Justified Beliefs by Justified Arguments

Davide Grossi and Wiebe van der Hoek
University of Liverpool

Abstract

The paper addresses how the information state of an agent relates to the arguments that the agent endorses. Information states are modeled in doxastic logic and arguments by recasting abstract argumentation theory in a modal logic format. The two perspectives are combined by an application of the theory of product logics, delivering sound and complete systems in which the interaction of arguments and beliefs is investigated.

1 Introduction

Leaving probabilistic approaches aside, one can recognize in the KR tradition two main paradigms in the formal representation of the knowledge or, more generally, of the information state of an agent. One describes knowledge as what holds in all the situations the agent considers possible—the epistemic alternatives. That is the perspective of epistemic logic (Meyer and van der Hoek 1995; Fagin *et al.* 1995). The other describes knowledge as some form of 'stable' theory about the world the agent endorses in the face of conflicting information. This is the tradition of non-monotonic reasoning and of its most successful recent development, abstract argumentation theory (Dung 1995).

The paper combines these two dimensions in a unified framework where questions concerning the interaction of arguments and beliefs can be systematically addressed, such as: Is the set of arguments supporting an agent's doxastic state 'justifiable' from the point of view of abstract argumentation (e.g., conflict-free, admissible, stable, etc.)?

Technically, the combination of the two perspectives—epistemic and argumentation-theoretic—is achieved by deploying techniques and results from the theory of product modal logics (Gabbay *et al.* 2003). The key idea consists in taking states in a doxastic (Kripke) model and arguments in a (Dung) attack graph as two orthogonal dimensions for the description of the information state of an agent. This intuition suggests the use of bi-dimensional structures for the study of argument-based beliefs. The logics obtained are studied in their completeness and finite model properties.

More generally, the paper lays a bridge between epistemic logic and argumentation theory. We hope that the results presented can foster further interaction between the two fields.

Related work To the best of our knowledge, the only works to date attempting to interface argumentation with epistemic logic are (Grossi 2012) and (Schwarzentruber *et al.* 2012). The first is concerned with the analysis, in dynamic epistemic logic (van Ditmarsch *et al.* 2007), of the fixpoint behavior of some argumentation theoretic notions, and the second enriches the standard framework of abstract argumentation by enabling arguers to hold beliefs about other arguers' available arguments. In its broad purposes, the present paper can be related to recent work (in particular, (Artemov 2008; van Benthem and Pacuit 2011)) aiming at explicitly modeling the 'justifications' or 'reasons' upon which agents base their information state. Our paper shows the viability of using product logics for this type of analysis.

Outline of the paper The following section prepares the ground recapitulating some basic notions from both epistemic logic and abstract argumentation. We then proceed in a modular fashion. First (in Section 3) we introduce a logic that combines the simplest modal logic of argumentation (logic K, cf. (Grossi 2010)) and the simplest modal doxastic logic ($\mathsf{KD45}$). Argument-belief interaction properties are then formalized and discussed in Section 4. Then, we allow the agent to have preferences among the arguments. Finally, Section 6 then introduces and studies a more expressive logic, able to formalize ample fragments of Dung's argumentation theory as well as a rich set of doxastic attitudes based on the set of arguments that an agent endorses. We conclude by pointing at future research directions at the interface of abstract argumentation and epistemic logic. All proofs (or proof sketches) are provided in the appendix.

2 Preliminaries

We start by introducing the basic building blocks of our analysis: simple structures for representing beliefs and for representing arguments and their attacks. We then move to motivate a specific way of combining the two: products. Although we give all necessary definitions, space limitations demand we do this succinctly: the reader may wish to consult (Meyer and van der Hoek 1995) for more background on doxastic logic, (Baroni and Giacomin 2009;

Baroni *et al.* 2011) for abstract argumentation, and (Gabbay and Shehtman 1998; Gabbay *et al.* 2003) for products in modal logic.

Doxastic structures

Definition 1 (Doxastic frame)**.** *A doxastic frame is a tuple* $\mathcal{D} = \langle S, B\rangle$ *where: i)* S *is a non-empty set of states; ii)* $B \subseteq S$*, s.t.,* $B \neq \emptyset$*. The class of all doxastic frames is denoted* $\mathfrak{D}$*.*

A doxastic frame represents the non-empty set B of states that an agent holds as most plausible among the set of all states S. We will refer to elements of B as *doxastic alternatives* or *doxastically accessible states*.

Given a doxastic frame, a Kripke model is obtained by adding a valuation function $\mathcal{V} : \mathbf{P} \longrightarrow \wp(S)$ interpreting a set of atoms $\mathbf{P}$. A doxastic modality $\diamondminus_{\mathsf{B}}$ can then be interpreted as follows:

$$\langle \mathcal{D}, \mathcal{V}\rangle, s \models \diamondminus_{\mathsf{B}}\varphi \quad \text{iff} \quad \exists s' \in B : \langle \mathcal{D}, \mathcal{V}\rangle, s' \models \varphi \qquad (1)$$

For any diamond $\Diamond$ in this paper, we define an associated $\Box$ as $\Box\varphi = \neg\Diamond\neg\varphi$. So $\boxminus_{\mathsf{B}}\varphi$ means that the agent believes φ.

Doxastic frames are somewhat simpler than the structures typically used to study beliefs, namely relational frames where the accessibility relation is transitive and euclidean (cf. (Meyer and van der Hoek 1995)). The two, however, can be proven equivalent for the purpose of this paper, using some standard modal logic arguments:

Theorem 1. *On the basic modal language, the class of doxastic frames is modally equivalent to the class of serial, transitive and euclidean relational frames.*

So the logic of the class of doxastic frames is completely axiomatized by the standard axiom system for logic $\mathsf{KD45}$ containing: the rules Modus Ponens ($\mathbf{MP}$: form φ and $\varphi \to \psi$, infer ψ) and Necessitation ($\mathbf{Nec}$: from φ, infer $\boxminus_{\mathsf{B}}\varphi$); the axioms $\mathbf{K} : \boxminus_{\mathsf{B}}(\varphi \to \psi) \to (\boxminus_{\mathsf{B}}\varphi \to \boxminus_{\mathsf{B}}\psi)$ representing the agent's ability to reason propositionally, $\mathbf{D} : \neg\boxminus_{\mathsf{B}}\bot$ (beliefs are consistent) and the axioms $\mathbf{4} : \boxminus_{\mathsf{B}}\varphi \to \boxminus_{\mathsf{B}}\boxminus_{\mathsf{B}}\varphi$ and $\mathbf{5} : \neg\boxminus_{\mathsf{B}}\neg\varphi \to \boxminus_{\mathsf{B}}\neg\boxminus_{\mathsf{B}}\neg\varphi$ representing positive and negative introspection, respectively. For later reference, recall that $\mathsf{S5}$ is the logic $\mathsf{KD45} + \mathbf{T} : \boxminus_{\mathsf{B}}\varphi \to \varphi$.

Argumentative structures

We start by the key structure of abstract argumentation:

Definition 2 (Attack graphs (Dung 1995))**.** *An attack graph is a tuple* $\mathcal{A} = \langle A, \to\rangle$ *where: i)* A *is a non-empty set of arguments; ii)* $\to \subseteq A^2$ *is a binary relation (*$a \to b$ *stands for* a *attacks* b*). The class of all attack graphs is denoted* $\mathfrak{A}$*.*

These relational structures are the building blocks of abstract argumentation theory. Once A is taken to represent a set of arguments, and $\to$ an 'attack' relation between arguments, the study of these structures provides general insights on how competing arguments interact, and structural properties of subsets of A can be taken to formalize how collections of arguments form 'justifiable' positions in an argumentation ((Baroni and Giacomin 2009; Baroni *et al.* 2011)).

Conflict-free	$X \subseteq \{x \mid \nexists y \in X : x \leftarrow y\}$
Self-defended	$X \subseteq \{x \mid \forall y : x \leftarrow y \Rightarrow \exists z \in X, y \leftarrow z\}$
Admissible	X is conflict-free and self-defended
Complete extension	X is conflict-free & $X = \{x \mid \forall y : x \leftarrow y \Rightarrow \exists z \in X, y \leftarrow z\}$
Stable extension	$X = \{x \mid \nexists y \in X : x \leftarrow y\}$

Table 1: Properties of a set of arguments X in a given attack graph $\mathcal{A}$.

In this paper we will touch upon the argumentation-theoretic notions of conflict-freeness, self-defense, admissibility, complete and stable extensions. Table 1 recapitulates these notions for the ease of the reader.

A Kripke model $\mathcal{M} = \langle A, \leftarrow, \mathcal{V}\rangle$ can be obtained from an attack graph by inverting the attack relation ($a \leftarrow b$ denotes that a is attacked by b) and by adding a valuation function $\mathcal{V} : \mathbf{P} \longrightarrow \wp(A)$ interpreting a set of propositional atoms $\mathbf{P}$. Consider now a modality $\diamondminus_{\mathsf{A}}$ with the following semantics:

$$\langle \mathcal{A}, \mathcal{V}\rangle, a \models \diamondminus_{\mathsf{A}}\varphi \ \text{iff} \ \exists b \in A : a \leftarrow b \ \& \ \langle \mathcal{A}, \mathcal{V}\rangle, b \models \varphi \quad (2)$$

An argument a satisfies $\diamondminus_{\mathsf{A}}\varphi$ iff some attacker b of a satisfies φ. The logic of $\diamondminus_{\mathsf{A}}$ defined by the class of attack graphs is, obviously, K. (Grossi 2010) shows that modal logic K can express a number of argumentation theoretic notions from (Dung 1995), such as: $\neg\diamondminus_{\mathsf{A}} p$, expressing that the current argument is not attacked by p; or $\boxminus_{\mathsf{A}}\diamondminus_{\mathsf{A}} p$ expressing that the current argument is 'defended' by p-arguments (i.e., its attackers are attacked by p-arguments). The logic K is axiomatized by rules $\mathbf{MP}$ and $\mathbf{Nec}$, and axiom $\mathbf{K}$.

Doxo-argumentative structures

Let us start with a simple motivating example:

Example 1 (After (Modgil 2009))**.** *Consider two individuals exchanging arguments about the weather forecast. Argument* a*: "Today will be dry in London since the BBC forecasts sunshine". And argument* a'*: "Today will be wet in London since CNN forecasts rain". We have two arguments (*a *and* a'*) concerned with whether the 'real' situation is a state* s *where the sun shines in London or in a state* s' *where it rains in London.*

In general, starting from a set of doxastic alternatives S and a set of 'arguments' A, we are after structures that can support the analysis of how elements of S interact with elements of A. We want to be able to express properties of state-argument pairs (s, a) such as, in particular, that "argument a supports state s" in the sense that a is ground for s or increases its likeliness.

Example 2 (After (Modgil 2009), continued)**.** *Let then* $S = \{s, s'\}$ *and* $A = \{a, a'\}$*. We can represent the simple scenario of Example 1 by the model on the left of Figure 1, where the dark circles indicate that the pairs at issue have a property of interest (in this case the property of 'supporting'):* a *supports* s *and does not support* s'*; vice versa* a' *supports* s' *and does not support* s*.*

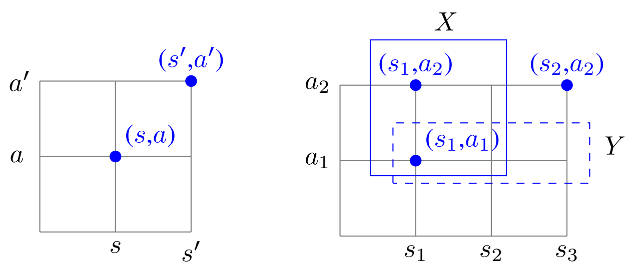

Figure 1: Two doxo-argumentative structures.

So our domain becomes the Cartesian product $S \times A$. Now, if S and A also come equipped with accessibility relations—the doxastic one in case of S and the attack one in case of A—then studying how the logics of these relations interact in $S \times A$ would allow one to talk about properties of state-argument pairs that are of a doxastic and argumentation-theoretic type in the same language. The paper takes this perspective and sets out to develop a formal theory of how arguments and their attacks relate to the doxastic state of an agent. The key tool in accomplishing this, is that of product logics.

Properties of states, arguments and their pairs

The above set up allows one to represent any property of state-argument pairs. In this paper, we will focus on one in particular, namely the property of *support* of an argument for a doxastic state. This will be denoted by symbol σ. Notice that the same argument can support several states and the same state can be supported by several arguments. If a set of states B is supported by a same argument a we say that a *supports* B.

Critically, the set up allows one to represent properties of states, or arguments alone. For instance, to express that a state s has a property X (e.g., 'the sun shines in London') it suffices for X to be true of all the state-argument pairs whose state is s, i.e., to be a set of columns in the Cartesian plane. Similarly, properties Y of arguments alone (e.g., 'being upheld by BBC') can be represented in the same fashion, i.e., by sets of rows of the Cartesian plane.[1]

Figure 1 (right) illustrates all these different properties. Dark circles indicate a relation (e.g., support) between arguments and states: a_2 supports both s_1 and s_3; s_1 is supported by both a_1 and a_2. Rectangle X represents a state-only property, of states s_1 and s_2, and rectangle Y (dashed line) represents an argument-only property, of argument a_1.

Product logics

The product of two (uni-)modal logics[2] is defined as follows (Gabbay and Shehtman 1998). The product $\mathcal{F} \times \mathcal{F}'$ between two frames $\mathcal{F} = \langle S, R\rangle$ and $\mathcal{F}' = \langle S', R'\rangle$ is the frame $\langle S \times S', H, V\rangle$ where:

$$(s,s')H(t,t') \quad \text{iff} \quad sRt \text{ AND } s' = t'$$
$$(s,s')V(t,t') \quad \text{iff} \quad s'R't' \text{ AND } s = t$$

Intuitively, the product of two frames can be depicted as a Cartesian plane where H is the relation on the 'horizontal' dimension consisting of the set S and V is the relation on the the 'vertical' one consisting of set S'. Following (Marx 1999), we will use $\diamondminus$ to denote the modality interpreted over H—'horizontal' modality—and $\diamondbar$ the modality interpreted over V—'vertical' modality.

The product of two classes of frames $\mathfrak{F}$ and $\mathfrak{F}'$ is $\{\mathcal{F} \times \mathcal{F}' \mid \mathcal{F} \in \mathfrak{F} \text{ AND } \mathcal{F}' \in \mathfrak{F}'\}$. Now, given two logics L and L' the product $\mathsf{L} \times \mathsf{L}'$ is the logic of the class of frames defined by the product of the two largest classes $\mathfrak{F}$ and $\mathfrak{F}'$ for which the two logics are complete. For instance, $\mathsf{K} \times \mathsf{K}$ is the logic of the class of all frames consisting of the product of two frames. Here we study products between logics L and L' where L—the 'horizontal' logic—is a doxastic logic and L'—the 'vertical' logic—is a modal logic for argumentation.

3 A simple product logic: $\mathsf{KD45} \times \mathsf{K}$

As a first framework in which to investigate interaction principles between arguments and doxastic states we consider the product of the simplest doxastic logic, namely $\mathsf{KD45}$, with the simplest modal logic of attack graphs, i.e., K.

Syntax and semantics

The language $\mathcal{L}(\diamondminus_{\mathsf{B}}, \diamondbar_{\mathsf{A}})$, has the following BNF:

$$p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \diamondminus_{\mathsf{B}}\varphi \mid \diamondbar_{\mathsf{A}}\varphi$$

where p belongs to the set of atoms $\mathbf{P}$. For any language $\mathcal{L}$ that we consider, the variant $\mathcal{L}^{\sigma}$ adds an atom σ to it, where σ intuitively says of (s,a) that s is supported by a. Semantics is given as follows. Let $\mathcal{D}$ be a doxastic frame on S and $\mathcal{A}$ an attack graph on A. A $\mathsf{KD45} \times \mathsf{K}$ model is a structure $\mathcal{M} = \langle \mathcal{D} \times \mathcal{A}, \mathcal{V}\rangle$ where $\mathcal{V} : \mathbf{P} \longrightarrow \wp(S \times A)$. The satisfaction relation is defined by the standard Boolean clauses plus the following clauses (cf. the expressions (1) and (2)):

$$\mathcal{M},(s,a) \models \diamondbar_{\mathsf{A}}\varphi \quad \text{iff} \quad \exists a' \in A : a \leftarrow a' \ \&\ \mathcal{M},(s,a') \models \varphi$$
$$\mathcal{M},(s,a) \models \diamondminus_{\mathsf{B}}\varphi \quad \text{iff} \quad \exists s' \in B : \mathcal{M},(s',a) \models \varphi$$

As usual, we denote the truth-set of a formula φ by $[\![\varphi]\!]$.

The expression $\mathcal{M},(s,a) \models \varphi$ can be interpreted as: given the 'actual' state is s and the 'currently entertained' argument is a, φ holds. So $\boxminus_{\mathsf{B}}\varphi$ expresses the property that, by keeping fixed the current argument, all pairs consisting of the current argument and a doxastically accessible state, satisfy φ. Intuitively: it is believed that φ holds of the current argument. Similarly, modalities $\diamondbar_{\mathsf{A}}$ and $\boxbar_{\mathsf{A}}$ express properties of the attack relation. So $\diamondbar_{\mathsf{A}}\varphi$ expresses the property that, by keeping the current state fixed, there exists a pair consisting of the current state and an argument that attacks the current argument, and this pair satisfies φ.

Remark 1 (Satisfaction in products)**.** *As customary in modal logic, formulae are interpreted on pointed models*

[1]Cf. (Sano 2011).

[2]The multi-modal case is a straightforward generalization. Cf. (Gabbay and Shehtman 1998).

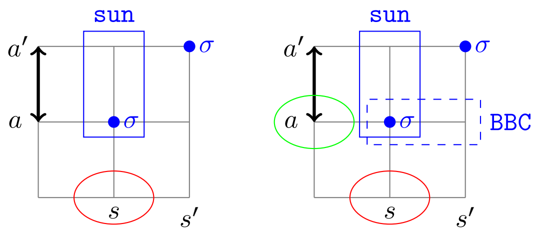

Figure 2: Rendering of Example 1 as a $\mathsf{KD45}\times\mathsf{K}$ model (left) and a DA model (right). Universal relations are not depicted.

"$\mathcal{M},(s,a)$". So, when we interpret a formula we fix both an argument and a state and $\mathcal{M},(s,a)\models\varphi$ can be interpreted as: given the 'actual' state is s and the 'currently entertained' argument is a, φ holds.

Example 3 (After (Modgil 2009), continued)**.** *We extend Example 1 by making explicit that the two arguments a and a' attack one another, and that the agent believes the actual state is s (the only one supported by argument a), so that the set of doxastic alternatives B is $\{s\}$ (left of Figure 2, where the ellipsis encloses the set of doxastic alternatives). Dark circles denote the truth set of atom σ (representing 'support') and the rectangle denotes the truth set of an atom* sun *(for 'sunshine'). Notice that* sun *is here a 'column' property (Remark 2). Arrows on the vertical dimension denote attack.*

Here are some formulae true at (s,a): (i) $\Box_{\mathsf{B}}\sigma$; (ii) $\Box_{\mathsf{B}}(\sigma\wedge\Box_{\mathsf{A}}\neg\sigma)$. Intuitively: (i) says that my beliefs are supported by the current argument, in the sense that all my doxastic alternatives are supported by the current argument; (ii) says that all my doxastic alternatives are supported by the current argument and all its attackers are not supported by my doxastic alternatives. Here are some of the formulae true at (s',a'): (i) $\neg\mathtt{sun}\wedge\Box_{\mathsf{B}}\mathtt{sun}$; (ii) $\Box_{\mathsf{B}}\neg\sigma$; (iii) $\Box_{\mathsf{B}}\Diamond_{\mathsf{A}}\sigma$. Intuitively: (i) expresses a standard false belief property 'I believe sun *of the current argument but* sun *is false'; (ii) expresses that no doxastic alternative is supported by the current argument; (iii) states that for all doxastic alternatives, the current argument is attacked by an argument supporting that alternative.*

Metalogical results

Axiomatization Logic $\mathsf{KD45}\times\mathsf{K}$ is the logic on $\mathcal{L}$ of the class of frames consisting of the product of one single-agent doxastic frame and one attack frame. It is axiomatized by taking $\mathsf{KD45}$ for $\Diamond_{\mathsf{B}}$, and K for $\Diamond_{\mathsf{A}}$, plus the two following axioms:

Com	$\Diamond_{\mathsf{A}}\Diamond_{\mathsf{B}}\varphi\leftrightarrow\Diamond_{\mathsf{B}}\Diamond_{\mathsf{A}}\varphi$
Con	$\Diamond_{\mathsf{B}}\Box_{\mathsf{A}}\varphi\rightarrow\Box_{\mathsf{A}}\Diamond_{\mathsf{B}}\varphi$

We will come back later to the intuitive meaning of these.

The completeness of this axiom system is established as a corollary of known general theorems (Gabbay and Shehtman 1998, Theorem 7.12) or (Gabbay *et al.* 2003, Theorem 5.9): we only need to check that the axioms for K and $\mathsf{KD45}$ are either without atoms (i.e., frame formulae), or else have a specific syntactic form (called *pseudo-transitivity*), which is the case.[3]

Decidability & Complexity A logic has the *product finite model property* w.r.t. class $\mathfrak{F}\times\mathfrak{F}'$ iff every satisfiable formula on that class can be satisfied on a model built on the product of finite frames in $\mathfrak{F}$ and $\mathfrak{F}'$.

Logic $\mathsf{KD45}\times\mathsf{K}$ has the (strong) product finite model property as every φ can be satisfied on a finite model of size exponential in the length of φ (Gabbay *et al.* 2003, Theorem 6.56). Logic $\mathsf{KD45}\times\mathsf{K}$ is therefore decidable and its satisfiability problem is NEXPTIME-complete (Gabbay *et al.* 2003, Theorem 6.57).

4 Interaction of attacks and beliefs

We now turn to the sort of insights that we gain by modeling the interaction of doxastic structures (Kripke frames) and argument structures (attack graphs) as a product, and what logic $\mathsf{KD45}\times\mathsf{K}$ allows us to say about such interaction.

Beliefs

Let us first define formally the classes of state- and argument-properties we have informally introduced in an earlier section. A state-property (column property) is expressed by a formula $\sigma\in\mathcal{L}(\Diamond_{\mathsf{B}},\Diamond_{\mathsf{A}})$ such that the following equation holds:

$$[\![\sigma]\!]=\{(s,a)\in S\times A\mid\exists a'\in A:(s,a')\in[\![\sigma]\!]\}\,.$$

Similarly, an argument-property (row property) is expressed by a formula $\gamma\in\mathcal{L}(\Diamond_{\mathsf{B}},\Diamond_{\mathsf{A}})$ such that the following equation holds:

$$[\![\alpha]\!]=\{(s,a)\in S\times A\mid\exists s'\in A:(s',a)\in[\![\alpha]\!]\}\,.^{4}$$

In the remaining of the paper we will denote state-properties by a variable σ and argument properties by a variable α.

Logic $\mathsf{KD45}\times\mathsf{K}$ supports the representation of two types of beliefs: the beliefs an agent can entertain about an 'external' state-of-affairs, and the beliefs an agent can entertain about how the states it considers possible interact with available arguments.

Beliefs of the first type concern state-properties: $\Box_{\mathsf{B}}\sigma$. Notice that beliefs of state-properties are themselves state-properties. That is, formulae $\Box_{\mathsf{B}}\sigma$ are independent of the vertical coordinate, so we have $\mathcal{M},(s,a)\models\Box_{\mathsf{B}}\sigma$ iff $\mathcal{M},(s,a')\models\Box_{\mathsf{B}}\sigma$ for any $a'\in A$. Put differently: the equivalence $\Box_{\mathsf{B}}\sigma\leftrightarrow\Box_U\Box_{\mathsf{B}}\sigma$ is valid.

Beliefs of the second type concern properties φ expressing some kind of relation between doxastic states and arguments. Both kinds of beliefs behave according to logic

[3]More precisely, the theorem states that the product $\mathsf{L}_H\times\mathsf{L}_V$ of two logics L_H and L_V whose axioms are either formulae from the frame language (i.e., without atoms, like $\Diamond\top$) or have the form $\triangledown\Box p\rightarrow\vartriangle p$ where $\triangledown$ is a sequence of possibly different diamonds and $\vartriangle$ a sequence of possibly different boxes (so-called *pseudo-transitive formulae*), is completely axiomatized by the axioms of L_H, the axioms of L_V, plus the **Com** and **Con** axioms for each pairs of modalities in the combined language. Notice that pseudo-transitive formulae are Sahlqvist formulae.

[4]Cf. Remark 2 below.

KD45, therefore validating the standard doxastic principles (e.g., positive and negative introspection). To this type of beliefs we turn in the next section.

Interaction properties

Example 3 has already shown interaction properties expressible in $\mathsf{KD45} \times \mathsf{K}$. Here are a few more examples, which we illustrate using the designated atom σ:

$$
\begin{array}{llll}
(a) & \boxbar_{\mathsf{A}} \diamondbar_{\mathsf{A}} \top & (b) & \sigma \wedge \diamondbar_{\mathsf{A}}\top \wedge \boxbar_{\mathsf{A}}\sigma \\
(c) & \diamondminus_{\mathsf{B}}\sigma & (d) & \sigma \wedge \boxbar_{\mathsf{A}}\neg\diamondminus_{\mathsf{B}}\sigma
\end{array}
$$

To understand these properties, let us see what they express once evaluated at a pointed model $\mathcal{M}, (s, a)$, where s denotes the current state and a the argument currently entertained. Formula (a) expresses that all attackers of a have an attacker, i.e., a is *defended* by some argument. Formula (b) states that the current argument a supports the current alternative s, but a has attackers and indeed all attackers of a support s. Intuitively, a is therefore a 'weak' argument for s. Formula (c) expresses that some doxastically accessible alternatives are supported by the current argument a and hence the beliefs of the agent are supported by that argument. Formula (d) states again that the current argument is supported by the current alternative and that all attackers of the current argument do not support any alternative. This property expresses a form of *safety* of a state-argument pair: the argument supports the alternative but there is no other alternative which is supported by an argument attacking the current one.

It is worth observing that that the latter property generalises a purely argumentative notion of safety as non-existence of attackers ($\boxbar_{\mathsf{A}}\bot$), by requiring that if there are attackers, then these are not effective as they rule out any doxastic alternative. So it makes the requirement of not having attackers relative to whether they support any further doxastic alternatives. The formula $\boxminus_{\mathsf{B}}(\sigma \rightarrow \boxbar_{\mathsf{A}}\neg\diamondminus_{\mathsf{B}}\sigma)$ would then express the agent's belief of safety of the current argument, across the set of doxastic alternatives.

Validities of $\mathsf{KD45} \times \mathsf{K}$ and their interpretation

Let us move to the interpretation of $\mathbf{Com}$ and $\mathbf{Con}$.[5]

Intuitively, $\diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{B}} \varphi \leftrightarrow \diamondminus_{\mathsf{B}} \diamondbar_{\mathsf{A}} \varphi$ states that the property 'there is an attacker of the argument I'm currently entertaining, and a state I consider doxastically possible, such that the first is in a relation φ with the latter' can be formulated independently of the order of the diamonds involved. The order of quantification over attacks and over doxastic alternatives does not matter.

As to $\diamondminus_{\mathsf{B}} \boxbar_{\mathsf{A}} \varphi \rightarrow \boxbar_{\mathsf{A}} \diamondminus_{\mathsf{B}} \varphi$, it expresses that, if I hold a state as doxastically possible which is in relation φ with all the attackers of an argument I currently entertain, then all the attackers of that argument I currently entertain are in relation φ with one of my doxastic alternatives.

We conclude this section by listing some further validities of $\mathsf{KD45} \times \mathsf{K}$. They concern the interaction properties of arguments and doxastic states and can all be proven by applications of axioms $\mathbf{Com}$ and $\mathbf{Con}$.

[5] $\mathbf{Con}$ is also known as the Church-Rosser property.

Proposition 1. *The following are validities of* $\mathsf{KD45} \times \mathsf{K}$*:*

$$
\boxminus_{\mathsf{B}} \diamondbar_{\mathsf{A}} \varphi \rightarrow \diamondminus_{\mathsf{B}} \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{B}}\varphi \tag{3}
$$

$$
\diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{B}} \varphi \rightarrow \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{B}}\varphi \tag{4}
$$

Again, to interpret these formulae it is useful to remember that they are evaluated at a given state and argument. Let us comment on the Formula (3): if all the states I hold possible are such that they are in relation φ (e.g., 'support') with an attacker of the argument I currently entertain, then I hold it possible that some attacker is in relation φ (e.g., 'support') with all the states I consider possible.

Neighborhoods

We conclude this section with a comment on related work. Evidence and justification for doxastic attitudes have been studied within neighborhood semantics (Chellas 1980) in (van Benthem and Pacuit 2011). It is worth, albeit only in passing, to notice a straightforward formal link between our two-dimensional models and neighborhood frames. The intuition is that the set of states supported by each argument can be viewed as the element of a neighborhood. So, any doxo-argumentative model $\mathcal{M} = \langle \mathcal{D} \times \mathcal{A}, \mathcal{V} \rangle$, where $\mathcal{V}$ evaluates a designated constant σ, induces a static neighborhood frame $\langle S, \mathcal{N} \rangle$ where: $\mathcal{N} = \{\{s \in S \mid \exists a \in A : \mathcal{M}, (s, a) \models \sigma\}\}_{a \in A}$ Vice versa, any static neighborhood frame $\langle S, \mathcal{N} \rangle$ induces a class of doxo-argumentative models such $\mathcal{M} = \langle \mathcal{D} \times \mathcal{A}, \mathcal{V} \rangle$ such that: the support of $\mathcal{D}$ is S; the support of $\mathcal{A}$ is $\mathcal{N}$; $\mathcal{V}$ is such that $\mathcal{M}, (s, a) \models \sigma$ iff $s \in a \in \mathcal{N}$. With respect to neighborhood frames, doxo-argumentative models contain extra structural information encoded in the attack relation.

5 Believing & endorsing

We now want to express properties about the interaction of a given set of doxastic alternatives and a given set of arguments representing the arguments the agent *endorses*. Logic $\mathsf{KD45} \times \mathsf{K}$ lacks this expressivity. So, to do this, we enrich the K component of the logic. We move from simple attack graphs $\langle A, \rightarrow \rangle$ to attack graphs designating a non-empty set E of 'endorsed arguments': $\mathcal{E} = \langle A, \rightarrow, E \rangle$. We call these frames *enriched attack graphs* and we denote their class by $\mathfrak{E}$. In this section we study the logic of the products of doxastic frames with enriched attack graphs and discuss the sort of insights the logic enables on the interaction between arguments and beliefs.

Syntax and semantics

The language $\mathcal{L}^E$ is defined by the following BNF:

$$
p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \diamondminus_{\mathsf{B}}\varphi \mid \diamondbar_{\mathsf{E}}\varphi \mid \diamondbar_{\mathsf{A}}\varphi \mid \diamondminus_{\mathsf{U}}\varphi \mid \diamondbar_{\mathsf{U}}\varphi
$$

with $p \in \mathbf{P}$. As before, $\mathcal{L}^{E,\sigma}$ extends $\mathcal{L}^E$ with the designated atom σ. Modalities $\diamondminus_{\mathsf{B}}$ and $\diamondbar_{\mathsf{A}}$ are as above. As to the others: $\diamondbar_{\mathsf{E}}$ means 'for some endorsed argument by keeping fixed the current state'; $\diamondminus_{\mathsf{U}}$ means 'for some state by keeping fixed the current argument'; $\diamondbar_{\mathsf{U}}$ means 'for some argument by keeping fixed the current state'. Notice that $\diamondminus_{\mathsf{U}}$ and $\diamondbar_{\mathsf{U}}$ are nothing but universal modalities for the horizontal and, vertical dimensions, respectively. We refer to the fragment of $\mathcal{L}^E$ containing only $\diamondminus_{\mathsf{B}}$, $\diamondminus_{\mathsf{U}}$ modalities as its *horizontal fragment*, and to the fragment of $\mathcal{L}^E$ containing only $\diamondbar_{\mathsf{E}}$, $\diamondbar_{\mathsf{A}}$ and $\diamondbar_{\mathsf{U}}$ modalities as its *vertical fragment*.

The semantics for $\mathcal{L}^E$ is defined as follows. Let $\mathcal{F}$ be a doxastic frame on S and $\mathcal{A}$ an enriched attack graph on A. A model is a structure $\mathcal{M} = \langle \mathcal{D} \times \mathcal{E}, \mathcal{V} \rangle$ where $\mathcal{V} : \mathbf{P} \longrightarrow \wp(S \times A)$. The satisfaction relation is defined as follows (clauses are limited to the newly introduced operators):

$$\begin{array}{lll} \mathcal{M}, (s,a) \models \diamondbar_{\mathsf{E}}\varphi & \text{iff} & \exists a' \in E : \mathcal{M}, (s,a') \models \varphi \\ \mathcal{M}, (s,a) \models \diamondminus_{\mathsf{U}}\varphi & \text{iff} & \exists s' \in S : \mathcal{M}, (s',a) \models \varphi \\ \mathcal{M}, (s,a) \models \diamondbar_{\mathsf{U}}\varphi & \text{iff} & \exists a' \in A : \mathcal{M}, (s,a') \models \varphi \end{array}$$

We call the logic on $\mathcal{L}^E$ defined by the class of the above models DA *doxastic argument logic*.

Remark 2 (‘Column’ and ‘row’ properties)**.** *Modalities $\diamondminus_{\mathsf{U}}$ and $\diamondbar_{\mathsf{U}}$ make it possible to express that given properties are state- or argument-properties. A state- or ‘column’ property is a formula σ that satisfies (in the given model) the equation: $\sigma \leftrightarrow \boxbar_{\mathsf{U}}\sigma$ (i.e., σ holds of the current pair independently of the argument). Similarly, an argument- or ‘row’ property is expressed by a formula α that satisfies the equation: $\alpha \leftrightarrow \boxminus_{\mathsf{U}}\alpha$ (i.e., α holds of the current pair independently of the state). Examples are, in the right model of Figure 2: $\boxbar_{\mathsf{U}}\mathtt{sun}$ of state s, and $\boxminus_{\mathsf{U}}\mathsf{BBC}$ of argument $\mathfrak{a}$. We can then express that the agent* believes *φ, in the sense that all its doxastc alternatives have column property φ by $\boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}} \varphi$ (in the example: $\boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}}\ \mathtt{sun}$). Similarly, we can express that all the arguments endorsed by the agent have row property φ by $\boxbar_{\mathsf{E}} \boxminus_{\mathsf{U}} \varphi$ (in the example: $\boxbar_{\mathsf{E}} \boxminus_{\mathsf{U}}\ \mathsf{BBC}$).*

Example 4 (After (Modgil 2009), continued)**.** *We expand Example 3 by recasting it as a DA model and making thus explicit that the agent endorses argument $\mathfrak{a}$, that is: $E = \{a\}$. The new model is depicted on the right of Figure 2 where the ellipsis on the vertical axis encloses the set of endorsed arguments. The following are validities of the above model (thus independent of the point of evaluation):*

$$\begin{array}{ll} (a) & \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{E}}\, \sigma \\ (b) & \boxbar_{\mathsf{E}} \boxminus_{\mathsf{B}} (\sigma \to \neg \diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{B}}\sigma) \\ (c) & \boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}}\, \mathtt{sun} \\ (d) & \boxbar_{\mathsf{U}} \boxminus_{\mathsf{U}} ((\sigma \wedge \mathtt{sun}) \to \neg \diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{U}}(\sigma \wedge \mathtt{sun})) \end{array}$$

Intuitively: (a) expresses that the agent’s beliefs are supported by arguments it endorses; (b) that for all pairs of endorsed arguments and doxastic alternatives, if the argument supports the alternative, then there is no attacker of that argument which supports some other alternative; (c) that all doxastic alternatives satisfy $\mathtt{sun}$*—i.e., the agent believes* $\mathtt{sun}$ *(cf. Remark 2); (d) that the state-argument pairs satisfying $\sigma \wedge \mathtt{sun}$ do not contain any attack between their arguments, that is, their arguments ar* conflict-free.[6] *Notice that none of these properties was expressible in the simpler language $\mathcal{L}(\diamondminus_{\mathsf{B}}, \diamondbar_{\mathsf{A}})$.*

[6]We come back to a general formalization of conflict-freeness and the other properties of Table 1 below.

Remark 3 (On the ‘support’ relation between states and arguments)**.** *We have taken a liberal view on the notion of support of a state by an argument. We have treated support as just one of the possible relations between arguments and states (others can be ‘incompatibility’, ‘weak support’, etc. depending on the application domain) and modeled it through a dedicated atom σ, whose interpretation has not been constrained. It is important to notice that meaningful classes of* DA *models can be isolated by strengthening our axiom system with axioms enforcing desirable properties on σ, such as: $\diamondminus_{\mathsf{U}}\sigma$ (every argument supports some state); $\boxminus_{\mathsf{B}} \diamondbar_{\mathsf{E}}\, \sigma$ (the agent considers possible only states that are supported by some argument).*

Metalogical results

We provide here a sound and complete axiom system for DA. The axiomatization is built from axiom systems for the horizontal and vertical components of the logic. We then establish the decidability of the logic and sample a few results on its expressivity.

Axiomatization: horizontal logic Let us first concern ourselves with axiomatizing the logic determined by the class of doxastic frames on the horizontal fragment of $\mathcal{L}^E$. Consider the logic—call it DA^H—defined by the rules and axioms of $\mathsf{KD45}$ for modality $\diamondminus_{\mathsf{B}}$, the rules and axioms of $\mathsf{S5}$ for modality $\diamondminus_{\mathsf{U}}$ plus:

$\mathbf{Inc}_{\mathsf{BU}}$	$\diamondminus_{\mathsf{B}}\varphi \to \diamondminus_{\mathsf{U}}\varphi$	$\mathbf{4}_{\mathsf{BU}}$	$\boxminus_{\mathsf{B}}\varphi \to \boxminus_{\mathsf{U}} \boxminus_{\mathsf{B}}\, \varphi$
$\mathbf{5}_{\mathsf{BU}}$	$\diamondminus_{\mathsf{B}}\varphi \to \boxminus_{\mathsf{U}} \diamondminus_{\mathsf{B}}\, \varphi$		

Lemma 1. DA^H *is sound and complete for class $\mathfrak{D}$.*

Axiomatization: vertical logic As to the logic determined by the class of enriched attack graphs on the vertical fragment of $\mathcal{L}^E$ we can proceed in a similar fashion. Notice that an enriched attack graph $\mathcal{E} = \langle A, \to, E \rangle$ can be viewed as a doxastic frame to which a binary relation $\to$ is added. So, consider the logic—call it DA^V—defined by: the rules and axioms of $\mathsf{KD45}$ for modality $\diamondbar_{\mathsf{E}}$, the rules and axioms of K for modality $\diamondbar_{\mathsf{A}}$ and the rules and axioms of $\mathsf{S5}$ for modality $\diamondminus_{\mathsf{U}}$, plus:

$\mathbf{Inc}_{\mathsf{AU}}$	$\diamondbar_{\mathsf{A}}\varphi \to \diamondbar_{\mathsf{U}}\varphi$	$\mathbf{Inc}_{\mathsf{EU}}$	$\diamondbar_{\mathsf{E}}\varphi \to \diamondbar_{\mathsf{U}}\varphi$
$\mathbf{4}_{\mathsf{EU}}$	$\boxbar_{\mathsf{E}}\varphi \to \boxbar_{\mathsf{U}} \boxbar_{\mathsf{E}}\, \varphi$	$\mathbf{5}_{\mathsf{EU}}$	$\diamondbar_{\mathsf{E}}\varphi \to \boxbar_{\mathsf{U}} \diamondbar_{\mathsf{E}}\, \varphi$

Lemma 2. DA^V *is sound and complete for class $\mathfrak{E}$.*

The proof proceeds as for Lemma 1.

An axiom system for DA Everything is now in place to prove the following result:

Theorem 2 (Completeness of DA)**.** *The logic defined by the axioms and rules of DA^H and of DA^V plus the following instances of* $\mathbf{Com}$ *and* $\mathbf{Con}$*:*

$\mathbf{Com}$	$\diamondminus \diamondbar\, \varphi \leftrightarrow \diamondbar \diamondminus\, \varphi$
$\mathbf{Con}$	$\diamondminus \boxbar\, \varphi \to \boxbar \diamondminus\, \varphi$

where $\diamondminus \in \{\diamondminus_{\mathsf{B}}, \diamondminus_{\mathsf{U}}\}$, and $\diamondbar \in \{\diamondbar_{\mathsf{E}}, \diamondbar_{\mathsf{A}}, \diamondbar_{\mathsf{U}}\}$, is sound and complete for the class $\mathfrak{D} \times \mathfrak{E}$ consisting of products of doxastic frames and enriched attack graphs.

(Product) Finite model property & decidability We finally establish the two following results concerning the decidability of DA:

Theorem 3. *Logic* DA *does not have the product finite model property.*

Theorem 4. *Logic* DA *is decidable.*

Expressivity on frames We conclude with some technical considerations about the expressivity of DA. The previous subsection has shown how basic notions from argumentation theory can be characterized at the level of models in DA. We now look at the feasibility of characterizations at frame level: are there formulae which characterize whether the set of endorsed arguments E is conflict-free and self-defended? This is a novel application of frame correspondence theory (van Benthem 1983) to abstract argumentation. We start by the following:

Proposition 2. *Let* $\mathcal{F} = \mathcal{D} \times \mathcal{E}$. *If* $\mathcal{F} \models \boxbar_{\mathsf{E}}(\boxminus_{\mathsf{U}} p \to \neg \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{U}} p)$ *then* E *in* $\mathcal{E}$ *is conflict free.*

We next show that Proposition 2 cannot be strengthened to a *characterization* of conflict-freeness:

Theorem 5. *There exists no formula* φ *of* DA *s.t.:* $\mathcal{F} \models \varphi$ *iff* E *is conflict-free, for* $\mathcal{F} \in \mathfrak{D} \times \mathfrak{E}$.

However, a characterization *does* exists for self-defense:

Proposition 3 (Characterization of self-defense of E)**.** *Let* $\mathcal{F} = \mathcal{D} \times \mathcal{E}$. *Then* $\mathcal{F} \models (\boxbar_{\mathsf{E}} p \wedge \diamondbar_{\mathsf{E}} \diamondbar_{\mathsf{A}} q) \to \diamondbar_{\mathsf{U}}(q \wedge \diamondbar_{\mathsf{A}} p)$ *if and only if* E *in* $\mathcal{E}$ *is self-defended.*

6 Justified beliefs & arguments in DA

This section concludes by showcasing DA as a rich framework for the study of argument-based beliefs.

From beliefs to argument-based beliefs

In standard doxastic logic beliefs are properties that are true of all (doxastically) accessible states. In DA, beliefs are properties that are true of all (doxastically) accessible state-argument pairs, *independently of the argument*. So, "I believe that φ" is formalized by $\boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}} \varphi$ (cf. Remark 2). But now, the argumentative structure available in DA allows us to differentiate between beliefs based on how they relate to underlying arguments, for instance by being supported by some such arguments. The following definitions illustrate the interaction between beliefs and endorsed arguments:

$$SB\varphi \;:=\; \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{U}}\sigma) \tag{5}$$

$$EB\varphi \;:=\; \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{E}}\sigma) \tag{6}$$

$$JB(\varphi, \psi) \;:=\; \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{E}}(\sigma \wedge \boxminus_{\mathsf{U}}\psi)) \tag{7}$$

The formulae appear in the order from logically weaker to logically stronger and may be taken as definitions of different types of argument-based beliefs in φ: supported belief (SB), endorsed supported belief (EB), and justified belief (JB) where the type of justification is expressed by ψ. Formula (5) expresses that the agent believes φ and its beliefs are supported by arguments—each doxastic alternative is supported by some argument. Formula (6) expresses that the agent believes φ and its beliefs are supported by arguments it endorses. Finally, Formula (7) states that the agent believes φ and its beliefs are supported by arguments it endorses and which have property ψ (notice again the use of $\boxminus_{\mathsf{U}}$ to express that such ψ is a 'row' property, *independent of the doxastic state*). In particular, such ψ can be chosen to express properties such as: "the argument belongs to a given conflict-free set", "the argument belongs to a given stable extension" and the like, to which we now turn.

Dung's argumentation theory in DA We show now how DA can capture some fundamental argumentation-theoretic properties (Table 1). These properties are properties of sets of arguments, that is, row properties (cf. Remark 2). For instance, we want to formalize the property that says that a given row property $\boxminus_{\mathsf{U}}\varphi$ identifies a set of admissible arguments. The strategy for obtaining these formalizations is based on (Grossi 2010). Let $\blacksquare := \boxminus_{\mathsf{U}}\boxbar_{\mathsf{U}}$:

(i)	φ is a conflict free set	$\blacksquare(\boxminus_{\mathsf{U}}\varphi \to \neg \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{U}}\varphi)$
(ii)	φ is a self-defended set	$\blacksquare(\boxminus_{\mathsf{U}}\varphi \to \boxbar_{\mathsf{A}} \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{U}}\varphi)$
(iii)	φ is a fixpoint of $\boxbar_{\mathsf{A}}\diamondbar_{\mathsf{A}}$	$\blacksquare(\boxminus_{\mathsf{U}}\varphi \leftrightarrow \boxbar_{\mathsf{A}} \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{U}}\varphi)$
(iv)	φ is an admissible set	$(i) \wedge (ii)$
(v)	φ is a complete extension	$(i) \wedge (iii)$
(vi)	φ is a stable extension	$\blacksquare(\boxminus_{\mathsf{U}}\varphi \leftrightarrow \neg \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{U}}\varphi)$

We comment on (i) and (ii). Formula (i) says that if an argument satisfies row property $\boxminus_{\mathsf{U}}\varphi$, then no attacker of that argument exists which also satisfies the same property. Formula (ii) states that all arguments satisfying row property $\boxminus_{\mathsf{U}}\varphi$ are such that all their attackers are defended by some argument satisfying the same property.

Example 5 (After (Modgil 2009), continued)**.** *Let us go back to the model on the right of Figure 2. It is a validity of the model that: the agent believes that* `sun` *is the case; that it has arguments for that belief, which it endorses; and that those arguments also have the property of belonging to the conflict-free set of arguments specified by property* `BBC`. *That is, the agent holds a belief to the effect that* `sun` *is the case and such belief is justified by the set of conflict-free* `BBC` *arguments. Formally:* $JB(\mathtt{sun}, \psi)$ *where* $\psi := \mathtt{BBC} \wedge \blacksquare(\boxminus_{\mathsf{U}}\mathtt{BBC} \to \neg \diamondbar_{\mathsf{A}} \boxminus_{\mathsf{U}}\mathtt{BBC})$. *It is also a validity of the model that the set of arguments supporting the agent's doxastic alternatives is an admissible set, that is:*

$$\blacksquare(\diamondminus_{\mathsf{B}}\sigma \to \neg \diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{B}}\sigma) \wedge \blacksquare(\diamondminus_{\mathsf{B}}\sigma \to \boxbar_{\mathsf{A}} \diamondbar_{\mathsf{A}} \diamondminus_{\mathsf{B}}\sigma)$$

where, notice, $\diamondminus_{\mathsf{B}}\sigma$ *is a row property* ($\diamondminus_{\mathsf{B}}\sigma \leftrightarrow \boxminus_{\mathsf{U}} \diamondminus_{\mathsf{B}} \sigma$).

Properties of argument-based beliefs

Having interfaced beliefs with properties of supporting arguments, the natural question arising is: how do the argument-based beliefs we have introduced in Formulae (5)-(7) behave with respect to standard principles of doxastic logic? In this final subsection we set out to provide a first answer to this question, and show that argument-based beliefs offer interesting insights into the failure of controversial principles such as, in particular, negative introspection.

All the above types of argument-based beliefs—viz., supported beliefs, endorsed beliefs and justified beliefs—obey the standard distribution and positive introspection principles, while they fail the negative introspection principle.

Proposition 4. *The following formulae are valid on $\mathfrak{D} \times \mathfrak{E}$:*

$$SB(\varphi \to \psi) \to (SB\varphi \to SB\psi) \quad (8)$$
$$EB(\varphi \to \psi) \to (EB\varphi \to EB\psi) \quad (9)$$
$$JB(\varphi \to \psi, \chi) \to (JB(\varphi, \chi) \to JB(\psi, \chi)) \quad (10)$$
$$SB\varphi \to SBSB\varphi \quad (11)$$
$$EB\varphi \to EBEB\varphi \quad (12)$$
$$JB(\varphi, \psi) \to JB(JB(\varphi, \psi), \psi) \quad (13)$$

Proposition 5 (Negative introspection failure)**.** *The following formulae are* not *valid on $\mathfrak{D} \times \mathfrak{E}$:*

$$\neg SB\varphi \to SB\neg SB\varphi \quad (14)$$
$$\neg EB\varphi \to EB\neg EB\varphi \quad (15)$$
$$\neg JB(\varphi, \psi) \to JB(\neg JB(\varphi, \psi), \psi) \quad (16)$$

A simple inspection of the proof (in the appendix) shows that negative introspection fails only in one specific case, namely when it is not the case that all doxastic states are supported by some argument (resp., supported by endorsed arguments, or supported by endorsed arguments with property ψ). That is, negative introspection for argument-based beliefs does not go through only when the to-be-introspected belief itself is not supported by arguments.

This suggests that, although negative introspection fails in its full form, it still holds in a restricted one:

Proposition 6. *The following formulae are valid on $\mathfrak{D} \times \mathfrak{E}$:*

$$(\neg SB\varphi \wedge \boxminus_{\mathsf{B}} \Diamond\!\!\!\mid_{\mathsf{U}} \sigma) \to SB\neg SB\varphi \quad (17)$$
$$(\neg EB\varphi \wedge \boxminus_{\mathsf{B}} \Diamond\!\!\!\mid_{\mathsf{E}} \sigma) \to EB\neg EB\varphi \quad (18)$$
$$(\neg JB\varphi \wedge \boxminus_{\mathsf{B}} \Diamond\!\!\!\mid_{\mathsf{E}} (\sigma \wedge \boxminus_{\mathsf{U}}\psi)) \to JB\neg JB\varphi \quad (19)$$

In other words, if I do not believe (with justification) that φ, but my beliefs are justified then I believe (with justification) that I do not believe (with justification) that φ.

7 Concluding remarks

Summary We proposed an approach based on product logics to study the interaction between the information state of an agent and the arguments the agent endorses. The approach has been illustrated by two product logics, which have been studied from a logical point of view, and have been used to model simple scenarios and formalize a number of interaction properties between beliefs and arguments.

Justification logic A related framework to the one we presented is *justification logic* (Artemov 2008), which formalizes the logic of statements of the type "t is a justification for φ", where t is called a, possibly complex, term. Justification logic provides an articulated solution to the logical omniscience problem of standard epistemic logic. In our framework arguments can well be seen as (atomic) terms but, unlike in justification logic, they are reified in the semantics of the logic and obtain the same 'ontological' status as doxastic states. While on the one hand this reintroduces logical omniscience, on the other hand it enables a classical modal approach for the study of (atomic) justifications viewed as arguments, which was our primary aim with this paper. Whether the reification of arguments/justifications in the semantics can be reconciled with a solution to the logical omniscience problem is an interesting open question.

Future work We point to three more research directions we consider especially worth pursuing. First, in our analysis we have assumed a primitive relation of 'support' between doxastic alternatives and arguments. This relation enabled us to say that a set of states (e.g., the truth-set of a column formula) was supported by a set of arguments whenever each state had an argument which was ground for it. Hence an argument can support a set of states only if it supports each alternative taken in isolation. This is clearly a strong requirement and is in part related to the logical omniscience issue just mentioned. Arguments might be thought of supporting sets of states without needing to support any of the states in the set taken in isolation. How to achieve this weakening while remaining within the set up of product logics is an interesting open question.

Second, we believe that studying our proposal in a multi-agent context would be another prominent venue for further research. For the beliefs, this would require that the alternatives for agents are modelled through a relation, rather than just identifying one set of states that are deemed doxastically possible. For the dimension representing the arguments, there is a range of possibilities one can consider. Are the arguments commonly known by all agents? Note that in our set-up, it is valid that $\blacklozenge \boxminus_{\mathsf{U}} \varphi \leftrightarrow \boxminus_{\mathsf{B}} \blacklozenge \boxminus_{\mathsf{U}} \varphi$ (with $\blacklozenge := \Diamond\!\!\!\!-_{\mathsf{U}} \Diamond\!\!\!\mid_{\mathsf{U}}$): there is an argument with row property φ if and only if the agent believes that. For the single agent-case this seems reasonable, but when reasoning about other agents' beliefs and justifications, there may be reasons to assume (we order them from weak to stronger) that (1) not all agents endorse the same arguments; (2) not all agents agree on the attack relation between arguments, and (3) not all agents need to be aware of the same arguments.[7] Once one agrees on some of those assumptions, one also needs to take into account that agents make such assumptions about each other (e.g., agent a believes that b's beliefs are justified by b's arguments, but not by arguments that a endorses).

Third, the paper has started from the simplest possible doxastic logic, KD45. Incorporating more sophisticated logics of belief constitutes another natural avenue of research. Apart from our interest to look at ways to cover the multi-agent case, other options that present themselves are incorporating richer order-theoretic models for doxastic logic (Baltag and Smets 2008); and studying the dyanmics supported by the horizontal and vertical dimensions with tools from dynamic epistemic logic (van Ditmarsch *et al.* 2007).

A Appendix: proofs

Proof of Theorem 1. Let $\mathfrak{F}$ the class of transitive and euclidian frames and $\mathfrak{F}_g$ the class of point-generated transitive and euclidean frames. We know that $\mathfrak{F}$ and $\mathfrak{F}_g$ are modally equivalent (Blackburn *et al.* 2001). [LEFT TO RIGHT] Assume $\mathfrak{D} \vDash \varphi$. Let $\mathcal{F}_s = \langle S_s, R_s \rangle \in \mathfrak{F}_g$ and $s \in S_s$. Define

[7] Cf. (Schwarzentruber *et al.* 2012).

$B = \{x \mid sR_s x\}$ and notice that $\mathcal{D}_s = \langle S_s, B\rangle \in \mathfrak{D}$ and hence $\mathcal{D}_s \vDash \varphi$. An easy induction shows that $\mathcal{D}_s$ and $\mathcal{F}_s$ are modally equivalent and thus $\mathfrak{F} \vDash \varphi$ [RIGHT TO LEFT] Assume $\mathfrak{F} \vDash \varphi$. Let $\mathcal{D} = \langle S, B\rangle \in \mathfrak{D}$. Define a relation R on S by xRy iff $y \in B$. It is easy to see that R is transitive and euclidean and hence $\mathcal{F} = \langle S, R\rangle \in \mathfrak{F}$. An easy induction proves that $\mathcal{D}$ and $\mathcal{F}$ are modally equivalent, entailing $\mathfrak{D} \vDash \varphi$. □

Proof sketch of Theorem 2. The result follows from Lemmata 1 and 2 by (Gabbay and Shehtman 1998, Theorem 7.12) (cf. our comment on completeness of $\mathsf{KD45} \times \mathsf{K}$). □

Proof sketch of Theorem 3. We exhibit a satisfiable formula from $\mathcal{L}^E$ which cannot be satisfied on a finite product model. Consider the following formula ψ from $\mathcal{L}^E$:

$$\boxbar_{\mathsf{U}} \diamondminus_{\mathsf{B}}\, p \wedge \boxbar_{\mathsf{U}} \boxminus_{\mathsf{B}} (p \to \diamondbar_{\mathsf{A}} \neg p) \wedge \boxbar_{\mathsf{U}} \boxminus_{\mathsf{B}} (\neg p \to \boxbar_{\mathsf{A}} \neg p).$$

Let $\mathcal{E} = \langle A, E, \to\rangle$ be such that $\langle A, \to\rangle$ consists of an infinite $\leftarrow$-ascending chain of elements $\langle x_n \mid 0 \leq n < \omega\rangle$ such that $a_i \not\to a_j$ for $i < j$. For any choice of E, this is clearly a frame in $\mathfrak{E}$. Let then $\mathcal{D}$ consist of a point s_0 accessing all other states (except itself) in the frame. Take an enumeration $s_0, s_1, \ldots$ of these states. This is clearly a frame in $\mathfrak{D}$. Set $\mathcal{V}(p) = \{(s_{n+1}, a_n) \mid 0 \leq n < \omega\}$. We have $\langle \mathcal{D} \times \mathcal{E}, \mathcal{V}\rangle, (s_0, a_0) \vDash \psi$. Clearly, no model on a finite frame in $\mathfrak{D} \times \mathfrak{E}$ satisfies ψ.[8] □

Proof sketch of Lemma 1. Soundness is straightforward. As to completeness, the axiom system can be shown to be complete with respect to the class of frames $\mathcal{F} = \langle S, R_{\mathsf{B}}, R_{\mathsf{U}}\rangle$ consisting of one equivalence relation R_{U} (axiomatized by $\mathsf{S5}$) which, within each of its equivalence classes, contains (containment is enforced by $\mathbf{Inc}_{\mathsf{BU}}$) a transitive and euclidean relation R_{B} (axiomatized by $\mathsf{KD45}$) with the additional property (enforced by axioms $\mathbf{4}_{\mathsf{BU}}$ and $\mathbf{5}_{\mathsf{BU}}$) that within each equivalence class all states have access to the same set of states: $\forall x, y, z$ (in each equivalence class): $xR_{\mathsf{B}}y$ iff $zR_{\mathsf{B}}y$ (in other words, there exists a set of states B all of which elements are R_{B}-accessible by all states in the class, cf. Definition 1). The latter property is a consequence of the fact that R_{B} is a subrelation of R_{U} (by $\mathbf{Inc}_{\mathsf{BU}}$) and that axioms $\mathbf{4}_{\mathsf{BU}}$ and $\mathbf{5}_{\mathsf{BU}}$—which, notice, are Sahlqvist—correspond to the following properties: $\forall x, y, z$ if $xR_{\mathsf{U}}y$ and $yR_{\mathsf{B}}z$ then $xR_{\mathsf{B}}z$; $\forall x, y, z$ if $xR_{\mathsf{U}}y$ and $xR_{\mathsf{B}}z$ then $yR_{\mathsf{B}}z$. Call now this class $\mathfrak{F}$ and consider the class $\mathfrak{F}_g$ of frames in $\mathfrak{F}$ which are point-generated by the equivalence relation R_{U}. Any $\mathcal{F} \in \mathfrak{F}_g$ is thus such that R_{U} is the universal relation on $\mathcal{F}$ and $\mathcal{F}$ contains one unique set B R_{B}-accessible by all elements of the frame. $\mathcal{F}$ is therefore modally equivalent (on the horizontal fragment of $\mathcal{L}^E$) to a $\mathcal{D} \in \mathfrak{D}$. Vice versa, any $\mathcal{D} \in \mathfrak{D}$ is modally equivalent to a $\mathcal{F} \in \mathfrak{F}_g$ (cf. proof of Theorem 1). So $\mathfrak{D}$ is modally equivalent to $\mathfrak{F}_g$ which is, by general results (Blackburn *et al.* 2001), modally equivalent to $\mathfrak{F}$. Therefore, the axiom system is complete for $\mathfrak{D}$. □

Proof sketch of Theorem 4. We provide a reduction of DA^H to $\mathsf{S5}$ and of DA^V to K^{U}, proving that the satisfiability of DA is reducible to the satisfiability of $\mathsf{S5} \times \mathsf{K}^{\mathsf{U}}$, which is decidable (Gabbay *et al.* 2003, Theorem 6.58). [DA^H to $\mathsf{S5}$]. Take a fresh atom p and define a translation (Boolean clauses omitted): $t(\diamondbar_{\mathsf{U}}\varphi) = \Diamond t(\varphi)$; $t(\diamondminus_{\mathsf{B}}\varphi) = \Diamond(p \wedge t(\varphi))$. We prove: φ is satisfiable in $\mathfrak{D}$ iff $t(\varphi)$ is satisfiable in the class of universal frames. [LEFT TO RIGHT] Assume $\mathcal{M}, x \vDash \varphi$. Build $\mathcal{M}' = \langle S', \mathcal{V}'\rangle$ with an extra atom p where: $S' = S$ and $\mathcal{V}' = \mathcal{V} \cup \{\langle p, S\rangle\}$. A simple induction shows that: $\mathcal{M}, x \vDash \varphi$ iff $\mathcal{M}', x \vDash t(\varphi)$. [RIGHT TO LEFT] Assume $\mathcal{M}, x \vDash t(\varphi)$. Build $\mathcal{M}' = \langle S', B, \mathcal{V}'\rangle$ where: $S' = S$, $B = \mathcal{V}(p)$ and $\mathcal{V}' = \mathcal{V} - \{\langle p, S\rangle\}$. A simple induction shows that: $\mathcal{M}, x \vDash t(\varphi)$ iff $\mathcal{M}' \vDash \varphi$. [$\mathsf{DA}^V$ to K^{U}] Similar. □

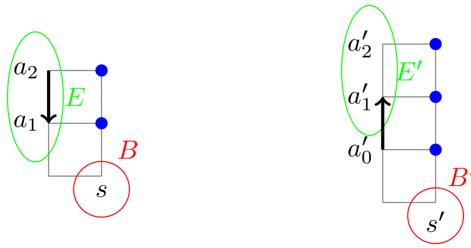

Figure 3: Two frames $\mathcal{F}$ and $\mathcal{F}'$ for the proof of Theorem 5.

Proof sketch of Theorem 5. Consider the frames $\mathcal{F}$ and $\mathcal{F}'$ of Figure 3. We have that $\forall \psi \in \mathcal{L}^E\ \mathcal{F}' \vDash \psi \Rightarrow \mathcal{F} \vDash \psi$, while at the same time, E' in $\mathcal{F}'$ is conflict free. □

Proof sketch of Proposition 4. Simple semantic arguments can be used to prove the claims. We do it for Formulae (8) and (13). [Formula (8)] Assume $\mathcal{M}, (s, a) \vDash \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}(\varphi \to \psi) \wedge \diamondbar_{\mathsf{U}}\sigma)$. It follows, by the semantics of $\boxminus_{\mathsf{B}}$ and $\boxbar_{\mathsf{U}}$ that $\mathcal{M}, (s, a) \vDash \boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}} (\varphi \to \psi) \wedge \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{U}}\, \sigma$ and then, by propositional reasoning, $\mathcal{M}, (s, a) \vDash \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{U}}\sigma) \to \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\psi \wedge \diamondbar_{\mathsf{U}}\sigma)$, which proves the claim. [Formula (13)] Assume $\mathcal{M}, (s, a) \vDash \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{E}}(\sigma \wedge \boxminus_{\mathsf{U}}\psi))$. It follows, by the semantics of $\boxminus_{\mathsf{B}}$ and $\boxbar_{\mathsf{U}}$ and propositional reasoning, that $\mathcal{M}, (s, a) \vDash \boxbar_{\mathsf{U}} \boxminus_{\mathsf{B}} (\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{E}}(\sigma \wedge \boxminus_{\mathsf{U}}\psi)) \wedge \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{E}} (\sigma \wedge \boxminus_{\mathsf{U}}\psi))$. From this, by again the semantics of $\boxminus_{\mathsf{B}}$ we conclude $\mathcal{M}, (s, a) \vDash \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}(\boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{E}}(\sigma \wedge \boxminus_{\mathsf{U}}\psi))) \wedge \diamondbar_{\mathsf{E}}(\sigma \wedge \boxminus_{\mathsf{U}}\psi))$, which proves the claim. □

Proof sketch of Proposition 5. It suffices to exhibit some counterexamples. We do that for Formula (14). Consider the model $\mathcal{M} = \langle\langle\{s\}, \{s\}\rangle \times \langle\{a\}, \varnothing, \varnothing\rangle, \mathcal{V}\rangle$ consisting of one pair (s, a) s.t. $(s, a) \in \mathcal{V}(p)$ and $(s, a) \notin \mathcal{V}(\sigma)$. We have that $\mathcal{M}, (s, a) \vDash \neg \boxminus_{\mathsf{B}} (\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{U}}\sigma)$ and $\mathcal{M}, (s, a) \nvDash \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}(\neg \boxminus_{\mathsf{B}} (\boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{U}}\sigma)) \wedge \diamondbar_{\mathsf{U}}\sigma)$. □

Proof sketch of Proposition 6. We prove the claim of Formula (17). The other two are analogous. Assume $\mathcal{M}, (s, a) \vDash \neg SB\varphi \wedge \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{U}}\, \sigma$. It follows by Formula (5) that $\mathcal{M}, (s, a) \vDash \neg \boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}}\varphi \wedge \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{U}}\, \sigma$, from which by the semantics of $\boxminus_{\mathsf{B}}$ we obtain $\mathcal{M}, (s, a) \vDash \boxminus_{\mathsf{B}}(\boxbar_{\mathsf{U}}(\neg \boxminus_{\mathsf{B}} \boxbar_{\mathsf{U}}\varphi \wedge \diamondbar_{\mathsf{U}}\sigma) \wedge \boxminus_{\mathsf{B}} \diamondbar_{\mathsf{U}}\, \sigma)$. Hence $\mathcal{M}, (s, a) \vDash SB\neg SB\varphi$. □

B Acknowledgments

The authors wish to thank Wesley Holliday for useful comments on an earlier draft of this paper.

[8] The proof goes through also for the weaker logic $\mathsf{KD45} \times \mathsf{K}^{\mathsf{U}}$.

References

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

[Baltag and Smets 2008] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. Van der Hoek, and M. Wooldridge, editors, *Logic and the Foundation of Game and Decision Theory—LOFT 07*, volume 3 of *Texts in Logic and Games*, pages 11–58. Amsterdam University Press, 2008.

[Baroni and Giacomin 2009] P. Baroni and M. Giacomin. Semantics of abstract argument systems. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artifical Intelligence*. Springer, 2009.

[Baroni *et al.* 2011] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.

[Blackburn *et al.* 2001] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.

[Chellas 1980] B. F. Chellas. *Modal Logic. An Introduction*. Cambridge University Press, Cambridge, 1980.

[Dung 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Fagin *et al.* 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about knowledge*. The MIT Press, Cambridge, 1995.

[Gabbay and Shehtman 1998] D. Gabbay and Shehtman. Products of modal logics, part 1. *Logic Journal of the IGPL*, 6(1):73–146, 1998.

[Gabbay *et al.* 2003] D. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. *Many-dimensional modal logics. Theory and applications*. Elsevier, 2003.

[Grossi 2010] D. Grossi. On the logic of argumentation theory. In W. van der Hoek, G. Kaminka, Y. Lespérance, and S. Sen, editors, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 409–416. IFAAMAS, 2010.

[Grossi 2012] D. Grossi. Fixpoints and iterated updates in abstract argumentation. In *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012)*, 2012.

[Marx 1999] M. Marx. Complexity of products of modal logics. *Journal of Logic and Computation*, 9(2):197–214, 1999.

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

[Modgil 2009] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[Sano 2011] K. Sano. Axiomatizing hybrid products of monotone neighborhood frames. *Electronic Notes in Theoretical Computer Science*, 273:51–67, 2011.

[Schwarzentruber *et al.* 2012] F. Schwarzentruber, S. Vesic, and T. Reinstra. Building an epistemic logic for argumentation. In *Proceedings of JELIA'12*, pages 359–371, 2012.

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

[van Benthem 1983] J. van Benthem. *Modal Logic and Classical Logic*. Monographs in Philosophical Logic and Formal Linguistics. Bibliopolis, 1983.

[van Ditmarsch *et al.* 2007] H. van Ditmarsch, B. Kooi, and W. van der Hoek. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer, 2007.