def}}{=} \bigwedge_{\alpha \in X'}\triangle_i\alpha \wedge \bigwedge_{\alpha \in X\setminus X'}\neg\triangle_i\alpha.$$

The plain implicit belief modality $\Box_i$ is a normal modal operator, satisfying the basic principles of modal logic K, including Axiom $\mathbf{K}_{\Box_i}$. The modality $\Box_i^{\emptyset}$ is an S5 modality, satisfying Axioms $\mathbf{K}_{\Box_i^{\emptyset}}$, $\mathbf{T}_{\Box_i^{\emptyset}}$, $\mathbf{4}_{\Box_i^{\emptyset}}$, and $\mathbf{5}_{\Box_i^{\emptyset}}$, as well as inference rule $\mathbf{Nec}_{\Box_i^{\emptyset}}$. In fact, it corresponds to a universal modality. Axiom $\mathbf{Int}_{\triangle_i,\Box_i}$ expresses the interaction between explicit belief and plain implicit belief: if something is explicitly believed, then it is also implicitly believed. Axiom $\mathbf{Int}_{\Box_i^{\emptyset},\Box_i}$ captures the interaction between the universal modality and plain implicit belief: if something is universally true, then it is implicitly believed. Axiom $\mathbf{Red}_{\Box_i^{X}}$ is a reduction axiom for focused implicit belief, reducing it to a formula involving only the universal modality. However, the formula on the right-hand side can be exponentially larger (in modal depth) than the one on the left. Finally, Axiom $\mathbf{Equiv}_{\Box_i^{\emptyset},\Box_j^{\emptyset}}$ expresses the agent-independence of the universal modality. It is also worth noting that the necessitation rule for plain implicit belief (i.e., $\frac{\varphi}{\Box_i\varphi}$) does not need to be included in the axiomatization, since it can be derived from inference rule $\mathbf{Nec}_{\Box_i^{\emptyset}}$ and Axiom $\mathbf{Int}_{\Box_i^{\emptyset},\Box_i}$.

The following is the central result about the proof-theoretic aspects of our framework.

Theorem 6. *The logic LB is sound and complete for the class* $\mathbf{M}$.

SKETCH OF PROOF. Soundness is straightforward. By combining a canonical model argument with the fact that every formula of the $\mathcal{L}$ can be tranformed in a provably equivalent formula containing only modalities of the form $\Box_i$ or $\Box_i^{\emptyset}$, we show that the logic LB is complete for the class $\mathbf{QPKB}$. Then, by Theorem 4, we conclude that LB is also complete for the class $\mathbf{M}$. ■

8 Perspectives

In this paper, we have laid out the conceptual and logical foundations for an epistemic notion of deductive argument. On the logical side, the novel contribution is the extension of the logic of explicit belief and plain implicit belief with focused implicit belief. This extension is necessary to represent deductive argument in the language. The notion of focused implicit belief has not been studied before in the epistemic logic literature. In this concluding section, we discuss two directions for future research.

Complexity analysis We have only provided a NExpTime upper bound for the satisfiability checking problem for the language $\mathcal{L}$ relative to the model class $\mathbf{M}$ (Theorem 5). Future work will be devoted to obtaining a tight complexity result. As for hardness, it is easy to show that there exists a poly-time reduction of the satisfiability checking problem for the multimodal logic K^n, extended with the universal modality, into our satisfiability checking problem. It is known that the former problem is EXPTIME-hard (Hemaspaandra 1996; Hemaspaandra 1993). Consequently, our problem is also EXPTIME-hard. We conjecture that our problem is in EXPTIME and thus EXPTIME-complete. Unfortunately, the reduction axiom $\mathbf{Red}_{\Box_i^X}$ of the logic LB cannot be used to obtain an efficient procedure, as it leads to an exponential blowup. Future work will be devoted to exploring alternative methods for proving this conjecture, including Hintikka set elimination.

Extension with non-normal modalities for reason In Section 6, we have provided a semantic analysis of the relation between deductive argument and substantive reason. However, the notion of substantive reason has only been defined at the semantic level. Future work will be devoted to extending our framework with a new family of non-normal modalities of the form $\langle [i] \rangle$, in order to express the notion of substantive reason in the language. The semantic interpretation of these modalities relative to a model is:

$$(S,U) \models \langle [i] \rangle \varphi \iff \exists \Theta \in Supp(i,S,U) \text{ such that } \forall S' \in \Theta, (S',U) \models \varphi,$$

where the formula $\langle [i] \rangle \varphi$ is read as "agent i has a reason to believe that φ". This modality, together with the universal modality of the language $\mathcal{L}$, allows us to express the notion of substantive reason as defined in Definition 11, namely: $\langle [i] \rangle \varphi \wedge \Diamond_i^{\emptyset} \varphi$. It is easy to check that the following formulas are valid with respect to the class $\mathbf{M}$:

$$\Big(\bigwedge_{\alpha \in X} \triangle_i \alpha \wedge \Diamond_i^X \top \Big) \rightarrow \langle [i] \rangle \Big(\bigwedge_{\alpha \in X} \alpha \Big) \tag{7}$$

$$\Diamond_i \top \rightarrow (\Box_i \varphi \leftrightarrow \langle [i] \rangle \varphi) \tag{8}$$

$$\langle [i] \rangle \top \tag{9}$$

$$\neg \langle [i] \rangle \bot \tag{10}$$

$$\Box_i^{\emptyset} (\varphi \rightarrow \psi) \rightarrow (\langle [i] \rangle \varphi \rightarrow \langle [i] \rangle \psi) \tag{11}$$

and that the following inference rule preserves validity:

$$\frac{\varphi \rightarrow \psi}{\langle [i] \rangle \varphi \rightarrow \langle [i] \rangle \psi} \tag{12}$$

Future work will be devoted to proving the conjecture that the validities and inference rule above, together with the principles of the logic LB given in Definition 13, provide a complete axiomatization of the language $\mathcal{L}$ extended with the reason modalities $\langle [i] \rangle$.

Dialogical argumentation In this paper, we have established the relationship between our epistemic logic framework and Besnard and Hunter's (B&H) notion of deductive argument. By linking our framework to B&H's notion, we can represent and infer the different types of attack described in their theory, including defeaters, undercuts, and rebuttals. A defeater occurs when the conclusion of one argument contradicts a premise of another. An undercut occurs when the conclusion of one argument directly negates a premise of another. A rebuttal occurs when two arguments have opposing conclusions. Once B&H's notion of attack is expressed within our framework, it can be applied to dialogical argumentation in a multi-agent setting, where each agent infers arguments from its private belief base. For example, it can model a dialogue between two agents defending opposing views about a given fact, each producing arguments and counterarguments in support of their position and against that of their opponent. The framework can also be used in multi-agent negotiation, where arguments are employed to persuade, justify positions, resolve conflicts, and reach agreements, as well as in persuasive human–machine dialogue, where a conversational agent must attribute beliefs to the human in order to persuade them by presenting arguments for a given claim. The formalization of the different notions of attack in line with B&H's theory, along with the extension of our analysis to dialogical argumentation, will be the subject of future work. This extension will allow us to fully address the criticism raised in (Betz 2016) of the so-called knowledge base interpretation of abstract argumentation, namely its focus on the arguments of a single participant and its inability to account for how arguments from others can influence that participant's beliefs. Our multi-agent belief base approach overcomes this limitation by modelling multiple agents who construct arguments privately but can also exchange, receive, and incorporate arguments from others into their own belief bases.

Collective arguments In the paper, we have not addressed doxastic group notions such as (explicit and implicit) distributed and common belief. These notions have been studied in previous work within the belief base framework (Lorini and Rapion 2022; Herzig et al. 2020). Extending our framework to include doxastic group attitudes would be an interesting direction for studying a notion of collective argument, that is, the idea that a group of agents collectively holds an argument supporting a certain conclusion.

Acknowledgments

This work supported by the ANR EpiRL project ANR-22-CE23-0029 and the TIRIS project CaRe.

References

Amgoud, L., and Demolombe, R. 2014. An argumentation-based approach for reasoning about trust in information sources. *Argument and Computation* 5(2-3):191–215.

Artemov, S. N. 2008. The logic of justification. *The Review of Symbolic Logic* 1(4):477–513.

Baltag, A.; Bezhanishvili, N.; Özgün, A.; and Smets, S. 2016. Justified belief and the topology of evidence. In *Proceedings of the 23rd International Workshop on Logic, Language, Information, and Computation (WOLLIC 2016)*, volume 9803 of *Lecture Notes in Computer Science*, 83–103. Berlin, Heidelberg: Springer Berlin Heidelberg.

Baumeister, D.; Neugebauer, D.; Rothe, J.; and Schadrack, H. 2018. Verification in incomplete argumentation frameworks. *Artificial Intelligence* 264:1–26.

Bench-Capon, T., and Dunne, P. E. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171(10-15):619–641.

Besnard, P., and Hunter, A. 2001. A logic-based theory of deductive arguments. *Artificial Intelligence* 128(1-2):203–235.

Besnard, P., and Hunter, A. 2008. *Elements of Argumentation*. Cambridge, MA: MIT Press.

Betz, G. 2016. Assessing the epistemological relevance of dung-style argumentation theories. *Annals of Mathematics and Artificial Intelligence* 78:303–321.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–357.

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

Groarke, L. 2017. Informal logic. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. Fall 2017 Edition.

Grossi, D., and van der Hoek, W. 2014. Justified beliefs by justified arguments. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 131–141. AAAI Press.

Hemaspaandra, E. 1993. *Complexity of Modal Logics*. Ph.d. thesis, University of Amsterdam.

Hemaspaandra, E. 1996. The price of universality. *Notre Dame Journal of Formal Logic* 37(2):174–203.

Herzig, A., and Yuste-Ginel, A. 2021a. Multi-agent abstract argumentation frameworks with incomplete knowledge of attacks. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, 1922–1928. IJCAI Organization.

Herzig, A., and Yuste-Ginel, A. 2021b. On the epistemic logic of incomplete argumentation frameworks. In *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning (KR 2021)*, 681–685. IJCAI Organization.

Herzig, A.; Lorini, E.; Perrotin, E.; Romero, F.; and Schwarzentruber, F. 2020. A logic of explicit and implicit distributed belief. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020)*, volume 325, 753–760. IOS Press.

Herzig, A.; Lorini, E.; and Maffre, F. 2018. Possible worlds semantics based on observation and communication. In van Ditmarsch, H., and Sandu, G., eds., *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*. Springer. 339–362.

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

Konolige, K. 1986. *A deduction model of belief*. Los Altos: Morgan Kaufmann Publishers.

Lorini, E., and Rapion, É. 2022. Logical theories of collective attitudes and the belief base perspective. In *Proceedings of the 21st International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2022)*, 833–841. International Foundation for Autonomous Agents and Multiagent Systems.

Lorini, E. 2018. In praise of belief bases: Doing epistemic logic without possible worlds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 1915–1922. AAAI Press.

Lorini, E. 2020. Rethinking epistemic logic with belief bases. *Artificial Intelligence* 282.

Mercier, H., and Sperber, D. 2011. Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences* 34(2):57–74.

Mercier, H., and Sperber, D. 2017. *The Enigma of Reason*. Harvard University Press.

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

Proietti, C., and Yuste-Ginel, A. 2021. Dynamic epistemic logics for abstract argumentation. *Synthese* 199(3):8641–8700.

Sakama, C., and Son, T. C. 2019. Epistemic argumentation framework. In *Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2019)*, volume 11671 of *Lecture Notes in Computer Science*, 718–732. Springer.

Sakama, C., and Son, T. C. 2020. Epistemic argumentation framework: Theory and computation. *Journal of Artificial Intelligence Research* 69:1103–1126.

Schwarzentruber, F.; Vesic, S.; and Rienstra, T. 2012. Building an epistemic logic for argumentation. In *Proceedings of the 13th European Conference on Logics in Artificial Intelligence (JELIA 2012)*, volume 7519 of *Lecture Notes in Computer Science*, 359–371. Springer.

Shi, C.; Smets, S.; and Velázquez-Quesada, F. R. 2017. Argument-based belief in topological structures. In *Proceedings Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2O17)*, volume 251 of *EPTCS*, 489–503.

Shi, C.; Smets, S.; and Velázquez-Quesada, F. R. 2023. Logic of justified beliefs based on argumentation. *Erkenntnis* 88(3):1207–1243.

Stalnaker, R. 2002. Common ground. *Linguistics and philosophy* 25(5/6):701–721.

van Benthem, J., and Pacuit, E. 2011. Dynamic logics of evidence-based beliefs. *Studia Logica* 99(1-3):61–92.

van Benthem, J.; Duque, D. F.; and Pacuit, E. 2014. Evidence and plausibility in neighborhood structures. *Annals of Pure and Applied Logic* 165(1):106–133.

Walton, D., and Krabbe, E. C. W. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press.

Wáng, Y. N., and Li, X. 2021. A logic of knowledge based on abstract arguments. *Journal of Logic and Computation* 31(8):2004–2027.

Özgün, A. 2017. *Evidence in Epistemic Logic: A Topo-*

logical Perspective. Phd thesis, University of Amsterdam, Amsterdam, The Netherlands.