

An Awareness Epistemic Framework for Belief, Argumentation and Their Dynamics

Alfredo Burrieza

Antonio Yuste-Ginel

Department of Philosophy,
University of Málaga, Spain

burrieza@uma.es

antonioyusteginel@gmail.com

The notion of argumentation and the one of belief stand in a problematic relation to one another. On the one hand, argumentation is crucial for belief formation: as the outcome of a process of arguing, an agent might come to (justifiably) believe that something is the case. On the other hand, beliefs are an input for argument evaluation: arguments with believed premisses are to be considered as strictly stronger by the agent to arguments whose premisses are not believed. An awareness epistemic logic that captures qualified versions of both principles was recently proposed in the literature. This paper extends that logic in three different directions. First, we try to improve its conceptual grounds, by depicting its philosophical foundations, critically discussing some of its design choices and exploring further possibilities. Second, we provide a (heretofore missing) completeness theorem for the basic fragment of the logic. Third, we study, using techniques from dynamic epistemic logic, how different forms of information change can be captured in the framework.

1 Introduction

Belief and argumentation are two central dimensions of humans' cognitive architecture. They have received attention from antiquity to nowadays, and from a broad range of disciplines. It is then unsurprising that formal researchers have undertaken the task of modelling both phenomena. Regarding beliefs, there is an important amount of options for capturing some of its formal aspects [22]. These models usually capture *what* kind of things are believed (typically, propositions or sentences); *who* believes them (intelligent agents); and, only sometimes, *how* strong or safe these beliefs are (for instance, in probabilistic models of belief or in plausibility structures [7]). However, most of them fail to capture *why* agents do believe certain things. This lack motivates the recent trial within the epistemic logic community of capturing the missing justification component. This enterprise has been approached from a variety of methods: justification logic [3, 4, 1, 2], evidence logics based on neighbourhood semantics [12, 11], and its further topological development [5], amongst others. Yet another natural candidate to model justification consists in using conceptual and technical tools coming from argumentation theory (as done, e.g. in [23, 36, 28, 15]).

As to argumentation theory, it is a well-established, interdisciplinary field of research [19]. Since the last few decades, formal argumentation has gained more and more attention within the field of artificial intelligence, and its general advantages have been highlighted several times [10, 32]. Within formal approaches to argumentation, it is frequent to distinguish between *abstract approaches* (those that consider arguments as primitive, atomic entities) and *structured approaches* (those that explicitly account for the structure of arguments). For expository purposes, we just mention the popular Dung's approach to abstract argumentation [18], based on so-called *abstract argumentation frameworks*, and the ASPIC family of formalisms for structured argumentation, e.g., ASPIC⁺ [30, 31], that will be the main argumentative resources used in this paper.

Recently, some works have taken the first steps to explore and exploit the relations among the two different traditions (epistemic logic and formal argumentation). These can be divided in two groups. On the one hand, there are works using epistemic logic tools to reason about argumentation frameworks [35, 34, 33]. On the other hand, there are works using argumentation tools to provide an (argumentatively inspired) notion of justified belief (the already mentioned [23, 36, 28, 15]). The current paper is inserted in the latter group, and it follows the ideas of [15] that, contrarily to [23, 36, 28], and according to more standard ideas in structured argumentation, decides to model arguments as *syntactic entities*.

We start by pointing out that the informal relation between argumentation and belief is itself problematic. Arguably, there is a tension between two intuitive principles governing belief formation and argument evaluation. These principles are:

- P1 *Beliefs are an input for argument evaluation*, meaning that arguments with believed premisses are better to those with contingent or even rejected premisses.¹
- P2 *Argumentation is an input for belief formation*, meaning that rational agents should believe sentences that are ground in good arguments.

The mentioned tension arises when one tries to embrace both principles without any restriction, leading to an infinite regress. A very similar problem can be found in the root of a long-standing debate about the structure of epistemic justification within contemporary epistemology. *Foundationalist* solutions to such a tension, to which we adhere here, consists in distinguishing between *basic (non-inferred) beliefs* and *non-basic (inferred) beliefs*, where the latter inherit the justification from the former [26]. This implies accepting qualified version of both principles, but giving some sort of priority to P1 over P2. Curiously enough, an analogous distinction can be found as one of the basis of the recent argumentative theory of reason advocated by Mercier and Sperber [29]. In this context, basic beliefs are called *intuitive beliefs* while inferred beliefs are called *reflective beliefs* (see [37] for a detailed exposition of the distinction).

In the rest of this paper, we follow up the work made in [15], by extending it in three different directions. First, and after recalling the logic introduced there, whose language allows talking about basic beliefs and structured arguments, we provide its sound and complete axiomatisation (Section 2). We then explain how to use this logic for reasoning about *explicit basic beliefs* and *argument-based belief*, discussing some of the design choices, as well as depicting some alternatives (Section 3). Finally, we extend the basic fragment of the logic so as to capture different kinds of informational dynamics, illustrating their effects on both types of beliefs (Section 4).

2 An awareness logic for belief and argumentation

Let us start by recalling the logic introduced in [15]. We follow the traditional order for presentation: syntax, semantics, and proof theory. We assume a countable set of *propositional letters* At as fixed from now on. The **language** \mathcal{L} is defined as the the pair (F, A) of *formulas* and *arguments* which are respectively generated by the following grammars:

$$\begin{aligned} \varphi &::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box\varphi \mid \text{aware}(\alpha) \mid \text{conc}(\alpha) = \varphi \mid \\ &\mid \text{strict}(\alpha) \mid \text{undercuts}(\alpha, \alpha) \mid \text{wellshap}(\alpha) \quad p \in At, \alpha \in A. \\ \alpha &::= \langle \varphi \rangle \mid \langle \alpha_1, \dots, \alpha_n \rightarrow \varphi \rangle \mid \langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle \quad \varphi \in F, n \geq 1. \end{aligned}$$

¹We use the term *contingent* in its doxastic sense, that is, a sentence is said to be contingent iff it is neither believed nor believed to be false.

The rest of Boolean operators ($\rightarrow, \vee, \leftrightarrow$) and constants (\top, \perp), as well as the dual of \Box (noted \Diamond), are defined as usual. Arguments of \mathcal{L} have the following informal readings. $\langle \varphi \rangle$ is an *atomic argument*. Note that this kind of arguments are rather strange in real-life examples, since they have one sole premise and conclusion, and there is not a proper inference step. Mathematically, they can be understood as a one-line proof from φ to φ . As for $\langle \alpha_1, \dots, \alpha_n \rightarrow \varphi \rangle$ (resp. $\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle$), it represents an argument claiming that φ follows deductively (resp. defeasibly) from the conclusions of arguments $\alpha_1, \dots, \alpha_n$. As an example of a complex argument consider $\langle \langle \langle \text{Bird} \rangle, \langle \text{Bird} \rightarrow \text{Wings} \rangle \rightarrow \text{Wings} \rangle \Rightarrow \text{Flies} \rangle$ that informally reads “This has wings, because it is a bird and all birds have wings. Moreover, since it has wings, it presumably (defeasibly) flies”.

Regarding formulas, elements of At represent factual, atomic propositions. $\Box\varphi$ means that the agent implicitly (ideally) believes that φ . $\text{aware}(\alpha)$ reads “the agent is aware of α ”. $\text{conc}(\alpha) = \varphi$ reads the “conclusion of α is φ ”. $\text{strict}(\alpha)$ means that α does not contain defeasible inference steps. $\text{undercuts}(\alpha, \beta)$ means that α undercuts β , that is, α attacks some defeasible inference link of β . Finally, $\text{wellshap}(\alpha)$ means that α has been constructed properly, that is, all its deductive inference steps are valid and all its defeasible inference steps are accepted by the agent.

We use $\text{SEQ}(F)$ to denote the *set of all finite sequences over F*. We denote an arbitrary sequence of $n+1$ elements over F as $((\varphi_1, \dots, \varphi_n), \varphi)$. Sequences of formulas are useful to represent inference steps in the meta-language. Although strongly connected from a conceptual point of view, the sequence $((\varphi_1, \dots, \varphi_n), \varphi)$ is *not* the same object as, for instance, the object language argument $\langle \langle \varphi_1 \rangle, \dots, \langle \varphi_n \rangle \Rightarrow \varphi \rangle$. Let $R = ((\varphi_1, \dots, \varphi_n), \varphi) \in \text{SEQ}(F)$ we use α^R as a shorthand for $\langle \langle \varphi_1 \rangle, \dots, \langle \varphi_n \rangle \Rightarrow \varphi \rangle$. We can see α^R as the *simplest argument using R*. As an example, consider the rule $R_1 = ((\text{Wings}), \text{Flies})$, we have $\alpha^{R_1} = \langle \langle \text{Wings} \rangle \Rightarrow \text{Flies} \rangle$, but note that there are infinitely many other arguments using R_1 , for instance $\langle \langle \langle \text{Bird} \rangle, \langle \text{Bird} \rightarrow \text{Wings} \rangle \rightarrow \text{Wings} \rangle \Rightarrow \text{Flies} \rangle$.

Let us define the following meta-syntactic functions for analysing an **argument’s structure**, taken from ASPIC^+ [30]:

$\text{Prem}(\alpha)$ returns the **premises** of α and it is defined as follows: $\text{Prem}(\langle \varphi \rangle) := \{\varphi\}$, $\text{Prem}(\langle \alpha_1, \dots, \alpha_n \leftrightarrow \varphi \rangle) := \text{Prem}(\alpha_1) \cup \dots \cup \text{Prem}(\alpha_n)$ where $\leftrightarrow \in \{\rightarrow, \Rightarrow\}$.

$\text{Conc}(\alpha)$ returns the **conclusion** of α and it is defined as follows $\text{Conc}(\langle \varphi \rangle) := \{\varphi\}$ and $\text{Conc}(\langle \alpha_1, \dots, \alpha_n \leftrightarrow \varphi \rangle) := \{\varphi\}$ where $\leftrightarrow \in \{\rightarrow, \Rightarrow\}$.

$\text{sub}_A(\alpha)$ returns the **subarguments** of α and it is defined as follows: $\text{sub}_A(\langle \varphi \rangle) := \{\langle \varphi \rangle\}$ and $\text{sub}_A(\langle \alpha_1, \dots, \alpha_n \leftrightarrow \varphi \rangle) := \{\langle \alpha_1, \dots, \alpha_n \leftrightarrow \varphi \rangle\} \cup \text{sub}_A(\alpha_1) \cup \dots \cup \text{sub}_A(\alpha_n)$ where $\leftrightarrow \in \{\rightarrow, \Rightarrow\}$.

$\text{TopRule}(\alpha)$ returns the **top rule** of α , i.e. the last rule applied in the formation of α . It is defined as follows: $\text{TopRule}(\langle \varphi \rangle)$ is left undefined, $\text{TopRule}(\langle \alpha_1, \dots, \alpha_n \rightarrow \varphi \rangle) = \text{TopRule}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle) := ((\text{Conc}(\alpha_1), \dots, \text{Conc}(\alpha_n)), \varphi)$.

$\text{DefRule}(\alpha)$ returns the set of **defeasible rules** of α and it is defined as $\text{DefRule}(\langle \varphi \rangle) := \emptyset$, $\text{DefRule}(\langle \alpha_1, \dots, \alpha_n \rightarrow \varphi \rangle) := \text{DefRule}(\alpha_1) \cup \dots \cup \text{DefRule}(\alpha_n)$ and $\text{DefRule}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle) := \{((\text{Conc}(\alpha_1), \dots, \text{Conc}(\alpha_n)), \varphi)\} \cup \text{DefRule}(\alpha_1) \cup \dots \cup \text{DefRule}(\alpha_n)$.

Let us also define **semantic propositional negations**, for any $\varphi, \psi \in F$: $\varphi = \sim \psi$ abbreviates $\text{wellshap}(\langle \langle \varphi \rangle \rightarrow \neg \psi \rangle) \wedge \text{wellshap}(\langle \langle \psi \rangle \rightarrow \neg \varphi \rangle)$.

Let us now move to semantics. A **model** for \mathcal{L} is a tuple $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, n, || \cdot ||)$ where:

- $W \neq \emptyset$ is a set of *possible worlds*.
- $\mathcal{B} \subseteq W$ and $\mathcal{B} \neq \emptyset$ is the set of *worlds that are doxastically indistinguishable for the agent*.
- $\mathcal{O} \subseteq A$ is the set of *available arguments*, also called the *awareness set of the agent*.

- $\mathcal{D} \subseteq \text{SEQ}(\mathcal{F})$ is a set of *accepted defeasible rules*. Moreover, for every $((\varphi_1, \dots, \varphi_n), \varphi) \in \mathcal{D}$ we require that:
 - $\{\varphi_1, \dots, \varphi_n, \varphi\} \not\vdash_0 \perp$ (defeasible rules are consistent), where \vdash_0 denotes the consequence relation of classical propositional logic, and
 - $\{\varphi_1, \dots, \varphi_n\} \not\vdash_0 \varphi$ (defeasible rules are not deductively valid).
- $n : \text{SEQ}(\mathcal{F}) \rightarrow \text{At}$ is a (possibly partial) *naming function* for rules, where $n(R)$ informally means “the rule R is applicable”.
- $\|\cdot\|$ is and an *atomic valuation*, i.e. a function $\|\cdot\| : \text{At} \rightarrow \wp(W)$.

Interpretation. In a given model $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|)$, \mathcal{O} represents the set of arguments that the agent entertains or is aware of. Whenever $\alpha \in \mathcal{O}$, we assume that (i) the agent can determine her doxastic attitude toward the premisses of α through non-inferential methods (for instance, through observations), and (ii) she knows the structure of α (either because α has been communicated to her, or because she has gone through the cognitive process of building α). Besides this, there is not semantic intuition underlying \mathcal{O} , so the agent can be perfectly aware of rather silly arguments, as $\langle\langle p \rangle \rightarrow q \rangle$, without accepting them in any sense. Moreover, rules in the set \mathcal{D} are interpreted as rules whose inference strength lies in their content, rather than as purely formal schemas (as deductive rules are). As an example, consider the rule “Peter’s bike is on the bike parking area, therefore he should be in his office”. The term *accepted* means that the agent considers them applicable if there are not good reasons against doing so. Note that $\alpha^R \in \mathcal{O}$ does not imply $R \in \mathcal{D}$ (informally corresponding to the intuition that an agent can be aware of a defeasible argument without accepting its rule). There are further restrictions that could be arguably adopted, but that we leave out for the sake of simplicity. For instance, we could require \mathcal{O} to be closed under subarguments, or that for any accepted defeasible rule, the agent is aware of at least an argument using it.

Let $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|)$ be a model for $\mathcal{L} = (\mathcal{F}, \mathcal{A})$. The **set of well-shaped arguments** $WS^M \subseteq \mathcal{A}$ (depending on \mathcal{D} in M) is the smallest set fulfilling the following conditions:

1. $\langle \varphi \rangle \in WS^M$ for any $\varphi \in \mathcal{F}$.
2. $\langle \alpha_1, \dots, \alpha_n \rightarrow \varphi \rangle \in WS^M$ iff both $\alpha_i \in WS^M$ for every $1 \leq i \leq n$ and $\{\text{Conc}(\alpha_1), \dots, \text{Conc}(\alpha_n)\} \vdash_0 \varphi$.
3. $\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle \in WS^M$ iff both $\alpha_i \in WS^M$ for every $1 \leq i \leq n$ and $((\text{Conc}(\alpha_1), \dots, \text{Conc}(\alpha_n)), \varphi) \in \mathcal{D}$.

We drop the superscript M from WS^M whenever there is no danger of confusion.

Let (M, w) be a pointed model for \mathcal{L} , that is, $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|)$ is a model and $w \in W$. The **truth** relation, relating pointed models and formulas, is given by:²

$M, w \models \Box \varphi$	iff	for all $w' \in W$: $w' \in \mathcal{B}$ implies $M, w' \models \varphi$.
$M, w \models \text{aware}(\alpha)$	iff	$\alpha \in \mathcal{O}$.
$M, w \models \text{conc}(\alpha) = \varphi$	iff	$\text{Conc}(\alpha) = \varphi$.
$M, w \models \text{strict}(\alpha)$	iff	$\text{DefRule}(\alpha) = \emptyset$.
$M, w \models \text{undercuts}(\alpha, \beta)$	iff	$\beta = \langle \beta_1, \dots, \beta_n \Rightarrow \psi \rangle$ and $\text{Conc}(\alpha) = \neg n(\text{TopRule}(\beta))$.
$M, w \models \text{wellshap}(\alpha)$	iff	$\alpha \in WS^M$.

²Note that we do not need to consider undercuts as a primitive operator, since it could be defined through a (simpler) operator that captures the meaning of n . We make this choice for the sake of succinctness, as well as for studying the axiomatic behaviour of undercuts.

A formula φ is said to be **valid** (noted $\models \varphi$) iff it is true at all pointed models. We use $\|\varphi\|_M$ to denote the **truth-set** of φ , i.e., the set of worlds of M where φ is true, and \mathcal{M} to denote the class of all models.

We now present a sound and complete axiomatisation of \mathcal{L} w.r.t. \mathcal{M} , a topic that was left out in [15], and that constitutes one of the main technical contributions of the current paper. Although our models provide a compact representation of the needed components for reasoning about basic and argument-based beliefs in a single-agent context, they are rather non-standard from a technical point of view. Besides the strongly syntactic character of some of their elements, their modal components are not defined as usual, therefore the definition of its canonical model cannot be extrapolated straightforwardly. Nevertheless, we can provide an indirect completeness proof (see Appendix A1 for details).

Theorem 1. *The axiom system \mathbb{L}^{BA} , defined in Table 1, is sound and complete for \mathcal{L} w.r.t. \mathcal{M} .*

3 Basic beliefs and argument-based beliefs

The logic introduced above can be used to study a rich repertoire of doxastic attitudes. We start by discussing *basic beliefs*, informally representing those that are not grounded on inferential processes. As mentioned, they can also be understood in terms of *intuitive beliefs*, i.e., those that the agent extracts from a sort of data-base, seen by her as completely trustworthy [37]. As usual in awareness epistemic logic, we have two versions of this notion. On the one hand, we have the implicit (ideal) version of basic beliefs, modelled through $\Box\varphi$, that suffers from the extensively discussed problem of logical omniscience (see e.g. [21, Chapter 9]). On the other hand, we have its explicit counterpart, for which we have chosen $\Box^e\varphi := \Box\varphi \wedge \text{aware}(\langle\varphi\rangle)$. Note that, like in other logics for implicit and explicit belief, it holds that $\models \Box^e\varphi \rightarrow \Box\varphi$. Moreover, under the current semantics, $\Box^e\varphi$ is equivalent to a schema that resembles another usual option for modelling explicit beliefs (e.g. [39]): $\models \Box^e\varphi \leftrightarrow \Box(\varphi \wedge \text{aware}(\langle\varphi\rangle))$.

Besides basic beliefs, we can also capture in \mathcal{L} a sort of deductive-explicit belief. Deductive-explicit beliefs are those rooted in a deductive argument s.t. the agent has a basic belief that all its premisses are true. Formally, and following [6], we define **doxastic argument acceptance** as

$$\text{accept}(\alpha) := \bigwedge_{\varphi \in \text{Prem}(\alpha)} \Box\varphi,$$

and **deductive-explicit belief** as

$$\text{B}^{\text{D}}(\alpha, \varphi) := \text{accept}(\alpha) \wedge \text{aware}(\alpha) \wedge \text{conc}(\alpha) = \varphi \wedge \text{strict}(\alpha) \wedge \text{wellshap}(\alpha).$$

Note that $\models \text{B}^{\text{D}}(\alpha, \varphi) \rightarrow \Box\varphi$ and $\models \Box^e\varphi \leftrightarrow \text{B}^{\text{D}}(\langle\varphi\rangle, \varphi)$. The first validity shows that deductive-explicit beliefs are a subset of basic-implicit beliefs. The second one shows that basic-explicit beliefs are an extreme case of deductive-explicit beliefs (those that are rooted in the trivial deduction that goes from φ to φ , i.e., in the atomic argument $\langle\varphi\rangle$).

Up to now, we have not gone far from the kind of attitudes that are usually discussed in the awareness logic literature (e.g. in [20, 13, 24, 25]). We now take a small detour through argumentation theory in order to define argument-based beliefs. Roughly speaking, argument-based beliefs are grounded in arguments that may involve non-deductive steps. They can be understood, at least to some extent, in terms of the *reflective beliefs* of [37]. Recall that we are after formalising the principle P2 presented in the introduction: the beliefs of a rational agent should be grounded in good arguments. But, what does it mean *good* in this context? Following [9], the very notion of argument strength can be analysed in three different layers or dimensions: the *support dimension* (how strong is the reason given by an argument to accept its conclusion), the *dialectic dimension* (how arguments attack and defeat each other), and the *evaluative dimension* (how the former conflicts are to be solved).

Modal core axioms	
(Ax0)	All propositional tautologies
(Ax1)	$KD45$ axioms for \Box
Introspection axioms	
(Ax2)	$\text{aware}(\alpha) \rightarrow \Box \text{aware}(\alpha)$
(Ax3)	$\neg \text{aware}(\alpha) \rightarrow \Box \neg \text{aware}(\alpha)$
(Ax4)	$\text{wellshap}(\alpha) \rightarrow \Box \text{wellshap}(\alpha)$
(Ax5)	$\neg \text{wellshap}(\alpha) \rightarrow \Box \neg \text{wellshap}(\alpha)$
(Ax6)	$\text{undercuts}(\alpha, \beta) \rightarrow \Box \text{undercuts}(\alpha, \beta)$
(Ax7)	$\neg \text{undercuts}(\alpha, \beta) \rightarrow \Box \neg \text{undercuts}(\alpha, \beta)$
Axioms for syntactic operators	
(Ax8)	$\text{conc}(\alpha) = \varphi$ whenever $\text{Conc}(\alpha) = \varphi$
(Ax9)	$\neg \text{conc}(\alpha) = \varphi$ whenever $\text{Conc}(\alpha) \neq \varphi$
(Ax10)	$\text{strict}(\alpha)$ whenever $\text{DefRule}(\alpha) = \emptyset$
(Ax11)	$\neg \text{strict}(\alpha)$ whenever $\text{DefRule}(\alpha) \neq \emptyset$
Wellshapedness axioms	
(Ax12)	$\text{wellshap}(\langle \varphi \rangle)$
(Ax13)	$\text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle) \rightarrow \bigwedge_{1 \leq i \leq n} \text{wellshap}(\alpha_i)$
(Ax14)	$\bigwedge_{1 \leq i \leq n} \text{wellshap}(\alpha_i) \rightarrow \text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle)$ whenever $\{\text{Conc}(\alpha_i) \mid 1 \leq i \leq n\} \vdash_0 \varphi$
(Ax15)	$\neg \text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle)$ whenever $\{\text{Conc}(\alpha_i) \mid 1 \leq i \leq n\} \not\vdash_0 \varphi$
(Ax16)	$\left(\bigwedge_{1 \leq i \leq n} \text{wellshap}(\alpha_i) \wedge \text{wellshap}(\langle \langle \text{Conc}(\alpha_1) \rangle, \dots, \langle \text{Conc}(\alpha_n) \rangle \Rightarrow \varphi \rangle) \right)$ $\leftrightarrow \text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle)$
(Ax17)	$(\text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle) \rightarrow \neg \text{wellshap}(\langle \langle \text{Conc}(\alpha_1) \rangle, \dots, \langle \text{Conc}(\alpha_n) \rangle, \langle \varphi \rangle \Rightarrow \perp \rangle))$
(Ax18)	$(\text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle) \rightarrow \neg \text{wellshap}(\langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle))$
Undercut axioms	
(Ax19)	$\text{undercuts}(\langle \neg p \rangle, \alpha^R) \rightarrow \neg \text{undercuts}(\langle \neg q \rangle, \alpha^R)$ whenever $q \neq p$
(Ax20)	$\neg \text{undercuts}(\alpha, \langle \varphi \rangle)$
(Ax21)	$\neg \text{undercuts}(\alpha, \langle \alpha_1, \dots, \alpha_n \Rightarrow \varphi \rangle)$
(Ax22)	$\neg \text{undercuts}(\alpha, \beta)$ whenever $\text{Conc}(\alpha) \neq \neg p$ for some $p \in \text{At}$
(Ax23)	$(\text{undercuts}(\langle \neg p \rangle, \langle \langle \text{Conc}(\beta_1) \rangle, \dots, \langle \text{Conc}(\beta_n) \rangle \Rightarrow \varphi \rangle) \wedge \text{conc}(\alpha) = \neg p) \rightarrow$ $\text{undercuts}(\alpha, \langle \beta_1, \dots, \beta_n \Rightarrow \varphi \rangle)$
(Ax24)	$(\text{undercuts}(\alpha, \langle \beta_1, \dots, \beta_n \Rightarrow \varphi \rangle) \wedge \text{conc}(\alpha) = \neg p) \rightarrow$ $\text{undercuts}(\langle \neg p \rangle, \langle \langle \text{Conc}(\beta_1) \rangle, \dots, \langle \text{Conc}(\beta_n) \rangle \Rightarrow \varphi \rangle)$
Rules	
(MP)	From $\varphi \rightarrow \psi$ and φ infer ψ
(Nec)	From φ , infer $\Box \varphi$

Table 1: Axiom system.

Hence, the first step is to set up a notion of argument strength regarding the *support dimension*. Formally, we seek to define a preference relation among the arguments of \mathcal{L} that takes into account P1 (arguments with believed premisses are to be preferred over arguments with premisses that are not believed). In [15], we showed how to do this by splitting all arguments in three preference classes that

were based on the basic doxastic attitude of the agent toward the premisses of the arguments. Here, we take a much simpler view, for the sake of brevity, and directly exclude arguments whose premisses are not believed. Both options makes P2 dependent on P1, since in the process of grounding arguments in beliefs and these in turn in new arguments, we arrive at good arguments that are good just because the agent has a basic belief that all their premisses hold. However, inference links must still play a role when determining the relative strength of two arguments, the simplest principle that can be adopted in this regard is captured by the following binary relation among arguments of \mathcal{L} : $\alpha \geq \beta := \text{strict}(\alpha) \vee \neg \text{strict}(\beta)$. This relation informally corresponds to the idea that, *ceteris paribus*, deductive arguments are to be preferred to non-deductive ones.

Regarding the *dialectic dimension* of argument strength, we capture two forms of argumentative defeat, namely, *undercutting* (attacking a defeasible inference step of any subargument) and *successful rebuttal* (attacking the conclusion of a less or equally preferred subargument). Formally,

- **Undercutting a subargument** undercuts^{*} $(\alpha, \beta) := \bigvee_{\beta' \in \text{sub}_A(\beta)} \text{undercuts}(\alpha, \beta')$.
- **Unrestricted successful rebuttal**
Urebutts $(\alpha, \beta) := \neg \text{strict}(\beta) \wedge \bigvee_{\beta' \in \text{sub}_A(\beta)} (\text{conc}(\alpha) = \varphi \wedge \text{conc}(\beta') = \psi \wedge \varphi = \sim \psi \wedge \alpha \geq \beta')$.
- **Defeat** defeat $(\alpha, \beta) := \text{undercuts}^*(\alpha, \beta) \vee \text{Urebutts}(\alpha, \beta)$.

As discussed in the formal argumentation literature, there is a more restrictive alternative for the notion of rebuttal, requiring the top rule of the attacked subargument to be defeasible³

- **Restricted successful rebuttal**
Rrebutts $(\alpha, \beta) := \neg \text{strict}(\beta) \wedge \bigvee_{\langle \beta_1, \dots, \beta_n \Rightarrow \varphi \rangle \in \text{sub}_A(\beta)} (\text{conc}(\alpha) = \psi \wedge \varphi = \sim \psi)$.

Argumentation frameworks and their semantics [18] are the most studied tool to capture the *evaluative dimension* of argument strength. We now explain how to incorporate them in the current approach. Let (M, w) be a pointed model for $\mathcal{L} = (F, A)$, we define its **associated argumentation framework** as $AF^M := (A^M, \rightsquigarrow)$, where $A^M := \{\alpha \in A \mid M, w \models \text{aware}(\alpha) \wedge \text{wellshap}(\alpha) \wedge \text{accept}(\alpha)\}$ and $\rightsquigarrow \subseteq A^M \times A^M$ is given by $\alpha \rightsquigarrow \beta$ iff $M, w \models \text{defeat}(\alpha, \beta)$. We stress the fact that in the domain of our frameworks (i.e., in A^M), basic beliefs act as a filter (in the clause $\text{accept}(\alpha)$), instantiating a qualified, unproblematic version of P1, namely P1': *basic beliefs* are an input for argument evaluation. Given a set of possibly conflicting arguments (an argumentation framework), we need a mechanism for the agent to decide which of the arguments are to be selected (an argumentation semantics). We say that a set $B \subseteq A^M$ is **conflict-free** iff there are no $\alpha, \beta \in B$ s.t. $\alpha \rightsquigarrow \beta$. Moreover, we say that $B \subseteq A^M$ **defends** $\alpha \in A^M$ iff for every $\gamma \in A^M$, $\gamma \rightsquigarrow \alpha$ implies that there is $\beta \in B$ s.t. $\beta \rightsquigarrow \gamma$. We say that $B \subseteq A^M$ is a **complete extension** iff it is conflict-free and it contains precisely the elements of A^M that it defends. We say that $B \subseteq A^M$ is the **grounded extension** of $AF^M = (A^M, \rightsquigarrow)$ iff it is the smallest (w.r.t. set inclusion) complete extension. We use $GE(AF^M)$ to denote the grounded extension of AF^M . As it is well-known, the grounded extension of an argumentation framework always exists and it is moreover unique [18]. The unfamiliar reader is referred to [8] for an extensive discussion on argumentation semantics.

Finally, we use the grounded extension to define the **argument-based beliefs** of the agent. First, let us extend $\mathcal{L} = (F, A)$ to $\mathcal{L}^{\text{AB}} = (F^{\text{AB}}, A)$ by adding a new kind of formulas $B(\alpha, \varphi)$ where $\alpha \in A$ and $\varphi \in F$. $B(\alpha, \varphi)$ means that the agent believes that φ based on argument α . We interpret the new language in the same class of models, by adding the truth clause:

$$M, w \models B(\alpha, \varphi) \quad \text{iff} \quad \alpha \in GE(AF^M) \quad \text{and} \quad \text{Conc}(\alpha) = \varphi.$$

³See [41] for a discussion about the two possible design choices. Note moreover that the other customary type of attack, i.e. *undermining* (attacking a premise), makes sense only when non-believed premisses are taken into account.

Note that $\models \text{B}^{\text{D}}(\alpha, \varphi) \rightarrow \text{B}(\alpha, \varphi)$ and $\models \Box^e \varphi \rightarrow \text{B}(\langle \varphi \rangle, \varphi)$.

We close this section by analysing our notion of argument-based belief under the view of [16]’s *rationality postulates*. In a nutshell, if no restrictions are imposed, our agent behaves according to a kind of *minimal rationality* (i.e. she does not explicitly believe in inconsistencies). If, however, we add some ideal assumptions, then she satisfies all [16]’s postulates.

Proposition 1. *Let (M, w) be a pointed model for $\mathcal{L} = (\text{F}, \text{A})$, where $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, \mathbf{n}, \|\cdot\|)$. Let AF^M be its associated argumentation framework, then:*

- *AF^M satisfies direct consistency, that is, there are no $\alpha, \beta \in \text{A}$ and $\varphi, \psi \in \text{F}$ s.t. $M, w \models \text{B}(\alpha, \varphi) \wedge \text{B}(\beta, \psi) \wedge \varphi \approx \sim \psi$.*
- *If restricted rebuttal is assumed and $\mathcal{O} = \text{A}$, then AF^M satisfies direct consistency; indirect consistency (that is, $\text{Conc}(GE(AF^M)) \not\vdash_0 \perp$); sub-argument closure (that is, $\alpha \in GE(AF^M)$ implies $\text{sub}_{\text{A}}(\alpha) \subseteq GE(AF^M)$); and strict closure (that is, $\text{Conc}(GE(AF^M)) \vdash_0 \varphi$ implies $\varphi \in \text{Conc}(GE(AF^M))$).*

Proof (sketched). For the first item, we suppose the contrary, that is, that there are arguments $\alpha, \beta \in GE(AF^M)$ s.t. $\text{Conc}(\alpha) = \varphi$, $\text{Conc}(\beta) = \psi$ and φ is propositionally equivalent to the negation of ψ . Then, we continue by cases on the shape of α and β (each of them can be either an atomic argument, or an argument whose last inference step is deductive (resp. defeasible)). From the nine different cases, three of them are redundant. From the six remaining cases, it is easy to arrive to $\alpha \rightsquigarrow \beta$ or $\beta \rightsquigarrow \alpha$ (which contradicts the assumption that they are in the grounded extension, because it is conflict-free).

For the second item, it suffices to show that under both assumptions (adopting the definition of restricted rebuttal and assuming $\mathcal{O} = \text{A}$), we are just working with an instance of *well-defined ASPIC+* frameworks (one constructed over a knowledge base where the set of ordinary premisses is empty), which is guaranteed to satisfy all [16]’s rationality postulates (see [31, Section 3.3] for details). \square

4 Dynamics of information

The current framework can throw some light on the relations between dynamics of information, argumentation and doxastic attitudes. We can distinguish several kinds of actions, that have different potential effects on basic and argument-based beliefs. The framework naturally allows for the use of tools imported from *dynamic epistemic logic* (DEL) [17]. In particular, we can describe these actions using dynamic modalities, for which complete axiomatisations can be then provided by finding a full list of *reduction axioms* [27, 17, 40]. In order to do so, one first need to show that the rule of replacement of proved equivalents is sound (it preserves validity) in the extended language (see [27] for details). Although this is *not* the case in \mathcal{L} , as it happens with other languages containing awareness operators [20, 24], we can restrict the domain of application of the rule, and it still does the job for axiomatizing certain dynamic extensions. More precisely, we will work with the rule:

$$\text{(RE)} \text{ From } \varphi \leftrightarrow \psi, \text{ infer } \delta \leftrightarrow \delta[\varphi/\psi],$$

with $\delta[\varphi/\psi]$ the result of replacing one or more non- \star occurrences of ψ in δ by φ .⁴ Semantically, this amounts to showing that each of the actions act we are about to discuss is well defined, in the sense that whenever we compute M^{act} (the result of executing action act in model M), we stay in the

⁴A non- \star of ψ in δ is just an occurrence of ψ in δ where ψ is not inside the scope of $\star \in \{\text{aware}, \text{conc}, \text{wellshap}, \text{undercuts}\}$. Note that we assume that φ is inside the scope of conc in the formula $\text{conc}(\alpha) = \varphi$.

intended class of models. When this does not happen, as it is the case with many DEL actions (e.g. public announcements [17]), one needs to find a set of *preconditions* for the action. Preconditions works as sufficient conditions for the action to be “safe” i.e., to secure that after executing it, we stay in the intended class of models.

Let us start by defining four different actions. Let $\mathcal{L} = (F, A)$ be given, let $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|)$ be an \mathcal{L} -model, let $\alpha \in A$, let $R \in \text{SEQ}(F)$, and let $\varphi \in F$. We define:

- The act of **acquiring argument** α (resp. **forgetting argument** α) produces the model $M^{\alpha+!} := (W, \mathcal{B}, \mathcal{O}^{\alpha+!}, \mathcal{D}, n, \|\cdot\|)$, where $\mathcal{O}^{\alpha+!} := \mathcal{O} \cup \{\alpha\}$ (resp. $M^{\alpha-!} := (W, \mathcal{B}, \mathcal{O}^{\alpha-!}, \mathcal{D}, n, \|\cdot\|)$, where $\mathcal{O}^{\alpha-!} := \mathcal{O} \setminus \{\alpha\}$).
- The act of **accepting the defeasible rule** R produces the model $M^{R+!} := (W, \mathcal{B}, \mathcal{O}, \mathcal{D}^{R+!}, n, \|\cdot\|)$, where $\mathcal{D}^{R+!} := \mathcal{D} \cup \{R\}$.
- The act of **publicly announcing** φ produces the model $M^{\varphi!} := (W^{\varphi!}, \mathcal{B}^{\varphi!}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|^{\varphi!})$, where $W^{\varphi!} := W \cap \|\varphi\|_M$; $\mathcal{B}^{\varphi!} := \mathcal{B} \cap \|\varphi\|_M$; and $\|p\|_M^{\varphi!} := \|p\|_M \cap \|\varphi\|_M$ for every atom p .

Interpretation. Note that the definition is far from being exhaustive, we analyse them because they are natural adaptations of other actions studied in the literature [24, 17]. The most basic argumentative change we can think of consists in adding an argument into the awareness of the agent. Informally, this can be thought as the result of a communicative event (e.g. an opponent advancing an argument), learning (the agent reading an argument in a book), or as the result of reflection (the own agent constructing an argument). Formally, the action is a direct generalization of the “consider” action defined for sentences in [24, 13]. Its straightforward counterpart is the act of forgetting an argument (i.e. dropping it from the awareness of the agent). As for the action $(\cdot)^{R+!}$, defeasible rules can also be learnt in different ways. For instance, an agent can learn the rule $((\text{Bird}), \text{Flies})$ because an ornithologist told her, because she observed repeatedly that birds fly, or because she read it in a textbook. Finally, public announcements are probably the most studied action in DEL (see e.g. [17, Chapter 4]). This kind of announcements are supposed to be truthful and coming from a completely reliable source.

We now define a **dynamic language**, in order to talk about the different actions. Let $\mathcal{L} = (F, A)$ be a language, formulas of the extended language $\mathcal{L}^! = (F^!, A)$ are given by:

$$\varphi ::= \psi \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid [\alpha+!]\varphi \mid [\alpha-!]\varphi \mid [R+!]\varphi \mid [\psi!]\varphi \quad \psi \in F, \alpha \in A, R \in \text{SEQ}(F).$$

Let $[\text{act}]$ be any of the dynamic modalities we have just defined, we use $\langle \text{act} \rangle$ as an abbreviation of $\neg[\text{act}]\neg$, with $\langle \text{act} \rangle \varphi$ informally meaning that action act can be executed and after executing it, φ holds.

Note that the class of all models \mathcal{M} is *not* closed under all defined actions. In particular, it is not closed under $(\cdot)^{R+!}$ nor under $(\cdot)^{\varphi!}$. For the former, the reason is that only rules that are consistent and non-deductive can be learnt as defeasible (see the definition of model in Section 2). For the latter, only truthful formulas that do not trivialize the beliefs of the agent (in the sense of making \mathcal{B} empty), can be announced. This inconvenience is solved by fixing **preconditions** (expressible in \mathcal{L}) for both actions.

Let $R = ((\varphi_1, \dots, \varphi_n), \varphi) \in \text{SEQ}(F)$ and $\varphi \in F$, we define:

$$\text{pre}(R) := \neg \text{wellshap}(\langle \langle \varphi_1 \rangle, \dots, \langle \varphi_n \rangle, \langle \varphi \rangle \rightarrow \perp \rangle) \wedge \neg \text{wellshap}(\langle \langle \varphi_1 \rangle, \dots, \langle \varphi_n \rangle \rightarrow \varphi \rangle);$$

$$\text{pre}(\varphi!) := \varphi \wedge \diamond \varphi.$$

It is almost immediate to check that, for any pointed model (M, w) , any $R = ((\varphi_1, \dots, \varphi_n), \varphi) \in \text{SEQ}(F)$, and any $\varphi \in F$ we have that:

$$\begin{aligned} & \{\langle \varphi_1, \dots, \varphi_n, \varphi \rangle \not\prec_0 \perp \text{ and } \{\langle \varphi_1, \dots, \varphi_n \rangle \not\prec_0 \varphi\} \text{ iff } M, w \models \text{pre}(R); \text{ and} \\ & (w \in \|\varphi\|_M \text{ and } \|\varphi\|_M \cap \mathcal{B} \neq \emptyset) \text{ iff } M, w \models \text{pre}(\varphi!). \end{aligned}$$

Moreover, note that $M, w \models \text{pre}(R)$ iff $M, u \models \text{pre}(R)$ for every $u \in W$. Let (M, w) be a pointed model with $M = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, \mathbf{n}, \|\cdot\|)$, we define the truth clause for the new kind of formulas:

$$\begin{aligned} M, w \models [\sigma+!]\varphi & \text{ iff } M^{\sigma+!}, w \models \varphi, \\ M, w \models [\sigma-!]\varphi & \text{ iff } M^{\sigma-!}, w \models \varphi, \\ M, w \models [R+!]\varphi & \text{ iff } M, w \models \text{pre}(R), \text{ implies } M^{R+!}, w \models \varphi, \\ M, w \models [\varphi!]\psi & \text{ iff } M, w \models \text{pre}(\varphi!) \text{ implies } M^{\varphi!}, w \models \psi. \end{aligned}$$

Finally, we establish a completeness result for $\mathcal{L}^!$ w.r.t. \mathcal{M} . Note that in Table 2, \pm denotes an arbitrary element of $\{+, -\}$.

Proposition 2. *The proof system $\mathsf{L}_{\text{BA}}^!$ that extends the one of Table 1 with all axioms of Table 2 and it is closed under (RE) is sound and complete for $\mathcal{L}^!$ w.r.t. \mathcal{M} .*

Proof. Soundness follows from the validity of all axioms and the validity-preserving character of (RE) in the extended language. Completeness follows from the usual reduction argument. In short, note that in the right-hand side of all axioms of Table 2, either the dynamic operator disappears or it is applied to a less complex formula than in the left-hand side. In the case of reduction axioms for $[R!+]\text{wellshap}(\alpha)$, either there are no dynamic modalities occurring in the right-hand side of the equivalence or they are applied to wellshap-formulas with less complex *arguments* than in the right-hand side. Therefore, we can define a meaning-preserving translation from $\mathsf{F}^!$ to F that, together with Theorem 1, provides the desired result. The validity-preserving character of (RE) in the extended language w.r.t. \mathcal{M} takes care of formulas with nested dynamic modalities. The reader is referred to [27] for details. \square

We close this section by modelling a toy example, inspired by [38], and illustrating how actions affect argument-based beliefs. Suppose that an agent is wondering whether another agent, Harry, is a British subject (br). Suppose that the only basic-explicit belief she holds at the beginning is that *Harry was born in Bermuda* (be). Other pieces of relevant information are: *Harry's parents are aliens* (a), and that *the rule "If Harry is born in Bermuda, then he is presumably a British subject" is applicable* (r1). Let $R_1 = ((\text{be}), \text{br})$. We start with the model $M_0 = (W, \mathcal{B}, \mathcal{O}, \mathcal{D}, \mathbf{n}, \|\cdot\|)$, where $W = \mathcal{B} = \{w_0, w_1, w_2, w_3\}$, $\mathcal{D} = \emptyset$, $\mathcal{O} = \{\langle \text{be} \rangle\}$, $\mathbf{n}(R_1) = r1$, $\|\text{be}\| = W$, $\|\text{br}\| = \|r1\| = \{w_0, w_2\}$, and $\|a\| = \{w_0, w_1\}$. It is then easy to check that $M_0, w_0 \models \Box^e \text{be}$. Moreover, we have that $M_0, w_0 \models [R1+!][\alpha^{R1+!}]B(\alpha^{R1}, \text{br})$. In words, after learning the rule R_1 and becoming aware of the simplest argument using it, i.e. $\langle \langle \text{be} \rangle \Rightarrow \text{br} \rangle$, the agent has an argument-based belief that Harry is a British subject. If, however the agent learns subsequently from a completely trustworthy source that Harry's parents are alien (a), together with the rule $R_2 = ((a), \neg r1)$, and the argument $\langle \langle a \rangle \Rightarrow \neg r1 \rangle$, then she revises her argument-based belief about Harry's nationality. In symbols, $M_0, w_0 \models [R1+!][\alpha^{R1+!}][a!][R2+!][\alpha^{R2+!}]\neg B(\alpha^{R1}, \text{br})$.

5 Concluding remarks

Closely related work. From all the works we have commented throughout the paper, it seems that [24, 25] and [36] are the closest one to our approach. Regarding [24, 25], we have somehow generalize their *awareness of rules* to our awareness of arguments (abstracting away from other forms of awareness treated there). As for [36], their choice of modelling arguments semantically (as opens of a topology), permits a transparent axiomatisation of their notion of argument-based beliefs, which is easily guaranteed to be consistent (two of the weaknesses of our approach). On the other hand, we naturally treat arguments as first-class citizens in our language, and the argument-based beliefs of our agent escape from every form of logical omniscience (while the beliefs of [36]'s agent are still closed under equivalent formulas).

$[\alpha\pm!]p \leftrightarrow p$	$[\varphi!]p \leftrightarrow (\text{pre}(\varphi!) \rightarrow p)$
$[\alpha\pm!]\neg\varphi \leftrightarrow \neg[\alpha\pm!]\varphi$	$[\varphi!]\neg\psi \leftrightarrow (\text{pre}(\varphi!) \rightarrow \neg[\varphi!]\psi)$
$[\alpha\pm!](\varphi \wedge \psi) \leftrightarrow ([\alpha\pm!]\varphi \wedge [\alpha\pm!]\psi)$	$[\varphi!](\delta \wedge \psi) \leftrightarrow ([\varphi!]\delta \wedge [\varphi!]\psi)$
$[\alpha\pm!]\Box\varphi \leftrightarrow \Box[\alpha\pm!]\varphi$	$[\varphi!]\Box\psi \leftrightarrow (\text{pre}(\varphi!) \rightarrow \Box[\varphi!]\psi)$
$[\alpha\pm!] \text{aware}(\beta) \leftrightarrow \text{aware}(\beta) \text{ for } \alpha \neq \beta$	$[\varphi!] \text{aware}(\beta) \leftrightarrow (\text{pre}(\varphi!) \rightarrow \text{aware}(\beta))$
$[\alpha+!] \text{aware}(\alpha) \leftrightarrow \top$	
$[\alpha-!] \text{aware}(\alpha) \leftrightarrow \perp$	
$[\alpha\pm!] \text{conc}(\beta) = \varphi \leftrightarrow \text{conc}(\beta) = \varphi$	$[\varphi!] \text{conc}(\beta) = \psi \leftrightarrow (\text{pre}(\varphi!) \rightarrow \text{conc}(\beta) = \psi)$
$[\alpha\pm!] \text{strict}(\beta) \leftrightarrow \text{strict}(\beta)$	$[\varphi!] \text{strict}(\beta) \leftrightarrow (\text{pre}(\varphi!) \rightarrow \text{strict}(\beta))$
$[\alpha\pm!] \text{undercuts}(\beta, \gamma) \leftrightarrow \text{undercuts}(\beta, \gamma)$	$[\varphi!] \text{undercuts}(\beta, \gamma) \leftrightarrow (\text{pre}(\varphi!) \rightarrow \text{undercuts}(\beta, \gamma))$
$[\alpha\pm!] \text{wellshap}(\beta) \leftrightarrow \text{wellshap}(\beta)$	$[\varphi!] \text{wellshap}(\beta) \leftrightarrow (\text{pre}(\varphi!) \rightarrow \text{wellshap}(\beta))$
$[R+!]p \leftrightarrow (\text{pre}(R) \rightarrow p)$	$[R+!] \text{aware}(\alpha) \leftrightarrow (\text{pre}(R) \rightarrow \text{aware}(\alpha))$
$[R+!]\neg\varphi \leftrightarrow (\text{pre}(R) \rightarrow \neg[R+!]\varphi)$	$[R+!] \text{conc}(\alpha) = \varphi \leftrightarrow (\text{pre}(R) \rightarrow \text{conc}(\alpha) = \varphi)$
$[R+!](\varphi \wedge \psi) \leftrightarrow ([R+!]\varphi \wedge [R+!]\psi)$	$[R+!] \text{strict}(\alpha) \leftrightarrow (\text{pre}(R) \rightarrow \text{strict}(\alpha))$
$[R+!]\Box\varphi \leftrightarrow \Box[R+!]\varphi$	$[R+!] \text{undercuts}(\alpha, \beta) \leftrightarrow (\text{pre}(R) \rightarrow \text{undercuts}(\alpha, \beta))$
$[R+!] \text{wellshap}(\langle\varphi\rangle) \leftrightarrow \top$	
$[R+!] \text{wellshap}(\langle\langle\varphi_1\rangle, \dots, \langle\varphi_n\rangle \twoheadrightarrow \varphi\rangle) \leftrightarrow (\text{pre}(R) \rightarrow \text{wellshap}(\langle\langle\varphi_1\rangle, \dots, \langle\varphi_n\rangle \twoheadrightarrow \varphi\rangle))$	
$[R+!] \text{wellshap}(\alpha^R) \leftrightarrow \top$	
$[R+!] \text{wellshap}(\alpha^{R'}) \leftrightarrow (\text{pre}(R) \rightarrow \text{wellshap}(\alpha^{R'})) \quad \text{whenever } R \neq R'$	
$[R+!] \text{wellshap}(\langle\alpha_1, \dots, \alpha_n \leftrightarrow \varphi\rangle) \leftrightarrow$ $\leftrightarrow (\text{pre}(R) \rightarrow (\bigwedge_{1 \leq i \leq n} [R+!] \text{wellshap}(\alpha_i) \wedge [R+!] \text{wellshap}(\langle\langle \text{Conc}(\alpha_1) \rangle, \dots, \langle \text{Conc}(\alpha_n) \rangle \leftrightarrow \varphi\rangle)))$	

Table 2: Reduction axioms for $\mathcal{L}^!$.

Future work. There are natural open paths for future work. An urgent task in the development of the logical aspects of the framework consists in axiomatizing (if possible) the argument-based belief operator $B(\cdot, \cdot)$. Moreover, the modal semantic apparatus of our models could be extended to plausibility structures [7], so as to model fine-grained preference between arguments, based on the agent's basic epistemic attitudes toward the premisses of the involved arguments (e.g. known premisses are to be preferred to strongly believed premisses, and the latter, in turn, are to be preferred to merely believed premisses). Finally, a multi-agent extension of the current framework could be used to model argument exchange in different kinds of scenarios (e.g. deliberation, persuasion dialogues or inquiry).

References

- [1] Sergei Artemov (2018): *Justification Awareness Models*. In Sergei Artemov & Anil Nerode, editors: *Logical Foundations of Computer Science, LNCS 10703*, Springer, pp. 22–36, doi:10.1007/978-3-319-72056-2_2.
- [2] Sergei Artemov & Melvin Fitting (2016): *Justification Logic*. In Edward N. Zalta, editor: *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.
- [3] Sergei Artemov & Elena Nogina (2005): *Introducing justification into epistemic logic*. *Journal of Logic and Computation* 15(6), pp. 1059–1073, doi:10.1093/logcom/exi053.
- [4] Sergei N Artemov (2012): *The ontology of justifications in the logical setting*. *Studia Logica* 100(1-2), pp. 17–30, doi:10.1007/s11225-012-9387-x.

- [5] Alexandru Baltag, Nick Bezhanishvili, Aybuke Özgün & Sonja Smets (2016): *Justified Belief and the Topology of Evidence*. In Jouko Väänänen, Åsa Hirvonen & Ruy de Queiroz, editors: *Logic, Language, Information, and Computation, LNCS 9803*, Springer, pp. 83–103, doi:10.1007/978-3-662-52921-8`6.
- [6] Alexandru Baltag, Bryan Renne & Sonja Smets (2012): *The Logic of Justified Belief Change, Soft Evidence and Defeasible Knowledge*. In Luke Ong & Ruy de Queiroz, editors: *Logic, Language, Information and Computation. WoLLIC 2012.*, LNCS 7456, Springer, pp. 168–190, doi:10.1007/978-3-642-32621-9`13.
- [7] Alexandru Baltag & Sonja Smets (2008): *A qualitative theory of dynamic interactive belief revision*. In Wiebe van der Hoek, Giacomo Bonanno & Michael Wooldridge, editors: *Logic and the foundations of game and decision theory (LOFT 7)*, Texts in Logic and Games 3, Amsterdam University Press, pp. 9–58.
- [8] Pietro Baroni, Martin Caminada & Massimiliano Giacomin (2018): *Abstract argumentation frameworks and their semantics*. In Pietro Baroni, Dov M. Gabbay, Massimiliano Giacomin & Leendert van der Torre, editors: *Handbook of formal argumentation*, College Publications, pp. 159–236.
- [9] Mathieu Beirlaen, Jesse Heyninck, Pere Pardo & Christian Straßer (2018): *Argument strength in formal argumentation*. *IfCoLog Journal of Logics and their Applications* 5(3), pp. 629–675.
- [10] Trevor JM Bench-Capon & Paul E Dunne (2007): *Argumentation in artificial intelligence*. *Artificial intelligence* 171(10-15), pp. 619–641, doi:10.1016/j.artint.2007.05.001.
- [11] Johan van Benthem, David Fernández-Duque & Eric Pacuit (2014): *Evidence and plausibility in neighborhood structures*. *Annals of Pure and Applied Logic* 165(1), pp. 106–133, doi:10.1016/j.apal.2013.07.007.
- [12] Johan van Benthem, David Fernández-Duque, Eric Pacuit et al. (2012): *Evidence Logic: A New Look at Neighborhood Structures*. *Advances in modal logic* 9, pp. 97–118.
- [13] Johan van Benthem & Fernando R Velázquez-Quesada (2010): *The dynamics of awareness*. *Synthese* 177(1), pp. 5–27, doi:10.1007/s11229-010-9764-9.
- [14] Patrick Blackburn, Maarten De Rijke & Yde Venema (2010): *Modal Logic*. Cambridge University Press, doi:10.1017/CBO9781107050884.
- [15] Alfredo Burrieza & Antonio Yuste-Ginel (2020): *Basic beliefs and argument-based beliefs in awareness epistemic logic with structured arguments*. In H. Prakken, S. Bistarelli, F. Santini & C. Taticchi, editors: *Proceedings of the COMMA 2020*, IOS Press, pp. 123–134, doi:10.3233/FAIA200498.
- [16] Martin Caminada & Leila Amgoud (2007): *On the evaluation of argumentation formalisms*. *Artificial Intelligence* 171(5-6), pp. 286–310, doi:10.1016/j.artint.2007.02.003.
- [17] Hans van Ditmarsch, Wiebe van der Hoek & Barteld Kooi (2007): *Dynamic epistemic logic*. Springer, doi:10.1007/978-1-4020-5839-4.
- [18] Phan Minh Dung (1995): *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games*. *Artificial Intelligence* 77(2), pp. 321–357, doi:10.1016/0004-3702(94)00041-X.
- [19] Frans H. van Eemeren, Bart Garssen, Erik C. W. Krabbe, A. Francisca Snoeck Henkemans, Bart Verheij & Jean H. M. Wagemans (2014): *Handbook of Argumentation Theory*. Springer, doi:10.1007/978-90-481-9473-5.
- [20] Ronald Fagin & Joseph Y Halpern (1987): *Belief, awareness, and limited reasoning*. *Artificial intelligence* 34(1), pp. 39–76, doi:10.1016/0004-3702(87)90003-8.
- [21] Ronald Fagin, Joseph Y Halpern, Yoram Moses & Moshe Vardi (2004): *Reasoning about knowledge*. MIT press, doi:10.7551/mitpress/5803.001.0001.
- [22] Konstantin Genin & Franz Huber (2021): *Formal Representations of Belief*. In Edward N. Zalta, editor: *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.
- [23] Davide Grossi & Wiebe van der Hoek (2014): *Justified Beliefs by Justified Arguments*. In Chitta Baral, Giuseppe De Giacomo & Thomas Eiter, editors: *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference*, AAAI Press, pp. 131–140, doi:10.5555/3031929.3031947.

- [24] Davide Grossi & Fernando R. Velázquez-Quesada (2009): *Twelve Angry Men: A Study on the Fine-Grain of Announcements*. In Xiangdong He, John Horty & Eric Pacuit, editors: *Logic, Rationality, and Interaction*, Springer, pp. 147–160, doi:10.1007/978-3-642-04893-7_12.
- [25] Davide Grossi & Fernando R. Velázquez-Quesada (2015): *Syntactic awareness in logical dynamics*. *Synthese* 192(12), pp. 4071–4105, doi:10.1007/s11229-015-0733-1.
- [26] Ali Hasan & Richard Fumerton (2018): *Foundationalist Theories of Epistemic Justification*. In Edward N. Zalta, editor: *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.
- [27] Barteld Kooi (2007): *Expressivity and completeness for public update logics via reduction axioms*. *Journal of Applied Non-Classical Logics* 17(2), pp. 231–253, doi:10.3166/jancl.17.231-253.
- [28] Xu Li & Yi N. Wang (2020): *A Logic of Knowledge and Belief Based on Abstract Arguments*. In Mehdi Dastani, Huimin Dong & Leon van der Torre, editors: *Logic and Argumentation*, Springer, pp. 116–130, doi:10.1007/978-3-030-44638-3_8.
- [29] Hugo Mercier & Dan Sperber (2011): *Why do humans reason? Arguments for an argumentative theory*. *Behavioral and brain sciences* 34(2), pp. 57–74, doi:10.1017/S0140525X10000968.
- [30] Sanjay Modgil & Henry Prakken (2013): *A general account of argumentation with preferences*. *Artificial Intelligence* 195, pp. 361–397, doi:10.1016/j.artint.2012.10.008.
- [31] Sanjay Modgil & Henry Prakken (2018): *Abstract rule-based argumentation*. In Pietro Baroni, Dov M. Gabbay, Massimiliano Giacomin & Leendert van der Torre, editors: *Handbook of formal argumentation*, College Publications, pp. 287–364.
- [32] Sanjay Modgil, Francesca Toni, Floris Bex, Ivan Bratko, Carlos I. Chesñevar, Wolfgang Dvořák, Marcelo A. Falappa, Xiuyi Fan, Sarah Alice Gaggl, Alejandro J. García, María P. González, Thomas F. Gordon, João Leite, Martin Možina, Chris Reed, Guillermo R. Simari, Stefan Szeider, Paolo Torroni & Stefan Woltran (2013): *The Added Value of Argumentation*, pp. 357–403. Springer, doi:10.1007/978-94-007-5583-3_21.
- [33] Carlo Proietti & Antonio Yuste-Ginel (2021): *Dynamic epistemic logics for abstract argumentation*. *Synthese*, doi:10.1007/s11229-021-03178-5.
- [34] Chiaki Sakama & Tran Cao Son (2020): *Epistemic Argumentation Framework: Theory and Computation*. *Journal of Artificial Intelligence Research* 69, pp. 1103–1126, doi:10.1613/jair.1.12121.
- [35] François Schwarzentruber, Srdjan Vesic & Tjitze Rienstra (2012): *Building an Epistemic Logic for Argumentation*. In Luis Fariñas del Cerro, Andreas Herzig & Jérôme Mengin, editors: *Logics in Artificial Intelligence, LNCS 7519*, Springer, pp. 359–371, doi:10.1007/978-3-642-33353-8_28.
- [36] Chenwei Shi, Sonja Smets & Fernando R. Velázquez-Quesada (2017): *Argument-based belief in topological structures*. In J. Lang, editor: *Proceedings TARK 2017. EPTCS*, Open Publishing Association, doi:10.4204/EPTCS.251.36.
- [37] Dan Sperber (1997): *Intuitive and reflective beliefs*. *Mind & Language* 12(1), pp. 67–83, doi:10.1111/j.1468-0017.1997.tb00062.x.
- [38] Stephen E. Toulmin ([1958] 2003): *The uses of argument*. Cambridge university press, doi:10.1017/CBO9780511840005.
- [39] Fernando R. Velázquez-Quesada (2014): *Dynamic epistemic logic for implicit and explicit beliefs*. *Journal of Logic, Language and Information* 23(2), pp. 107–140, doi:10.1007/s10849-014-9193-0.
- [40] Yanjing Wang & Qinxiang Cao (2013): *On axiomatizations of public announcement logic*. *Synthese* 190(1), pp. 103–134, doi:10.1007/s11229-012-0233-5.
- [41] Zhe Yu, Kang Xu & Beishui Liao (2018): *Structured argumentation: Restricted rebut vs. unrestricted rebut*. *Studies in Logic* 11(3), pp. 3–17.

Appendix (Proof sketch of Theorem 1)

The outline of the proof is as follows: we first define a new class of (non-standard) models for our language (which are Kripke models where the syntactic components –awareness, accepted rules and names of rule– are maintained throughout the accessibility relation). We then show two things: (i) we can go from pointed Kripke models to its generated submodels without loosing \mathcal{L} -information (just as in the general modal case) and; (ii) we can transform systematically Kripke generated submodels into our models (again, without loosing \mathcal{L} -information). Finally, we prove completeness w.r.t. the class of non-standard models and apply (i) and (ii) to obtain the desired result. Let us unfold some of the details.

First of all, we define a **Kripke model for** $\mathcal{L} = (F, A)$ as a tuple $S = (W, \mathcal{R}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|)$ where: $W \neq \emptyset$ is a set of *possible worlds*; $\mathcal{R} \subseteq W \times W$ is a serial, transitive and euclidean relation; $\mathcal{O} : W \rightarrow \wp(A)$ is a function assigning an awareness set $\mathcal{O}(w)$ to each world w ; $\mathcal{D} : W \rightarrow \wp(\text{SEQ}(F))$ (with $n \in \mathbb{N}$) is a function assigning a set of *accepted defeasible rules* $\mathcal{D}(w)$ to each world w ; $n : (W \times \text{SEQ}(F)) \rightarrow \text{At}$ is a (possibly partial) *naming function* for defeasible rules, where $n(w, R)$ informally means “the defeasible rule R is applicable at w ”; and $\|\cdot\| : \text{At} \rightarrow \wp(W)$ is a valuation function. Moreover, we assume that for every $w, w' \in W$, $w\mathcal{R}w'$ implies $\mathcal{O}(w) = \mathcal{O}(w')$, $\mathcal{D}(w) = \mathcal{D}(w')$, and $n(w, R) = n(w', R)$. We also assume that if $(\{\varphi_1, \dots, \varphi_n\}, \varphi) \in \mathcal{D}(w)$, then $\{\varphi_1, \dots, \varphi_n, \varphi\} \not\vdash_0 \perp$ and $\{\varphi_1, \dots, \varphi_n\} \not\vdash_0 \varphi$.

Note that now WS sets depend on both the model and the world at which we are looking (since \mathcal{D} may vary from one world to another). Consequently, we use $WS^S(w)$ to denote the set of well-shaped arguments at (S, w) .

Truth w.r.t. pointed Kripke models is denoted by \models_k and defined as follows (the missing clauses are as expected):

$$\begin{aligned} S, w \models_k \Box \varphi & \text{ iff } w\mathcal{R}v \text{ implies } S, v \models_k \varphi \\ S, w \models_k \text{aware}(\alpha) & \text{ iff } \alpha \in \mathcal{O}(w) \\ S, w \models_k \text{wellshap}(\alpha) & \text{ iff } \alpha \in WS^S(w) \\ S, w \models_k \text{undercuts}(\alpha, \beta) & \text{ iff } \beta = \langle \beta_1, \dots, \beta_n \Rightarrow \varphi \rangle \text{ and } \text{Conc}(\alpha) = \neg n(w, \text{TopRule}(\beta)). \end{aligned}$$

We say that a Kripke model $S = (W, \mathcal{R}, \mathcal{O}, \mathcal{D}, n, \|\cdot\|)$ is **uniform** iff for every $w, w' \in W$ it holds that: (i) $\mathcal{O}(w) = \mathcal{O}(w')$; (ii) $\mathcal{D}(w) = \mathcal{D}(w')$; and (iii) $n(w, R) = n(w', R)$ for every $R \in \mathcal{D}$. \mathcal{K} denotes the class of all pointed Kripke models, and \mathcal{K}^u denotes the class of all uniform pointed Kripke models. We abuse of notation and use \mathcal{M} to denote the class of all *pointed* models (the standard ones that we defined in Section 2).

Transformation lemmas. Now, we need a couple of lemmas. The first one says that we can go from pointed Kripke models to Kripke uniform pointed models without loosing \mathcal{L} -information, by taking generated submodels. We use S^w to denote the submodel of S generated by w (see [14, Chapter 2]).

Lemma 1. *Let $(S, w) \in \mathcal{K}$. We have that:*

- i) $(S^w, w) \in \mathcal{K}^u$,
i.e. each pointed-generated submodel of a Kripke model is a uniform Kripke model.
- ii) For every $\varphi \in F$, $(S, w) \models_k \varphi$ iff $(S^w, w) \models_k \varphi$, i.e. truth is preserved under generated submodels.

Item i) follows easily from the definition of generated submodel and uniform Kripke model. Item ii) can be proved by induction on φ .

The second lemma says that we can go from Kripke uniform models to our models (the standard ones, defined in Section 2) without loosing \mathcal{L} -information.

Lemma 2. *For every uniform pointed Kripke model $(S, v) \in \mathcal{K}^u$, there is a pointed model $(M, w) \in \mathcal{M}$ s.t. for every $\varphi \in F$:*

$$S, v \models_k \varphi \quad \text{iff} \quad M, w \models \varphi.$$

Let us define the function τ for each uniform Kripke model as follows $\tau(S, w) = (\tau(S), \tau(w))$ where $\tau(w) = w$ and $\tau(S) = (\tau(W), \tau(\mathcal{R}), \tau(\mathcal{O}), \tau(\mathcal{D}), \tau(\mathfrak{n}), \tau(\|\cdot\|))$ s.t.:

$$\begin{aligned} \tau(W) &:= \{w\} \cup \mathcal{R}[w], \\ \tau(\mathcal{R}) &:= \mathcal{R}[w], \\ \tau(\mathcal{O}) &:= \mathcal{O}(w), \\ \tau(\mathcal{D}) &:= \mathcal{D}(w), \\ \tau(\mathfrak{n}) &:= \{(R, p) \in \text{SEQ}(F) \times \text{At} \mid \mathfrak{n}(w, R) = p\}, \\ \tau(\|\cdot\|) &:= \|\cdot\| \cap \tau(W) \text{ for every } p \in \text{At}. \end{aligned}$$

Now, it is easy to check that, for every $(S, w) \in \mathcal{K}^u$: $\tau((S, w)) \in \mathcal{M}$, that is $\tau: \mathcal{K}^u \rightarrow \mathcal{M}$. Once this is done, we can show that, for every $\varphi \in F$, it holds that:

$$S, w \models_k \varphi \quad \text{iff} \quad \tau(S, w) \models \varphi.$$

The proof of the last assertion is by induction on φ where the step for $\varphi = \text{wellshap}(\alpha)$ is another inductive argument (on the construction on α).

Completeness w.r.t. Kripke models. We can now define the **canonical Kripke** model for \mathcal{L} as:

$$S^c = (W^c, \mathcal{R}^c, \mathcal{O}^c, \mathcal{D}^c, \mathfrak{n}^c, \|\cdot\|^c),$$

where the definition of W^c , \mathcal{R}^c y $\|\cdot\|^c$ is as usual in modal logic [14]; while the definition of the rest of the elements mimics the one of awareness operators [20]:

$$\begin{aligned} \mathcal{O}^c(\Gamma) &:= \{\alpha \in A \mid \text{aware}(\alpha) \in \Gamma\}, \\ \mathcal{D}^c(\Gamma) &:= \{R \in \text{SEQ}(F) \mid \text{wellshap}(\alpha^R) \in \Gamma\}, \\ ((\Gamma, R), p) \in \mathfrak{n}^c &\quad \text{iff} \quad \text{undercuts}(\langle \neg p, \alpha^R \rangle) \in \Gamma. \end{aligned}$$

Now, we need to prove:

Lemma 3 (Canonicity). *S^c is a Kripke model for \mathcal{L} .*

For showing that S^c satisfies all conditions, we reason using maximally-consistent set properties and our axiom system. As illustrations: semantic restrictions on the accessibility relations follows from (Ax1) (see e.g. [21] or [14]), while (Ax19) permits showing that \mathfrak{n}^c is a function.

Lemma 4 (Truth). *For every $\varphi \in F$: $\varphi \in \Gamma$ iff $S^c, \Gamma \models_k \varphi$.*

The proof proceeds by induction on φ . The Boolean and modal cases are standard [14]. The cases for operators aware, conc and strict are straightforward (they actually do not make use of the induction hypothesis, due to their syntactic character). The cases for $\varphi = \text{undercuts}(\alpha, \beta)$ and $\varphi = \text{wellshap}(\alpha)$ are slightly more compromised. For the latter, another inductive argument on the structure of α is required.

Completeness w.r.t. standard models. Finally, completeness w.r.t. standard models can be proved as follows. Suppose $\Gamma \not\models \varphi$, then $\Gamma \cup \{\neg\varphi\}$ is consistent. By Lindenbaum, we have that there is a $\Gamma^+ \in W^c$ s.t $\Gamma \cup \{\neg\varphi\} \subseteq \Gamma^+$. By the Truth Lemma we have that $S^c, \Gamma^+ \models_k \Gamma \cup \{\neg\varphi\}$. By item *ii*) of Lemma 1 we have that $S^{c\Gamma^+}, \Gamma^+ \models_k \Gamma \cup \{\neg\varphi\}$ and by item *i*) we have that $S^{c\Gamma^+}, \Gamma^+$ is a pointed uniform Kripke model. Then by Lemma 2 we know that $\tau(S^{c\Gamma^+}, \Gamma^+) \models \Gamma \cup \{\neg\varphi\}$ which implies by definition of semantic logical consequence that $\Gamma \not\models \varphi$.