



Dynamic epistemic logics for abstract argumentation

Carlo Proietti^{1,2} · Antonio Yuste-Ginel³ 

Received: 18 July 2020 / Accepted: 24 April 2021 / Published online: 15 May 2021
© The Author(s) 2021

Abstract

This paper introduces a multi-agent dynamic epistemic logic for abstract argumentation. Its main motivation is to build a general framework for modelling the dynamics of a debate, which entails reasoning about goals, beliefs, as well as policies of communication and information update by the participants. After locating our proposal and introducing the relevant tools from abstract argumentation, we proceed to build a three-tiered logical approach. At the first level, we use the language of propositional logic to encode states of a multi-agent debate. This language allows to specify which arguments any agent is aware of, as well as their subjective justification status. We then extend our language and semantics to that of epistemic logic, in order to model individuals' beliefs about the state of the debate, which includes uncertainty about the information available to others. As a third step, we introduce a framework of dynamic epistemic logic and its semantics, which is essentially based on so-called event models with factual change. We provide completeness results for a number of systems and show how existing formalisms for argumentation dynamics and unquantified uncer-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 748421. Carlo Proietti conducted a substantial portion of this research at the Institute for Logic, Language and Computation of the University of Amsterdam. The research activity of Antonio Yuste-Ginel is funded by the predoctoral Grant MECD-FPU 2016/04113 of the Spanish Ministry of Universities. Both authors have contributed equally to the current paper. We are indebted to Andreas Herzig for useful comments and technical insights. Special thanks to anonymous reviewer 2 for this journal, whose comments helped to significantly improve this paper. We also thank the audience of the LiLaC's group seminar at the Institut de Recherche en Informatique de Toulouse for useful feedback at an earlier stage of this work.

✉ Antonio Yuste-Ginel
antonioyusteginel@gmail.com

Carlo Proietti
carlo.proietti@ilc.cnr.it

- ¹ Institute for Computational Linguistics “A. Zampolli” (ILC), National Research Council of Italy (CNR), Genova, Italy
- ² Institute for Logic, Language, and Computation (ILLC), University of Amsterdam, Amsterdam, The Netherlands
- ³ Department of Philosophy, Universidad de Málaga, Málaga, Spain

tainty can be reduced to their semantics. The resulting framework allows reasoning about subtle epistemic and argumentative updates—such as the effects of different levels of trust in a source—and more in general about the epistemic dimensions of strategic communication.

Keywords Abstract argumentation · Dynamic epistemic logic · Awareness logics · Multi-agent argumentation frameworks · Persuasion · Strategic Argumentation

Contents

1	Introduction	8642
2	Historical background and general motivations	8645
3	Multi-agent argumentation frameworks	8647
4	Encoding argumentative notions	8650
5	Epistemic logics for abstract argumentation	8653
6	Epistemic and argumentative dynamics	8659
7	Modelling persuasion, sceptic updates and conditional trust	8666
8	Relation to other formalisms	8669
8.1	Incomplete AFs	8670
8.1.1	From incomplete AFs to \mathcal{EA} -models	8670
8.1.2	From \mathcal{EA} -models to incomplete AFs	8671
8.2	Control AFs	8673
8.3	Reasoning about opponent models	8677
9	Discussion, open problems and future research	8678
	Appendix	8681
A1.	Encodings	8681
	Proposition 1	8681
A2.	Proof of Theorem 1	8684
A3.	Proofs of Sect. 6	8688
	Lemma 1 (Closure)	8688
	Lemma 2	8689
	Details of the proof of Theorem 2	8690
A4.	Preserving doxastic relations via an enhanced truth condition	8691
A5.	Incomplete and control AFs	8694
	References	8697

1 Introduction

When engaging in a debate, we do not only exchange arguments, we also reason about information available to others, and both things play crucial roles. On the one hand, acquiring and communicating new arguments can shift one’s point of view on the issue of the debate, or make it more robust. On the other hand, beliefs about someone else’s background information determine which arguments one is willing to put on the table and in which order, like in a game of incomplete information. To understand how argumentation unfolds in real-life debates we need to reason, at least, about *goals*, *beliefs*, and *information change*. The latter involves *communication moves* of the speaker (sender)—choosing and disclosing certain piece of information—and *information updates* by the hearer (receiver)—incorporating that piece into her

knowledge base.¹ Our running example illustrates how strongly these elements interact with each other.

Example 1 Charlie wants to convince his mother that *he has right to have a chocolate candy (a)*. Mom rebuts that *too much chocolate is not good for his teeth (b)*. Charlie may counterargue that *he didn't have chocolate since yesterday (d)*. Unfortunately for him, Mom has seen him *grabbing chocolate from the pantry just a few hours ago (e)*—by the way, she wrongly thinks that Charlie noticed this. Alternatively, Charlie may quote *scientific evidence from a journal paper on Pscience that eating chocolate is never too much (c)*. Mom does not know that *this paper has been retracted (f)* and, in principle, this would be a safe move for Charlie.²

Charlie's goal is to make *a* justified in the eyes of his mother. To achieve this goal he needs to rebut *b*. He has several options to do so: he may put forward *d*, or *c* or both, i.e. he has to select a communication move. To choose his strategy, he needs clues on Mom's background information, i.e. to form beliefs about her beliefs. Finally, success also depends on Mom's attitude towards the information she receives, i.e. her updating policy.

Logical languages and semantics provide a powerful tool to reason about these aspects of argumentation.³ Here we aim to show that dynamic epistemic logic (DEL) can serve as a general framework to deal with many conceptual aspects of argumentation which are of interest in general argumentation theory and its more recent developments in AI and computer science, specifically in the study of computational models of argument (see Sect. 8).

We can see the language of DEL as structured in three layers. The first layer consists of the propositional language. The one we adopt enables to encode the state of a multi-agent debate, which semantically constitutes a propositional valuation. Using tools from abstract argumentation (Dung 1995), such states are modelled here as *multi-agent argumentation frameworks* (MAF). They include (a) the description of a *universal argumentation framework* consisting of all the arguments (and conflicts among them) that are potentially available to the agents, and (b) the specific information of each agent, i.e. the part of the universal framework each agent is aware of. Languages of propositional logic are widely used to encode argumentation frameworks, see Besnard et al. (2020) for a survey. In many cases such encodings employ minimal resources as they are designed with efficiency in mind, e.g. to reduce computational problems in abstract argumentation to SAT-solving problems (Cerutti et al. 2013). The language and semantics we adopt are not tailored for computational purposes and are rather rich instead. On the other hand, they allow us to encode fine-grained argumentative notions such as the agents' subjective justification status of specific arguments, which, as we will see, is needed to talk about their goals.

¹ In more general words, we should touch upon the procedural and dialogical aspects of argumentation. This entails going beyond the somehow traditional divide between the *logical*, *dialectical* and *rhetorical* levels of argumentation as outlined e.g. by Wenzel (1992).

² This example was inspired by Mercier and Sperber (2017).

³ To avoid confusion, here we see logic as a descriptive tool to talk about mathematical structures such as abstract multi-agent systems. As it will be clear later on, see in particular Sect. 2, this attitude is radically different from a normative view, which sees logic as a guidance to 'correct' reasoning or 'good' argumentation.

The modal part of the language constitutes the second layer and includes epistemic (resp. doxastic) operators for knowledge (resp. belief). With these operators it is possible to express individual attitudes at any level of nesting, such as the second level attitude ‘Charlie believes that Mom believes that argument a is justified for Charlie’. At this stage, the language is interpreted in standard Kripke-style semantics where states are MAFs. The plurality of states serves to capture the uncertainty of agents about the actual state of the debate. As mentioned, modelling uncertainty is relevant to analyze the strategic aspects of argumentation. Recent approaches in formal argumentation model uncertainty by means of *incomplete argumentation frameworks* (Baumeister et al. 2018a,b), *control argumentation frameworks* (Dimopoulos et al. 2018), and *opponent models* (Oren and Norman 2009; Rienstra et al. 2013; Hadjinikolis et al. 2013; Thimm 2014; Black et al. 2017). These approaches provide efficient solutions for computational and application purposes, such as building automated persuasion systems (Hunter 2018). Our goal here is mainly one of conceptual analysis, for which we seek to achieve generality. Indeed, we show in Sect. 8 that it is possible to translate the central notions of these approaches by means of our language and semantics. Moreover, having the expressive power for talking about epistemic attitudes at any level, our language is able to frame agents’ goals of complex kinds. In our running example, Charlie’s goal amounts to inducing Mom to believe that a is justified, i.e. a first-level attitude. However, we shall see in Sect. 7 that goals and strategies for action may entail more articulated nestings. Furthermore, although we frame our main examples in contexts of strategic and persuasive argumentation, this framework is not conceptually limited to such contexts. Other uses of argumentation entail different kinds of goals but, insofar as they can be phrased in terms of individual or collective beliefs, the DEL approach is useful there too. This holds, for example, for collective *inquiry*, where the aim is to reach *common knowledge* or *shared belief*.⁴

The third layer of the language includes dynamic modalities to reason about the effect of argumentative actions (e.g. communicating an argument) and different belief updates by the agents. Here again, while dynamics of argument communication is the focus of a well-established tradition in abstract argumentation (see Doutre and Maily 2018 for a survey), belief updates are mostly confined to the tradition going from AGM belief revision (Alchourrón et al. 1985) to DEL (van Ditmarsch et al. 2007; van Benthem 2011). To the best of our knowledge, there is no unified logical framework for treating both aspects.⁵ A general framework for reasoning about argumentative and epistemic actions becomes relevant insofar agents are liable to revise their knowledge base in different ways, as it is the case for Mom in our running example. For this purpose, we use a rather expressive language, the one of *DEL with factual change*

⁴ The influential classification by Walton (1984), Walton and Krabbe (1995) distinguishes six types of dialogical contexts depending on their respective goals, namely persuasion, negotiation, information seeking, deliberation, inquiry and quarrel. Some of them require other conceptual ingredients to be framed, such as desires and intentions, but beliefs are essential for all of them.

⁵ Some works (Booth et al. 2013; de Saint-Cyr et al. 2016 among others) integrate abstract argumentation with belief revision theory to incorporate an epistemic dimension. Modelling techniques from DEL provide however two additional features which are relevant in this context and left opaque in belief revision: (1) expressing higher order beliefs of agents and (2) reasoning explicitly—in the object language—about how agents perceive changes.

(van Ditmarsch et al. 2005; van Benthem et al. 2006) which comes at the price of a blow in computational complexity.⁶

The rest of this paper is organized as follows. In Sect. 2 we illustrate the background and the general motivations for our work. Section 3 presents the preliminary tools from abstract argumentation and introduces the notion of MAF. There are indeed several alternative ways to represent a multi-agent scenario of debate. Here we take a specific option and leave critical discussion of other possibilities to Sect. 9. In Sect. 4 we introduce a propositional language to encode MAFs and prove soundness for this encoding in Proposition 1. In Sect. 5, we develop the epistemic fragment for reasoning about knowledge and belief in abstract argumentation. We introduce the general semantics of *epistemic argumentative models* (Definition 8). After this, we isolate specific subclasses of models that capture a number of constraints on the awareness of arguments and attacks, as well as on epistemic accessibility. Then we provide axiomatizations for these subclasses and show their soundness and completeness in Theorem 1. In Sect. 6 we introduce the full language of DEL for argumentative models. Semantics are given in terms of event models and product updates as in Baltag and Moss (2004). Here we show how to model basic communication moves and information updates under full trust. We then provide completeness results via reduction axioms (Theorem 2). In Sect. 7, we exploit event models to encode the effects of more subtle policies of communication and information update under mixed trust. In Sect. 8 we show how this framework relates to other formalisms developed in the area of computational argumentation. We conclude with Sect. 9, by discussing conceptual alternatives to our modelization as well as open problems and future work. Given the length of the proofs of most of our results, and the substantial amount of tools they involve, we leave them for the final “Appendix”, where we also prove additional results for an extended modal language.

2 Historical background and general motivations

By bringing together two different formal traditions such as epistemic modal logic and abstract argumentation, we aim not only to provide results of interest for both, but also to show that their respective toolboxes provide powerful conceptual resources to think both traditions in a different light. At least since Aristotle, logic and the study of argumentation ran along separated lines, the latter being the exclusive competence of rhetoric. This separation contributed to crystalize the notion of deductive inference from classical logic as the golden standard of correct reasoning. Classical inference is non-defeasible and typically abstracts away from the dialogical/adversarial dimension in which real-life argumentation takes place. From the philosophical side, major criticisms of this paradigm came in the twentieth century from the works of Toulmin (2003), Perelman and Olbrechts-Tyteca (1958), and Hamblin (1970). Yet, formal

⁶ Although decidable, the satisfiability and the model-checking problem for DEL (without factual change) are known to be, respectively, NEXPTIME-complete and PSPACE-complete (Aucher and Schwarzentruher 2013). The solution of many reasoning tasks in abstract argumentation are based on model-checking algorithms (Sect. 8). This, combined with the use of the of an expressive propositional encoding, is unlikely to provide efficient algorithms, and this will not be the purpose of the present work.

research was still mastered by traditional approaches, at least until the new-born field of artificial intelligence undertook modelling human-like reasoning, and eventually converged in the definition of systems of non-monotonic logics (Reiter 1980) and defeasible reasoning (Pollock 1987, 1991). A turning point has been the introduction of abstract argumentation by Dung (1995). Here the main tool are argumentation frameworks, i.e. directed graphs which represent debates at an abstract level, where arguments are nodes and attacks from one argument to another—e.g. undercuts or rebuttals—are directed edges. The key semantic notion in abstract argumentation is that of a *solution*, i.e. a set of arguments that constitutes an acceptable opinion as the outcome of a debate. It turns out that the most relevant semantics for non-monotonic and defeasible reasoning can be expressed in terms of solution concepts for argumentation frameworks (Dung 1995), which thus provide a powerful mathematics for defeasible reasoning in dialogical scenarios. Abstract argumentation can be seen as a very general theory of conflict that, in the words of Dung, captures the fact that

the way humans argue is based on a very simple principle which is summarized succinctly by an old saying: “The one who has the last word laughs best” (Dung 1995, p. 322).

For our purposes, argumentation frameworks are a first adequate building block to model scenarios like Example 1, where solution concepts provide the essentials for defining agents’ (defeasible) *justification* of an argument and their *goals*.

From the beginning of the 1980s—in the wake of the “dynamic turn” pushed by the introduction of propositional dynamic logic (Fischer and Ladner 1979)—logicians have dedicated increasing interest to information change, the study of how information states transform under new data. The early approach that dominated the field was AGM belief revision (Alchourrón et al. 1985), later joined by DEL (Plaza 1989; Gerbrandy and Groeneveld 1997; Baltag et al. 2016). Dynamic epistemic logics, endowed with plausibility models and operators of conditional belief, allow a systematic treatment of AGM-style belief revision and can model a wide range of information updates (van Benthem 2007; Baltag and Smets 2008). A dominant part of the work in both areas has been shaped by a *normative* approach to the study of information change. AGM belief revision typically focuses on postulates encoding the properties that an update operation should satisfy to be considered rational. Although DEL has the flexibility to model a wide range of epistemic transformations, including the effects of lying and deception (Baltag and Moss 2004; van Ditmarsch et al. 2007), it is fair to say that the mainstream focus has been the update of information under new evidence, where the latter is intended as truthful information made available to the agent. The typical belief upgrades studied in DEL applied to belief revision—such as public announcement $!P$, lexicographic upgrade $\uparrow P$ and minimal upgrade $\uparrow P$ —implicitly assume that the source of information is trusted as infallible (public announcement) or at least believed to be trustworthy (minimal upgrade) (Rodenhäuser 2014). However, most situations of real-life information exchange among individuals are of mixed trust: the source of information is taken to be trustworthy to a limited, or at least context-dependent, extent: we may trust Professor Bertrand Russell on logic matters, probably less so when he predicts the outcome of the next horse race. With the exception of Rodenhäuser

(2014), mixed trust of this and other kinds deserved limited attention in DEL. We will handle situations of mixed trust with our formal machinery in Sect. 7.

From a normative perspective, many interesting real-life mechanisms of information update are deemed “descriptive” and left to psychologists, when not discarded as reasoning flaws of an imperfect reasoner. This holds for *confirmation bias* (Wason 1960), more adequately called *myside bias* (Perkins et al. 1986)—that is the tendency to strictly evaluate information disconfirming our prior opinions and, vice versa, loosely filter and search for confirming evidence—and for the operation by which we reduce *cognitive dissonance* upon receiving information which is inconsistent with our prior beliefs (Festinger 1957). Scholars in logic can hardly be blamed for this attitude, since it is supported by most psychology of reasoning, as the extensive debate on, e.g., the Wason selection task witnesses (Wason 1966). More recently, Mercier and Sperber’s argumentative theory of reasoning advances a different view, according to which these purported flaws are rather *features* of reasoning, having an evolutionary explanation in the social context of human communication (Mercier and Sperber 2011, 2017). The argumentative theory of reasoning is a naturalized approach that sees reasoning as a specific cognitive module which “evolved for the production and evaluation of arguments in communication” (Mercier and Sperber 2011, p. 58) rather than to perform sound logical and probabilistic inferences, or to enhance individual cognition. Seen from this angle, the myside bias serves the goal of convincing others and keeping *epistemic vigilance*. Indeed, what we often blame as a bad attitude in everyday confrontations is a common—and mostly healthy—practice in scientific debate over new theories and explanations (Kelly 2008). In general, an argumentation-based approach to reasoning and communication can explain collective dynamics like *groupthink* and *opinion polarization*. When individuals with similar opinions on a given issue discuss, they tend to mutually reinforce their views by providing each other novel and persuasive arguments towards the same direction.⁷ A further step in this direction is to investigate the triggering effect of more subtle mechanisms of information update, akin to the myside bias. Sect. 7 shows that DEL can be used for this purpose. Indeed, the notion we characterize as *sceptic update* provides one possible way of understanding biased assimilation of new arguments. Before getting there, a careful logical construction is needed though, that we begin in the next section.

3 Multi-agent argumentation frameworks

The fundamental notion we employ is that of an argumentation framework, which is no more and no less than a directed graph.

Definition 1 [*Argumentation framework* (Dung 1995)] An *argumentation framework* (AF) is a pair $F = (A, R)$ where $A \neq \emptyset$ is a set of *arguments* and $R \subseteq A \times A$ is called the *attack relation*. We adopt the infix notation aRb to abbreviate $(a, b) \in R$. Given

⁷ This explanation has been formulated and tested by social psychologists under the name of *persuasive arguments theory* (Vinokur and Burstein 1974). More recently, Mäs and Flache (2013) investigated the polarizing effects of argumentative exchange by combining lab experiments with multi-agent computer simulations, showing that a small degree of homophily—the tendency to communicate with like-minded individuals—is able to generate clusters of opposite polarizing opinions in groups.

a set of arguments $B \subseteq A$, we denote by B^+ the *set of arguments attacked by B*, that is $B^+ := \{a \in A \mid \exists b \in B: bRa\}$.

An AF represents a full debate seen from a third-person point of view, where all potential arguments and attacks are on the table. Clearly, at a given moment of a debate, each participant is aware of a specific subset of arguments and attacks, i.e. her subjective information about the debate. This calls for the definition of *multi-agent AF*. A number of alternative options is available in the literature, and many others are there. Each choice depends on specific assumptions about the common ground of the debate and the awareness constraints on the agents' information. In our approach we assume the following:

- (a) the set of arguments that are potentially available to agents is finite;
- (b) it is fixed in advance;
- (c) there is an objective matter of fact; independent from subjective views, by which an argument attacks another;
- (d) agents can only be aware of arguments in set (a), i.e. there are no *non-existing* or *virtual* arguments (cf. Schwarzenrüber et al. 2012; Rienstra et al. 2013);
- (e) agents can be aware of an attack between a and b only if they are aware of both a and b ;
- (f) if an agent is aware of an attack then this attack holds;
- (g) if an objective attack holds between two arguments and some agent is aware of both, then she is also aware of the attack.

Together, (f) and (g) imply that agents have a (locally) *sound and complete awareness of attacks* (SCAA). In general, each of these choices has alternatives, and this gives a very large combinatorics of possibilities for design, which we critically discuss in Sect. 9. It may seem at first sight that constraints (a)–(g) impose strict limitations on the agent' uncertainty, but we shall see (Sect. 5) that this is not quite so, since the modal component of our framework allows to recapture all sorts of uncertainty. Based on our assumptions we define a multi-agent argumentation framework as follows:

Definition 2 (*Multi-agent argumentation framework*) A *multi-agent argumentation framework* (MAF) for a non-empty and finite set of agents Ag is a 4-tuple $(A, R, \{A_i\}_{i \in \text{Ag}}, \{E_1, \dots, E_n\})$ such that (A, R) is a finite AF (the *universal argumentation framework*, UAF), $A_i \subseteq A$ is the set of arguments agent i is currently aware of, and $\{E_1, \dots, E_n\}$ is a specific enumeration of the subsets of A , which we assume as fixed from now on. Given a MAF and an agent $i \in \text{Ag}$, agent i 's *partial information* is defined as (A_i, R_i) where $R_i := R \cap (A_i \times A_i)$.

Having A and R finite and fixed captures the constraints from (a) to (c). Constraint (d) amounts to $A_i \subseteq A$. Finally, the definition of R_i subsumes (e)–(g). The enumeration $\{E_1, \dots, E_n\}$ of $\wp(A)$ is an important device for encoding, the use of which will be clarified in Sect. 4. Figure 1 provides a pictorial representation of a two-agent MAF describing Example 1.

Solution concepts from abstract argumentation are a key to subjective justification and goals. A *solution* is a set of arguments that meets intuitive constraints to constitute

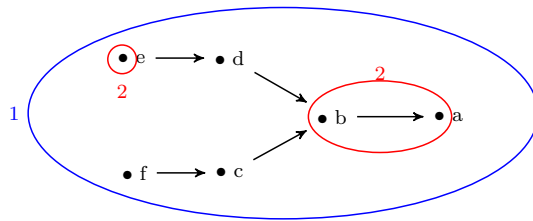


Fig. 1 A MAF for Example 1 (Charlie and Mom). The universal argumentation framework consists of nodes a – f and the corresponding attacks, as described in the Example. Agent 1 (Charlie) is aware of the entire universal argumentation framework (area in blue) while agent 2 (Mom) is only aware of $\{a, b, e\}$ and the attacks between them (red ellipses). We omit the representation of an enumeration of $\wp(A)$

an acceptable point of view.⁸ Several solution concepts have been introduced by Dung (1995) and subsequent work in abstract argumentation, see Baroni et al. (2018) for an extensive state-of-the-art. For the sake of presentation, we focus on preferred solutions, but our approach can be straightforwardly extended to other admissibility-based semantics (i.e., grounded, complete and stable).⁹

Definition 3 (Defence and preferred solutions) Given an AF $F = (A, R)$, a set of arguments $B \subseteq A$, and an argument $a \in A$: B *defends* a iff for every $c \in A$: if cRa then $c \in B^+$. Moreover, B is said to be a *complete solution* iff (1) it is *conflict-free*, i.e. $B \cap B^+ = \emptyset$ and (2) it contains precisely the arguments that it defends, i.e. $b \in B$ iff B defends b . B is a *preferred solution* iff it is a maximal (w.r.t. set inclusion) complete solution. Given an AF $F = (A, R)$ we denote by $\text{Pr}(F)$ the set of all its preferred solutions.

In the UAF of Fig. 1, the only preferred solution is $\{b, e, f\}$. This also corresponds to agent’s 1 preferred solution, as his awareness set A_1 coincides with the entire framework. When we relativize to agent’s 2 awareness set A_2 , we obtain instead $\{b, e\}$ as the unique preferred solution. An AF may have more than one preferred solution. Plurality of solutions allows to define—following Wu and Caminada (2010)—the *fine-grained justification status* of an argument relative to an AF. The latter is key to express *graded* notions of acceptability (Beirlaen et al. 2018; Baroni et al. 2019) for reasoning about agents’ goals and the degree of their opinion about the debated issue.¹⁰

⁸ Alternative names for solutions in the literature are *extensions* or *semantics*, which are almost interchangeable. The latter are indeed more standard in the field of abstract argumentation. We opt for the more neutral “solution” to avoid confusion with homonymous notions in logic.

⁹ Research with methods of experimental psychology (Rahwan et al. 2010) suggests that preferred solutions are, among those introduced by Dung (1995), the best predictor for human argument acceptance. More extended experiments by Cramer and Guillaume (2018) confirm this finding, but further speak in favour of so-called *naive-based* semantics such as CF2 (Baroni et al. 2005). However, the authors carefully warn that all the results may be influenced by the specific thematic contexts of the natural-language arguments chosen in the experimental setting (news reports, arguments based on scientific publications, and arguments based on the precision of a calculation tool). In particular, only one context (scientific publications) was used for the comparison between preferred and CF2 solutions.

¹⁰ This can be seen as generalization of the classical concepts of credulous/sceptic acceptance and is probably the most immediate way to define graded acceptability (Baroni et al. 2019) of arguments. Moreover, it provides an elegant way of dealing with the phenomenon of floating conclusions (see Wu and Caminada 2010 for details).

We follow the *extension-based* characterization of this notion provided by Baroni et al. (2018).¹¹

Definition 4 (*Fine-grained justification status*) Given an AF $F = (A, R)$ and an argument $a \in A$, then a is said to be:

- *strongly (or sceptically) accepted* iff $\forall E \in \text{Pr}(F) a \in E$;
- *weakly accepted* iff $(\exists E \in \text{Pr}(F): a \in E, \exists E \in \text{Pr}(F): a \notin E, \text{ and } \forall E \in \text{Pr}(F) a \notin E^+)$;
- *weakly rejected* iff $(\exists E \in \text{Pr}(F): a \in E^+, \exists E \in \text{Pr}(F): a \notin E^+, \text{ and } \forall E \in \text{Pr}(F) a \notin E)$;
- *strongly rejected* iff $\forall E \in \text{Pr}(F) a \in E^+$; and
- *borderline* otherwise.¹²

Note that the justification status of an argument is always relative to an AF $F = (A, R)$, but we omit an explicit reference to F when the context is clear enough. Again, the notion can be straightforwardly relativised to agents. For instance, given $\text{MAF} = (A, R, \{A_i\}_{i \in \text{Ag}}, \{E_1, \dots, E_n\})$ we say that $a \in A$ is strongly accepted by agent j iff $a \in A_j$ and a is strongly accepted w.r.t. (A_j, R_j) . As an example, argument b of Fig. 1 is strongly accepted by 1 and 2, and argument a is strongly rejected by both agents.

4 Encoding argumentative notions

Logical languages are a general tool to describe mathematical structures, and multi-agent AFs are one of these. Compared to others, a propositional language has minimal descriptive power though.¹³ However, it turns out that, in the finite case, its expressivity is sufficient for our purpose to encode the notions introduced in the previous section.¹⁴ Furthermore, since we construct a Kripke semantics where multi-agent AFs are states (Sect. 5), a propositional language provides a natural fit with the techniques of epistemic logic.

The set of propositional variables $\mathcal{V}_{\text{Ag}}^A$, where A is a set of arguments (intuitively, the domain of the UAF) and Ag is a set of agents, is defined as the union of the following sets:

$$\mathcal{AT} := \{a \rightsquigarrow b \mid (a, b) \in A \times A\};$$

¹¹ The original one by Wu and Caminada (2010) is presented in terms of *labellings* (Caminada 2006; Caminada and Gabbay 2009). Both ways are equivalent (Baroni et al. 2018).

¹² The class of borderline arguments allows for further distinctions (Wu and Caminada 2010). We overlook them here as it is not our primary interest to provide a full classification. In the rest of our example we indeed only make use of strong acceptance and strong rejection, which are the only possible statuses of an argument in a well-founded graph as that of Fig. 1.

¹³ There is already a significant amount of work on encoding abstract argumentation semantics with logical languages. Typical candidates are propositional logic (Besnard and Doutre 2004; Besnard et al. 2014; Doutre et al. 2014, 2017), modal logic (Grossi 2010a, b; Caminada and Gabbay 2009), first-order logic (de Saint-Cyr et al. 2016) and second order logic (Dvořák et al. 2012).

¹⁴ To the best of our knowledge, ours is the first encoding of the (fine-grained) justification status of an argument in propositional logic.

$$\mathcal{O} := \{aw_i(a) \mid i \in \text{Ag}, a \in A\}; \text{ and}$$

$$\mathcal{B} := \{a \in E_k \mid a \in A, E_k \subseteq A\}.$$

Each variable $a \rightsquigarrow b$ reads “argument a attacks b ” and $aw_i(a)$ stands for “agent i is aware of a ”. The informal reading of the third kind of variables $a \in E_k$ is “argument a belongs to subset E_k ”. These variables are needed because the definition of (fine-grained) justification status quantifies over sets (Definition 4). The language $\mathcal{L}(\mathcal{V}_{\text{Ag}}^A)$ is built from $\mathcal{V}_{\text{Ag}}^A$ using Boolean functors $\neg, \wedge, \vee, \rightarrow$ and \leftrightarrow as usual. A given MAF = $(A, R, \{A_i\}_{i \in \text{Ag}}, \{E_1, \dots, E_n\})$ determines unequivocally its associated set of variables $\mathcal{V}_{\text{Ag}}^A$.

The semantics of this propositional language is defined, as standard, by means of valuations of its propositional variables. Given a valuation $v \subseteq \mathcal{V}_{\text{Ag}}^A$ and a propositional variable p , we say that p is true at v iff $p \in v$. A valuation recursively determines the truth value of any formula $\varphi \in \mathcal{L}(\mathcal{V}_{\text{Ag}}^A)$ in the usual way. $v \models \varphi$ stands for “ φ is true at v ”.

Definition 5 (*Associated valuation and theory of a MAF*) Given MAF = $(A, R, \{A_i\}_{i \in \text{Ag}}, \{E_1, \dots, E_n\})$, we define its unequivocally associated valuation as $v_{\text{MAF}} := \{a \rightsquigarrow b \mid (a, b) \in R\} \cup \{aw_i(a) \mid a \in A_i\}_{i \in \text{Ag}} \cup \{a \in E_k \mid a \in E_k \text{ for every } 1 \leq k \leq n\}$. Furthermore, the following Boolean formula Th_{MAF} , called the *theory of MAF*, encodes MAF, in the sense that v_{MAF} is the unique valuation such that $v_{\text{MAF}} \models \text{Th}_{\text{MAF}}$:

$$\text{Th}_{\text{MAF}} := \bigwedge_{(a,b) \in R} a \rightsquigarrow b \wedge \bigwedge_{(a,b) \notin R} \neg(a \rightsquigarrow b) \wedge \bigwedge_{1 \leq k \leq n} \left(\bigwedge_{a \in E_k} a \wedge \bigwedge_{a \notin E_k} \neg a \right)$$

$$\wedge \bigwedge_{i \in \text{Ag}} \left(\bigwedge_{a \in A_i} aw_i(a) \wedge \bigwedge_{a \notin A_i} \neg aw_i(a) \right).$$

Example 2 Let $\text{MAF}_0 = (A, R, \{A_1, A_2\}, \{E_1, E_2, E_3, E_4\})$ s.t. $A = \{a, b\}$, $R = \{(b, a)\}$, $A_1 = \{a\}$, $A_2 = \{b\}$, $E_1 = \emptyset$, $E_2 = \{a\}$, $E_3 = \{b\}$, $E_4 = \{a, b\}$; we have that $\text{Th}_{\text{MAF}_0} = \neg a \rightsquigarrow a \wedge \neg a \rightsquigarrow b \wedge b \rightsquigarrow a \wedge \neg b \rightsquigarrow b \wedge (\neg a \in E_1 \wedge \neg b \in E_1) \wedge (a \in E_2 \wedge \neg b \in E_2) \wedge (\neg a \in E_3 \wedge b \in E_3) \wedge (a \in E_4 \wedge b \in E_4) \wedge aw_1(a) \wedge \neg aw_1(b) \wedge \neg aw_2(a) \wedge aw_2(b)$.

For what follows it is relevant to note that not every valuation is a valuation for a MAF. The reason is that subset variables may fail to represent a proper enumeration of subsets, in the sense of the following definition.

Definition 6 Let A be a finite set of arguments with $|\wp(A)| = n$, we say that a valuation $v \subseteq \mathcal{V}_{\text{Ag}}^A$ represents an enumeration of $\wp(A)$ iff for all $k, m: 1 \leq k < m \leq n$ it holds that $\{x \in A \mid x \in E_k \in v\} \neq \{x \in A \mid x \in E_m \in v\}$.

The inequality of two sets E_k and E_m can be expressed in our propositional language:

$$E_k \neq E_m := \bigvee_{x \in A} \neg(x \in E_k \leftrightarrow x \in E_m).$$

This allows to encode the representation of an enumeration by the following formula

$$\bigwedge_{1 \leq k < m \leq n} E_k \overset{\bullet}{\neq} E_m \quad (\text{subset enumeration}).$$

Clearly, for any MAF it holds that

$$v_{\text{MAF}} \models \bigwedge_{1 \leq k < m \leq n} E_k \overset{\bullet}{\neq} E_m$$

Most importantly, based on this language and semantics we can provide encodings for the relevant notions introduced in the previous section, as by the following list:

- $E_k \sqsubseteq E_l := \bigwedge_{a \in A} (a \in E_k \rightarrow a \in E_l)$,
- $E_k \sqsubset E_l := E_k \sqsubseteq E_l \wedge \bigvee_{a \in A} (a \in E_l \wedge \neg a \in E_k)$,
- $\text{conf_free}_i(E_k) := \bigwedge_{a \in A} \left(a \in E_k \rightarrow \left(\text{aw}_i(a) \wedge \neg \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a) \right) \right)$,
- $\text{complete}_i(E_k) := \text{conf_free}_i(E_k) \wedge \bigwedge_{a \in A} \left(a \in E_k \leftrightarrow \bigwedge_{b \in A} \left((\text{aw}_i(b) \wedge b \rightsquigarrow a) \rightarrow \bigvee_{c \in A} (c \in E_k \wedge c \rightsquigarrow b) \right) \right)$,
- $\text{preferred}_i(E_k) := \text{complete}_i(E_k) \wedge \neg \bigvee_{1 \leq l \leq n} (\text{complete}_i(E_l) \wedge (E_k \sqsubset E_l))$,
- $\text{stracc}_i(a) := \bigwedge_{1 \leq k \leq n} (\text{preferred}_i(E_k) \rightarrow a \in E_k)$,
- $\text{wekacc}_i(a) := \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge a \in E_k) \wedge \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge \neg a \in E_k) \wedge \neg \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a))$,
- $\text{strej}_i(a) := \bigwedge_{1 \leq k \leq n} (\text{preferred}_i(E_k) \rightarrow \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a))$,
- $\text{wekrej}_i(a) := \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a)) \wedge \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge \bigwedge_{b \in A} (b \in E_k \rightarrow \neg b \rightsquigarrow a)) \wedge \bigwedge_{1 \leq k \leq n} (\text{preferred}_i(E_k) \rightarrow \neg a \in E_k)$, and
- $\text{border}_i(a) := \neg \text{stracc}_i(a) \wedge \neg \text{wekacc}_i(a) \wedge \neg \text{strej}_i(a) \wedge \neg \text{wekrej}_i(a)$.

The shorthand $E_k \sqsubseteq E_l$ (resp. $E_k \sqsubset E_l$) stands for “ E_k is a subset (resp. a proper subset) of E_l ”. $\text{conf_free}_i(E_k)$ (resp. $\text{complete}_i(E_k)$, $\text{preferred}_i(E_k)$) means “the set E_k is conflict-free (resp. complete, preferred) for agent i (i.e. w.r.t. (A_i, R_i))”. $\text{stracc}_i(a)$ encodes “argument a is strongly accepted by agent i ” (Definition 4). Analogously, $\text{wekacc}_i(a)$, $\text{strej}_i(a)$, $\text{wekrej}_i(a)$ and $\text{border}_i(a)$ stand respectively for “argument a is weakly accepted, strongly rejected, weakly rejected, borderline for agent i ”.¹⁵

The following proposition shows that our encoding is sound, following the *satisfiability approach* of Besnard et al. (2014), in the sense that MAF has a given property if and only if its encoding is true at v_{MAF} .

¹⁵ We omit to encode the absolute notions of conflict-freeness, completeness etc., i.e. the ones relative to (A, R) , since they can be easily reconstructed from those relative to agents by carefully omitting awareness variables.

Proposition 1 Let $MAF = (A, R, \{A_i\}_{i \in Ag}, \{E_1, \dots, E_n\})$ be a MAF, let $\mathcal{L}(\mathcal{V}_{Ag}^A)$ be the propositional language for MAF. The following holds, where $1 \leq k, l \leq n, i \in Ag$, and $a \in A$:

1. $v_{MAF} \models E_k \sqsubseteq E_l$ (resp. $v_{MAF} \models E_k \sqsubset E_l$) iff $E_k \subseteq E_l$ (resp. $E_k \subset E_l$).¹⁶
2. $v_{MAF} \models \text{conf_free}_i(E_k)$ iff E_k is conflict free w.r.t. (A_i, R_i) (that is, iff $E_k \subseteq A_i$ and E_k is conflict-free).
3. $v_{MAF} \models \text{complete}_i(E_k)$ iff E_k is complete w.r.t. (A_i, R_i) .
4. $v_{MAF} \models \text{preferred}_i(E_k)$ iff E_k is preferred w.r.t. (A_i, R_i) .
5. $v_{MAF} \models \text{stracc}_i(a)$ (resp. $\text{wekacc}_i(a)$, $\text{wekrej}_i(a)$, $\text{strrej}_i(a)$, $\text{border}_i(a)$) iff a is strongly accepted (resp. weakly accepted, weakly rejected, strongly rejected, borderline) by i .

Proof See “Appendix A1”. □

As mentioned, this is a fundamental step to talk about *goals of communication*, when these involve the justification status of a specific argument (the *issue* of the debate) that the speaker wants to induce in the hearer.

5 Epistemic logics for abstract argumentation

As our initial example shows, agents need to form beliefs about the awareness set of other agents, and these beliefs may be more or less accurate. Agents may also have different capacities to detect whether an argument attacks another. To reason about agents’ uncertainty we need to expand our language with epistemic modalities \Box_i , which stand for “agent i believes that” or sometimes “agent i knows that”. For reasons explained below, we do not need to choose between the two readings at this stage.

Definition 7 Formulas of the language $\mathcal{L}(\mathcal{V}_{Ag}^A, \Box)$ are given by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi \quad p \in \mathcal{V}_{Ag}^A, i \in Ag$$

Other Boolean connectives ($\vee, \rightarrow, \leftrightarrow$) and constants (\top, \perp) are defined as usual and \Diamond_i is defined as $\neg\Box_i\neg$ with the informal meaning “agent i consider epistemically possible that...”. In some axiomatisations, we will make use of the *mutual belief (knowledge) modality*, defined as $\Box_{Ag}\varphi := \bigwedge_{i \in Ag} \Box_i\varphi$, which reads “everyone in Ag believes (knows) that φ ”.

Standard Kripke-style semantics, where states are MAFs over a given set A , provides a natural interpretation of this language and allows to model uncertainty about other agents’ information and about the presence of attacks. Uncertainty is captured by the accessibility of different states. Intuitively, each state is an alternative of the *actual* MAF, based on the same pool of arguments A and the same enumeration of its subsets, but with possibly different objective attacks, and where agents may be aware of different arguments. We name them *epistemic argumentative models* and define them as follows.

¹⁶ Or equivalently, $\text{Th}_{MAF} \wedge E_k \sqsubseteq E_l$ (resp. $\text{Th}_{MAF} \wedge E_k \sqsubset E_l$) is satisfiable (precisely because it its true at v_{MAF}) iff $E_k \subseteq E_l$ (resp. $E_k \subset E_l$).

Definition 8 (Model) An *epistemic argumentative model* (\mathcal{EA} -model) for $\mathcal{V}_{\text{Ag}}^A$ is a tuple $M = (W, \mathcal{R}, V)$. Here, $W \neq \emptyset$ is a set of states, $\mathcal{R}: \text{Ag} \rightarrow \wp(W \times W)$ is a function assigning an *epistemic accessibility relation* \mathcal{R}_i to each agent $i \in \text{Ag}$, and $V: \mathcal{V}_{\text{Ag}}^A \rightarrow \wp(W)$ is a *valuation function*. We denote by $\hat{V}: W \rightarrow \wp(\mathcal{V}_{\text{Ag}}^A)$ the *dual valuation function* of V , which is defined as $\hat{V}(u) := \{p \in \mathcal{V}_{\text{Ag}}^A \mid u \in V(p)\}$. We also denote by $A_i(w) := \{a \in A \mid w \in V(\text{aw}_i(a))\}$ the *awareness set* of agent i at world w .¹⁷ Similarly we define the *set of attacks that hold at w* as $R(w) := \{(a, b) \in A \times A \mid w \in V(a \rightsquigarrow b)\}$. The valuation V should satisfy the following additional constraints:

- ER for some $w \in W$, $\hat{V}(w)$ represents an enumeration of $\wp(A)$ (see Definition 6) (*enumeration representation*);
- SU for every $a \in E_k \in \mathcal{B}$: $V(a \in E_k) = W$ or $V(a \in E_k) = \emptyset$ (*subset uniformity*).

The class of all \mathcal{EA} -models is denoted by \mathcal{EA} . When no confusion is possible we simply refer to them as models.

Condition ER guarantees that some state w in the model has an unequivocally associated $\text{MAF}_w := (A, R(w), A_1(w), \dots, A_n(w), \{E_1, \dots, E_n\})$ s.t. $\hat{V}(w) = v_{\text{MAF}_w}$. Condition SU guarantees that the enumeration of subsets is constant over the whole model. Taken together, ER and SU guarantee that every state $u \in W$ is unequivocally associated with $\text{MAF}_u = (A, R(u), A_1(u), \dots, A_n(u), \{E_1, \dots, E_n\})$, where the only elements that vary with respect to MAF_w are $R(u)$ and $A_i(u)$ for $i \in \{1, \dots, n\}$.

Given $M = (W, \mathcal{R}, V)$ we sometimes denote W by $M[W]$. A *pointed model* is a pair (M, w) where $w \in M[W]$ is a specific world representing the actual state of affairs. A pointed model for a given MAF is just a pointed model $((W, \mathcal{R}, V), w)$ such that $\hat{V}(w) = v_{\text{MAF}}$. As for the interpretation of formulas, truth in pointed models is defined recursively as usual:

Definition 9 (Truth) Given an \mathcal{EA} -model $M = (W, \mathcal{R}, V)$ and a state $w \in W$, define the relation \models as the smallest one satisfying the following clauses:

$$\begin{aligned} M, w \models p & \text{ iff } w \in V(p) \\ M, w \models \neg\varphi & \text{ iff } M, w \not\models \varphi \\ M, w \models \varphi \wedge \psi & \text{ iff } M, w \models \varphi \text{ and } M, w \models \psi \\ M, w \models \Box_i\varphi & \text{ iff } \forall u \in W \ w \mathcal{R}_i u \text{ implies } M, u \models \varphi \end{aligned}$$

Note that, given a pointed model (M, w) for MAF, it holds that $M, w \models \text{Th}_{\text{MAF}}$. Let \mathcal{C} be a class of models, a formula φ is said to be *valid in \mathcal{C}* , denoted as $\models_{\mathcal{C}} \varphi$ iff $\forall M \in \mathcal{C}, \forall w \in M[W]: M, w \models \varphi$. A formula φ is said to be \mathcal{C} -*consequence* of a set Γ , denoted as $\Gamma \models_{\mathcal{C}} \varphi$ iff $\forall M \in \mathcal{C}, \forall w \in M[W]: M, w \models \Gamma$ implies $M, w \models \varphi$.

Remark 1 (Unawareness of attacks) Note that, according to Definition 8, it is possible to build a model M with a world $w \in M[W]$ at which: 1. agent i is not aware of

¹⁷ It is worth mentioning that the awareness component of \mathcal{EA} -models is way much simpler than the one in standard awareness epistemic models (Fagin and Halpern 1987). While in the latter agents can be aware of any formula, in the former they are only aware of arguments. The fact that arguments are primitive in abstract argumentation avoids the discussion about the closure properties that awareness sets should satisfy (e.g. closed under subformulas).

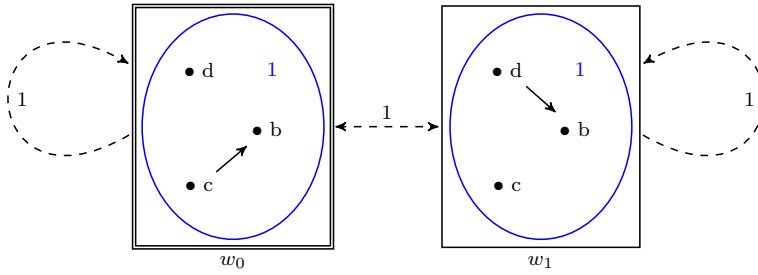


Fig. 2 A single-agent pointed \mathcal{EA} -model (M_0, w_0) for the single-agent argumentation framework $(A, R, A_1, \{E_1, \dots, E_8\})$, capturing uncertainty about the attack relation. The actual world w_0 is within a double-line frame

a (i.e. $w \notin V(\text{aw}_i(a))$) and 2. she considers possible a state u (i.e. $wR_i u$) at which $a \rightsquigarrow b$ holds (i.e. $u \in V(a \rightsquigarrow b)$). Although this could seem a defect of Definition 8, it is not. The key is that, in the intended interpretation of \mathcal{EA} -models, once sound and (locally) complete awareness of attacks (SCAA) is assumed, i is simply not aware of the attack $a \rightsquigarrow b$ (although this attack holds at u as a matter of fact). More formally, recall that, since we are assuming SCAA, we use R_i defined as $R \cap (A_i \times A_i)$ to denote the attacks that agent i is aware of in a multi-agent AF. Note that R_i can be easily captured in our object language as $a \rightsquigarrow_i b := a \rightsquigarrow b \wedge \text{aw}_i(a) \wedge \text{aw}_i(b)$, and then we have $M, u \not\models a \rightsquigarrow_i b$. However, we do not need to make this distinction explicit in our object language, since it is already captured in the syntactic definitions of solution concepts/justification status for a given agent (see complete_i , preferred_i , stracc_i , etc in page 12).

General \mathcal{EA} -models tell us very little about the constraints on agents’ awareness of arguments and attacks. Even if SCAA holds at every point, agents may still be uncertain about attacks if they are not able to distinguish between two points with radically different underlying universal frameworks, as in the following example:

Example 3 Figure 2 depicts a pointed \mathcal{EA} -model (M_0, w_0) for the single-agent argumentation framework $(A, R, A_1, \{E_1, \dots, E_8\})$, where $A = \{b, c, d\}$, $R = \{(c, b)\}$, $A_1 = \{b, c, d\}$, $E_1 = \emptyset$, $E_2 = \{b\}$, $E_3 = \{c\}$, $E_4 = \{d\}$, $E_5 = \{b, c\}$, $E_6 = \{c, d\}$, $E_7 = \{b, d\}$ and $E_8 = \{b, c, d\}$. Here, the valuation is as indicated by Fig. 2 and the enumeration, i.e. $\hat{V}(w_0) = \{c \rightsquigarrow b\} \cup \{\text{aw}_1(b), \text{aw}_1(c), \text{aw}_1(d)\} \cup \{x \in E_k \mid x \in E_k, 1 \leq k \leq 8\}$ and $\hat{V}(w_1) = \{d \rightsquigarrow b\} \cup \{\text{aw}_1(b), \text{aw}_1(c), \text{aw}_1(d)\} \cup \{x \in E_k \mid x \in E_k, 1 \leq k \leq 8\}$. Note that the valuation of attack variables is not uniform. The reader can check satisfiability of some interesting facts as $M_0, w_0 \models \neg \Box_1(c \rightsquigarrow b) \wedge \neg \Box_1(d \rightsquigarrow b) \wedge \Box_1 \text{strrej}_1(b)$. Informally, agent 1 is not sure about which argument attacks b but he knows that its justification status is strong rejection. To see that the last part of the conjunction is true, note that both MAF_{w_0} and MAF_{w_1} have a unique preferred solution i.e. $\{c, d\}$ and that, in both cases, it attacks b , hence $M_0 \models \text{strrej}_1(b)$ holds by applying Definition 4 and Proposition 1.

\mathcal{EA} -models can then be seen as minimal semantic devices for joint reasoning about argumentation and epistemic attitudes. We qualify them as minimal because they

capture no assumption about the reasoning/awareness introspection capabilities of the formalised agents. We shall mostly focus on particular subclasses of \mathcal{EA} which incrementally combine additional constraints.

Definition 10 (*Properties of models*) Let $M \in \mathcal{EA}$, $i, j \in \text{Ag}$, $w, u \in M[W]$, $a, b \in A$, and $a \rightsquigarrow b \in \mathcal{AT}$. We say that M satisfies:

- AU (*attack uniformity*) iff $V(a \rightsquigarrow b) = W$ or $V(a \rightsquigarrow b) = \emptyset$;
- PIAw (*positive introspection of awareness*) iff $w\mathcal{R}_i u$, then $A_i(w) \subseteq A_i(u)$;
- NIAw (*negative introspection of awareness*) iff $w\mathcal{R}_i u$, then $A_i(u) \subseteq A_i(w)$; and
- GNIAw (*generalized negative introspection of awareness*) iff $w\mathcal{R}_i u$, then $A_j(u) \subseteq A_i(w)$.

Condition AU amounts to assuming that attacks are the same through all the states and therefore SCAA is common knowledge (belief). PIAw and NIAw are adapted versions of the introspective properties for general awareness (Fagin and Halpern 1987). Condition PIAw dictates that if one is aware of a specific argument, then he cannot consider it possible that he is not. Conversely, NIAw amounts to saying that if one is not aware of a specific argument then he cannot think it possible that he is. They are respectively captured by axioms $\text{aw}_i(a) \rightarrow \Box_i \text{aw}_i(a)$ and $\neg \text{aw}_i(a) \rightarrow \Box_i \neg \text{aw}_i(a)$. GNIAw is a stronger constraint, saying that if one is not aware of a specific argument then he cannot think it possible that other agents are, and therefore NIAw is just a special case of GNIAw.¹⁸ GNIAw is captured by the axiom $\neg \text{aw}_i(a) \rightarrow \Box_i \neg \text{aw}_j(a)$ or, maybe more intuitively, by its contrapositive $\Diamond_j \text{aw}_j(a) \rightarrow \text{aw}_i(a)$.

We denote by \mathcal{AoA} (*awareness of arguments*) the class of all \mathcal{EA} -models satisfying AU, PIAw and GNIAw and refer to its elements as \mathcal{AoA} -models. Clearly, the one in Fig. 2 is not a \mathcal{AoA} -model. However, the class \mathcal{AoA} is general enough to subsume scenarios like our Example 1.

Example 4 Figure 3 represents a pointed \mathcal{AoA} -model (M_1, w_0) capturing the relevant epistemic features of Example 1. We assume that (M_1, w_0) is an \mathcal{AoA} -model for the MAF of Fig. 1, that is $\hat{V}(w_0) = v_{\text{MAF}}$. Again, we assume some enumeration E of the set $\wp(A)$ to be given and that the valuation of M_1 represents that enumeration. Condition AU in the definition of \mathcal{AoA} -models allows dispensing with the graphical representation of the valuation of attack variables (as far as we keep in mind what is the underlying universal framework), since attack variables are uniform throughout the model. In the case of model M_1 , depicted in Fig. 3, we assume that V matches the structure of the UAF of Fig. 1, i.e. $R(w) = \{(b, a), (d, b), (e, d), (c, b), (f, c)\}$ for every $w \in M_1[W]$. Moreover, following Schwarzenrüber et al. (2012), we represent in a compact way also the valuation of awareness variables; e.g. 1: $\{a, b, c\}$ inside the w_3 -rectangle means that $A_1(w_3) = \{a, b, c\}$ or, equivalently, that $w_3 \in V(\text{aw}_1(a))$, $w_3 \in V(\text{aw}_1(b))$, $w_3 \in V(\text{aw}_1(c))$ and $\forall x \in A \setminus \{a, b, c\}: w_3 \notin V(\text{aw}_1(x))$. The reader can check that $M_1, w_0 \models \Box_1 \text{strrej}_1(a) \wedge \Box_2 \text{strrej}_2(a)$, i.e. both agents agree about

¹⁸ Conditions PIAw and GNIAw are provided by Schwarzenrüber et al. (2012) and adapted to our notation. GNIAw makes perfect sense in a *de re* reading of the awareness operator: as soon as I attribute to someone else awareness of a fully specified item, then I should myself be aware of it. This reading is adopted in strategic argumentation (Rienstra et al. 2013). Things are different when such item is left unspecified or vague, e.g. in sentences like “I know you have some proof for theorem X and I don’t”.

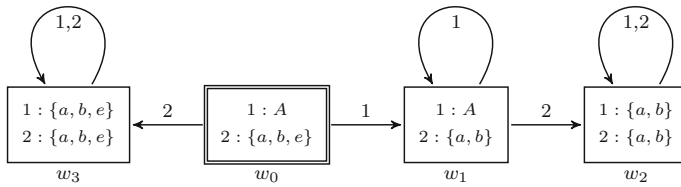


Fig. 3 A pointed $\mathcal{A}o\mathcal{A}$ -model (M_1, w_0) representing the agents’ uncertainty in Example 1. The actual world w_0 is within a double-line frame

the justification status of a . However, this agreement is based on different reasons: Agent’s 1 strong rejection is based on full awareness of the universal framework and is therefore not defeasible, while Agent’s 2 rejection is based on partial awareness and is defeasible by new information.

We did not discuss specific properties of \mathcal{R}_i so far, since we want to provide a comprehensive approach, taking into account both knowledge and belief. Moreover, there is no universal agreement about the properties of both notions.¹⁹ We do not intend to take a stand in this debate and we are content to show that the different constraints on \mathcal{R}_i do not pose any technical problem for completeness. Accordingly, given a class of models \mathcal{C} , we denote by $S4(\mathcal{C})$ (resp. $S5(\mathcal{C})$, $\mathcal{KD}45(\mathcal{C})$) the subclass of \mathcal{C} where every \mathcal{R}_i is a preorder (resp. an equivalence relation; a serial, transitive and euclidean relation).

We now provide sound and strongly complete axiomatistions for the relevant classes of models. Let us first define the corresponding proof systems:

Definition 11 (*Proof systems*)

- EA is the proof system containing all instances of Taut, K, PIS (*positive introspection of subsets*), NIS (*negative introspection of subsets*), ER (*enumeration representation*) and both inference rules from Table 1.²⁰ $S4(EA)$ (resp. $S5(EA)$, $\mathcal{KD}45(EA)$) extends EA with axioms T and 4 (resp. T, 4 and 5; D, 4 and 5) from Table 1.
- AoA (*Awareness of Arguments*) is the system extending EA with PIAt (*positive introspection of attacks*), NIAt (*negative introspection of attacks*), PIAw (*positive introspection of awareness*) and GNIAw (*generalized negative introspection of*

¹⁹ Concerning knowledge, computer scientists usually model it as an equivalence relation (Fagin et al. 2004; Meyer and van der Hoek 1995)—which informally corresponds to factive, fully introspective knowledge. On the other hand, since Hintikka (1962), philosophers have argued against some counter-intuitive consequences of assuming euclideaness—which informally corresponds to *negative introspection*, see Stalnaker (2006) for a more detailed discussion. For belief, the situation is slightly more unequivocal. It is mostly agreed that serial, transitive and euclidean relations capture the relevant features of belief—which informally corresponds to consistent and fully introspective beliefs. Nonetheless, consistency of (rational) beliefs has been questioned in some cases, as e.g. in Parikh (2008, §2).

²⁰ PIS and NIS are implied by SU. Although they capture the weaker condition that subsets are uniform along the agents’ indistinguishability relation, they are together sufficient for proving completeness.

Table 1 Axioms for \mathcal{EA} -models and \mathcal{AoA} -models capturing different constraints on \mathcal{R}_i

<i>Axioms</i>			
All propositional tautologies	(Taut)	$\Box_i(\varphi \rightarrow \psi) \rightarrow (\Box_i\varphi \rightarrow \Box_i\psi)$	(K)
$a \in E_k \rightarrow \Box_{Ag}a \in E_k$	(PIS)	$\neg a \in E_k \rightarrow \Box_{Ag}\neg a \in E_k$	(NIS)
$a \rightsquigarrow b \rightarrow \Box_{Ag}a \rightsquigarrow b$	(PIAt)	$\neg a \rightsquigarrow b \rightarrow \Box_{Ag}\neg a \rightsquigarrow b$	(NIAt)
$aw_i(a) \rightarrow \Box_iaw_i(a)$	(PIAw)	$\neg aw_i(a) \rightarrow \Box_i\neg aw_j(a)$	(GNIAw)
$\bigwedge_{1 \leq k < m \leq n} E_k \neq E_m$	(ER)		
$\neg \Box_i \perp$	(D)	$\Box_i\varphi \rightarrow \varphi$	(T)
$\Box_i\varphi \rightarrow \Box_i\Box_i\varphi$	(4)	$\neg \Box_i\varphi \rightarrow \Box_i\neg \Box_i\varphi$	(5)
<i>Rules</i>			
From $\varphi \rightarrow \psi$ and φ , infer ψ	MP	From φ infer $\Box_i\varphi$	NEC \square

Table 2 Proofs systems (first column), their corresponding class of models (second column), and the axioms they contain (third column)

L	\mathcal{C}^L	Axioms
EA	\mathcal{EA}	Taut, K, PIS, NIS, and ER
S4(EA)	$S4(\mathcal{EA})$	EA extended with T, and 4
S5(EA)	$S5(\mathcal{EA})$	S4(EA) extended with 5
KD45(EA)	$\mathcal{KD45}(\mathcal{EA})$	EA extended with D, 4, and 5
AoA	\mathcal{AoA}	EA extended with PIAt, NIAt, and PIAw, GNIAw
S4(AoA)	$S4(\mathcal{AoA})$	AoA extended with T, and 4
S5(AoA)	$S5(\mathcal{AoA})$	S4(AoA) extended with 5
KD45(AoA)	$\mathcal{KD45}(\mathcal{AoA})$	AoA extended with D, 4, and 5

Recall that all systems are closed under both MP and NEC \square

awareness).²¹ S4(AoA) (resp. S5(AoA), KD45(AoA)) extends AoA with axioms T and 4 (resp. T, 4 and 5; D, 4 and 5) from Table 1.

Let L be any of the proof systems defined above, we denote by \mathcal{C}^L the corresponding class of models according to Table 2. For instance $\mathcal{C}^{S4(\text{AoA})} = S4(\mathcal{AoA})$.

Theorem 1 *Let L be any of the proof systems defined above, then L is sound and strongly complete w.r.t. \mathcal{C}^L .*

Proof See “Appendix A2”. □

Although the details of the proof are left for the “Appendix”, some remarks are in order. Soundness results are straightforward by induction on the length of derivations, given that all axioms are valid and that rules preserve validity (in their corresponding class of models). Strong completeness will be proved using the canonical model

²¹ Again, PIAt and NIAt are two properties entailed by *attack uniformity* and suffice to prove completeness with respect to models that satisfy AU. Note, moreover, that AoA is a conservative extension of Schwarzen-truber et al. (2012)’s \mathcal{L}_1 .

technique. Note however that the canonical model for EA is not an \mathcal{EA} -model—hence this problem is inherited by every system extending EA. More concretely, SU is violated by the canonical model for EA. This inconvenience is circumvented by taking its generated submodels, which, thanks to the constraints encoded by PIS and NIS, turn out to be \mathcal{EA} -models (see Theorem 7.3. of Blackburn et al. 2002 for a similar proof). Furthermore, truth is preserved under generated submodels for our language and semantics, just as in the general modal case (Blackburn et al. 2002, Prop. 2.6.), even if we are not working with normal modal logics in the sense of Blackburn et al. (2002), because the rule of uniform substitution is not sound here.

6 Epistemic and argumentative dynamics

Standard approaches to the dynamics of AFs focus almost exclusively on changes generated by addition and deletion of arguments and/or attacks, leaving epistemic updates aside (Doutre and Mailly 2018). Here, we present a framework where both dynamics (epistemic and argumentative) are encompassed. Moreover, this framework allows reasoning about different *communication moves* and complex *information updates*. For presentational purposes, we focus on completeness results for dynamic extensions of EA, AoA, S4(AoA), KD45(AoA) and, semantically, on transformations of the corresponding classes of models. Completeness proofs and conceptual considerations concerning the dynamic extensions of other systems can be easily extrapolated and are therefore not discussed. The main technical idea of our dynamic approach is to use *event models* (Baltag et al. 2016; Baltag and Moss 2004) enriched with *propositional assignments* or *substitutions* (van Benthem et al. 2006; van Ditmarsch and Kooi 2008) to capture both kind of dynamics.²² A key notion to define these models is that of *propositional substitution*.

Definition 12 (*Substitutions*) A *propositional EA-substitution* (or an *EA-substitution*, for short) is a function $\sigma : \mathcal{V}_{Ag}^A \rightarrow \mathcal{V}_{Ag}^A \cup \{\perp, \top\}$ s.t.:

- (i) for every $p \in \mathcal{B}$ it holds that $\sigma(p) = p$ (i.e. subset variables are not substituted); and
- (ii) for every $p \in \mathcal{AT} \cup \mathcal{O}$ either $\sigma(p) = p$ or $\sigma(p) = \top$ or $\sigma(p) = \perp$.²³

We use SUB^{EA} to denote the set of all EA-substitutions, and λ to denote the identity substitution. Moreover, an AoA-substitution is an *EA-substitution* s.t.:

- (iii) for every $p \in \mathcal{AT}$ it holds that $\sigma(p) = p$ (*persistence of attacks*).

We use SUB^{AoA} to denote the set of all AoA-substitutions.

Intuitively, condition *i* ensures that the enumeration is kept fixed under update.²⁴ In the general case of EA-substitutions, condition *ii* allows to modify both awareness and

²² These structures are sometimes called *event models with factual change* [e.g. in van Benthem et al. (2006), van Ditmarsch and Kooi (2008)], but this term can be deceiving in the actual context, where the informal meaning of some propositional variables is agent-related.

²³ As shown by van Ditmarsch and Kooi (2008), this is equivalent to use more general substitutions w.r.t. $\mathcal{AT} \cup \mathcal{O}$, that is functions of the form $\sigma : \mathcal{AT} \cup \mathcal{O} \rightarrow \mathcal{L}(\mathcal{V}_{Ag}^A, \square)$.

²⁴ This will be relevant for complete axiomatisations.

attack variables. The modification of awareness variables corresponds to addition or deletion of arguments from the agents' awareness set. Modification of attack variables is of interest in order to contextualize other formalisms we deal with in Sect. 8. Since modification of attacks is not relevant for our main focus, we will mostly deal with AoA-substitutions, where this is forbidden by condition *iii*. We can also represent EA-substitutions (resp. AoA-substitutions) as maps of the form $\{p_1 \mapsto *_1, \dots, p_n \mapsto *n\}$ where for every $0 \leq k \leq n$ we have that: $p_k \in \mathcal{AT} \cup \mathcal{O}$ (resp. $p_k \in \mathcal{O}$); $*_k \in \{\top, \perp\}$; and for every $0 \leq m \leq n, k \neq m$ implies $p_k \neq p_m$. With this notion at hand, we define event models as follows:

Definition 13 (*Event model*) An \mathcal{EA} -event model (or an event model, for short) for a given language $\mathcal{L}(\mathcal{V}_{\text{Ag}}^A, \square)$ is a tuple $E = (S, \mathcal{T}, \text{pre}, \text{pos})$ where $S \neq \emptyset$ is a finite set of events; $\mathcal{T}: \text{Ag} \rightarrow \wp(S \times S)$ assigns to each agent i an indistinguishability relation \mathcal{T}_i between events (intended to represent uncertainty of agent i about which changes are happening); $\text{pre}: S \rightarrow \mathcal{L}(\mathcal{V}_{\text{Ag}}^A, \square)$ is a function assigning a precondition to each event and $\text{pos}: S \rightarrow \text{SUB}^{\text{EA}}$ assigns a substitution to each event, indicating its effect on awareness and attacks.

Given an event model $E = (S, \mathcal{T}, \text{pre}, \text{pos})$ we sometimes use $E[S]$ to denote S . A pointed event model is a pair (E, s) where $s \in E[S]$. We denote by ea the class of all \mathcal{EA} -event models. The next definition explains how \mathcal{EA} -models and event models interact through action execution.

Definition 14 (*Product update*) Given an \mathcal{EA} -model $M = (W, \mathcal{R}, V)$ and an event model $E = (S, \mathcal{T}, \text{pre}, \text{pos})$, their *product update model* is defined as $M \otimes E := (W', \mathcal{R}', V')$ where:

- $W' := \{(w, s) \mid M, w \models \text{pre}(s)\}$;
- $(w, s) \mathcal{R}'_i (w', s')$ iff $w \mathcal{R}_i w'$ and $s \mathcal{T}_i s'$; and
- $V'(p) := \{(w, s) \in W' \mid M, w \models \text{pos}(s)(p)\}$.

Informally, product update is meant to provide a new \mathcal{EA} -model where the possible states are pairs (w, s) , accessibility holds between pairs iff it holds coordinatewise and the valuation of variables is updated according to the substitution labelling s as its postcondition.

Remark 2 Note that if M is an \mathcal{EA} -model, $M \otimes E$ is not guaranteed to be an \mathcal{EA} -model, since W' might be empty. To see this, take for instance $\text{pre}(s) = \perp$ for every $s \in E[S]$. When $W' \neq \emptyset$, we say that $M \otimes E$ is *defined*.

We use the symbols \bullet, \circ, Δ to name events. Let us now look at two examples of event models.

Example 5 (*Public addition of an argument*) Let us first consider the event model for the *public addition of an argument* a , defined as $\text{Pri}_i^a := (S, \mathcal{T}, \text{pre}, \text{pos})$ where:

- $S = \{\Delta\}$,
- $\mathcal{T}_k = \{(\Delta, \Delta)\}$ for every $k \in \text{Ag}$,
- $\text{pre}(\Delta) = \top$, and

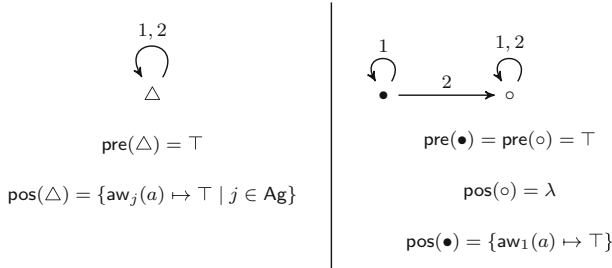


Fig. 4 Representation of Pub^a (left.) and Pri₁^a (right.) for two agents

– pos(△) = {aw_k(a) ↦ ⊤ | k ∈ Ag}.

Pub^a is graphically represented in the left-hand side of Fig. 4, for the special case where Ag = {1, 2}.

Example 6 (*Private addition of an argument*) We define the event model for *i* privately adding of an argument *a* as Pri_{*i*}^a := (S, T, pre, pos) where:

- S = {●, ○}, where ● is the action of *i* learning *a* while ○ is the “nothing happens” event;
- if *k* = *i*, then T_{*k*} = {(●, ●), (○, ○)}; else T_{*k*} = {(●, ○), (○, ○)};
- pre(●) = pre(○) = ⊤; and
- pos(●) = {aw_{*i*}(a) ↦ ⊤} and pos(○) = λ.

In this case, the definition of T captures the intuition of a completely private learning action for *i*; meaning that, after the execution of (Pri_{*i*}^a, ●) everyone (except *i*) believes that nothing has happened. Pri_{*i*}^a is pictorially represented in the right-hand side of Fig. 4 for the special case Ag = {1, 2}.

Both event models represent the same (well-studied) action of adding an argument to an argumentation framework (Cayrol et al. 2010), but DEL modelling allows to account for the distinction between public and private communication, thus adding a relevant epistemic dimension.²⁵ As an example of the product model execution, Fig. 5 illustrates the operation M ⊗ Pub^d, that we discuss in Example 7.

More in general, given a set of arguments *B*, the public addition of the whole set is captured by the action Pub^B, which only modifies the definition of Pub^a in that pos(△) := {aw_{*j*}(*b*) ↦ ⊤ | *b* ∈ *B*, *j* ∈ Ag}. Analogously, private addition of *B* by *i* is Pri_{*i*}^B and works as in Fig. 4 (right) with pos(●) := {aw_{*i*}(*b*) ↦ ⊤ | *b* ∈ *B*}.

The effects of updating EA-models with actions are described by the following dynamic languages:

Definition 15 (*Dynamic languages*) Let V_{Ag}^A be a set of propositional variables, and let ★ ⊆ ea be a class of event models for L(V_{Ag}^A, □). The formulas of the language

²⁵ These actions can also be seen as special cases of the *consider* action of van Benthem and Velázquez-Quesada (2010).

$\mathcal{L}^*(\mathcal{V}_{\text{Ag}}^A, \square)$ (or simply \mathcal{L}^* when the context is clear) are given by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \square_j\varphi \mid [E, s]\varphi \quad p \in \mathcal{V}_{\text{Ag}}^A, j \in \text{Ag}, E \in \star, s \in E[S]$$

where $[E, s]$ reads: “after executing (E, s) , φ holds”.²⁶ We extend the truth relation for the new kinds of formulas as follows:

$$M, w \models [E, s]\varphi \quad \text{iff} \quad M, w \models \text{pre}(s) \text{ implies } M \otimes E, (w, s) \models \varphi$$

The flexibility of event models is well-known. In their actual epistemic-argumentative reading, they can be used to model the effects of acts of *information exchange*. As mentioned, these acts have two sides: (i) how the hearer decides to update her knowledge base with new information (*information update*) and (ii) what argument(s) the speaker decides to communicate in order to fulfill his goal (*communication moves*). In order to persuade their interlocutors, smart players choose (ii) based on their expectations about (i). Let us now illustrate this through the simplest combination of (i) and (ii) of Example 1. A deeper analysis is left for Sect. 7.

Example 7 We assume that (i) is as follows: Mom behaves credulously. This means that whenever Charlie communicates an argument, she simply adds it to her knowledge base. This is modelled through the event model for public addition of an argument (left-hand-side of Fig. 4). As for (ii), we assume that Charlie thinks that Mom is indeed behaving credulously. Recall that Charlie has three options: communicating c , communicating d or communicating $\{c, d\}$. Hence, his way of selecting the best set of arguments to communicate consists in reasoning about the effects of all options. Note that although one of the three moves will not work ($M_1, w_0 \models [\text{Pub}^d, \Delta]\text{strrej}_2(a)$, see Fig. 5), Charlie thinks that they are equally good $M_1, w_0 \models \square_1([\text{Pub}^c, \Delta]\text{stracc}_2(a) \wedge [\text{Pub}^d, \Delta]\text{stracc}_2(a) \wedge [\text{Pub}^{\{c,d\}}, \Delta]\text{stracc}_2(a))$. Therefore, under these assumptions, the success of Charlie is a simple matter of luck.

Remark 3 Here communication of an argument x to everybody is modelled by the operation of public addition and not, as common in DEL, as the *public announcement* of the formula $\text{aw}_i(x)$ (where i the speaker) or of $\bigwedge_{i \in \text{Ag}}(\text{aw}_i(x))$.²⁷ The usual event model for public announcement of a formula φ is based on the same single-event structure of Fig. 4 (left), but with φ (instead of \top) as precondition and with no postconditions. If we did so, agents could never learn arguments whose (collective) awareness is not considered as a doxastic possibility before communication takes place; but this fails to capture what actually happens in most real-life debates. For instance, if we modelled communication of d by Charlie as the public announcement of $\text{aw}_1(d)$ (resp. as the

²⁶ Note that each dynamic language is parametrised not only by Ag and A but also by \star . This is useful when defining *action-friendly* logics, in the sense of Baltag and Renne (2016, Sect. 3.3.). Moreover, we fixed the range of pre in Definition 13, so that our event models always have static formulas as preconditions. This is by no means an essential limitation of the current framework, but rather a simplifying assumption for presentation purposes. The interested reader is referred to van Ditmarsch et al. (2007, Sect. 6) and Baltag and Renne (2016, Appendix H) for further information.

²⁷ See van Ditmarsch et al. (2007, sections 4 and 6).

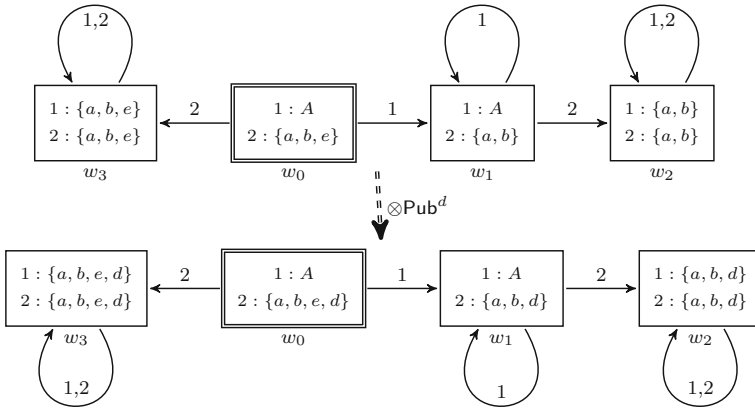


Fig. 5 Product update $M_1 \otimes \text{Pub}^d$ capturing the effects of Charlie communicating d and Mom updating her AF credulously. See Example 4 for an explanation of the graphical representation of \mathcal{AoA} -models

public announcement $\text{aw}_1(d) \wedge \text{aw}_2(d)$ in the previous example, then the beliefs of Mom (resp. everyone) would become inconsistent with no apparent reason.

We give axiomatisations and prove completeness for the dynamic extensions of EA, AoA, $S4(\text{AoA})$ and $KD45(\text{AoA})$.²⁸ For this we use reduction axioms and an *inside-out reduction* (as described e.g. in Wang and Cao 2013). That is to say, we don’t use axioms for event model composition but we show how to eliminate all dynamic operators starting from their innermost occurrences. To do so, we need to prove that the rule of substitution of proven equivalents is sound w.r.t. all the systems considered. From a semantic perspective, this implies showing that the class of models we are working with is closed under product update. It is easy to show that this is in general not the case for \mathcal{AoA} , $S4(\mathcal{AoA})$ and $S5(\mathcal{AoA})$. One of the possible solutions to this shortcoming is to restrict the class of “allowed” event models, so as to ensure that we remain in the targeted class after the execution of the product update.

The general case of updating \mathcal{EA} models with \mathcal{EA} event models does not present any problem. Indeed, $a \in E_k$ -variables do not change their truth values by the constraint i of Definition 12, and this guarantees that *subset uniformity* (SU) and *enumeration representation* (ER) are trivially preserved. Updating an \mathcal{AoA} -model with an event model that only uses AoA -substitutions guarantees further that *attack uniformity* (AU) is preserved by the constraint *iii* of Definition 12, since \rightsquigarrow -variables are also left untouched. The problem for updating \mathcal{AoA} -models lies in the awareness constraints PIAw and GNIAw . We can however provide a set of sufficient conditions for their preservation. For this we need to introduce some additional notation. Let E be an event model and let $s \in E[S]$; define $\text{pos}_i^+(s) := \{a \in A \mid \text{pos}(s)(\text{aw}_i(a)) = \top\}$ and $\text{pos}_i^-(s) := \{a \in A \mid \text{pos}(s)(\text{aw}_i(a)) = \perp\}$. Informally, $\text{pos}_i^+(s)$ (resp. $\text{pos}_i^-(s)$) denotes the set of arguments gained (resp. lost) by i as a consequence of executing s . Furthermore, let E be an action model, we say that E satisfies:

²⁸ As mentioned, we omit completeness proofs for the dynamic extensions of $S4(\text{EA})$, $KD45(\text{EA})$ and $S5(\text{EA})$ and $S5(\text{AoA})$, since they are analogous to the ones we are about to present and thus lack technical interest.

- EM_1 iff for all $s, t \in E[S]$: if $s\mathcal{T}_i t$, then $\text{pos}_i^+(s) \subseteq \text{pos}_i^+(t)$ and $\text{pos}_i^-(t) \subseteq \text{pos}_i^-(s)$; and
- EM_2 iff for all $s, t \in E[S]$: if $s\mathcal{T}_i t$, then $\forall j \in \text{Ag}$: $\text{pos}_i^-(s) \subseteq \text{pos}_j^-(t)$ and $\text{pos}_j^+(t) \subseteq \text{pos}_i^+(s)$.

Let us explain these conditions informally. In an event model satisfying EM_1 , if we suppose that s is the event that actually happens, then EM_1 implies that any event t that agent i cannot tell from s is one where he gains at least the same new arguments and does not lose any argument he actually keeps. It is intuitive to see that EM_1 preserves $PIAw$. Indeed, suppose that i is aware of a after the execution of s (antedecedent $\text{aw}_i(a)$ of $PIAw$). Two things are possible. Either a is a newly acquired argument (by the execution of s). Then, since any state accessible after the update is “filtered” by some indistinguishable event t , the condition $\text{pos}_i^+(s) \subseteq \text{pos}_i^+(t)$ forces a to be acquired at that state too, and therefore the consequent $\Box_i \text{aw}_i(a)$ is satisfied. Or else, i was already aware of a before the execution of s , and therefore he has not lost it. Here $\text{pos}_i^-(t) \subseteq \text{pos}_i^-(s)$ guarantees that a is not lost at any state accessible after the execution of the event. An analogous informal reading, generalized to other agents, can be given for EM_2 : at any indistinguishable event any other agent loses at least the same arguments as i and gains no more. By the same pattern as before we can extrapolate that this condition preserves $GNIAw$ (see Lemma 1 for a detailed proof).

Let us now define some relevant classes of event models:

Definition 16 (*Classes of event models*) We denote by em12 , emS4 and pure the following classes of event models:

- em12 is the class of event models satisfying EM_1 , EM_2 and assigning AoA -substitutions (see Definition 12) to all their events. In other words, $E = (S, \mathcal{T}, \text{pre}, \text{pos}) \in \text{em12}$ iff E satisfies EM_1 , EM_2 and $\text{pos} : S \rightarrow \text{SUB}^{AoA}$.
- emS4 is the subclass of em12 where every \mathcal{T}_i is a preorder.
- pure is the subclass of em12 s.t. $\text{pre}(s) = \top$ for every $s \in E[S]$ and every \mathcal{T}_i of E is serial, transitive and euclidean.²⁹

Remark 4 Note that both Pub^a and Pri_i^a (see Examples 5, 6 and Fig. 4) are purely argumentative event models (i.e. they belong to pure) and, a fortiori, they also belong to em12 .

We can then prove the following result:

Lemma 1 (*Closure*) Let $M = (W, \mathcal{R}, V)$ be an \mathcal{EA} -model and let $E = (E, \mathcal{T}, \text{pre}, \text{pos})$ be an event model, then:

- (i) If $M \otimes E$ is defined, then $M \otimes E \in \mathcal{EA}$.
- (ii) If $M \in AoA$, $E \in \text{em12}$, and $M \otimes E$ is defined, then $M \otimes E \in AoA$.
- (iii) If $M \in S4(AoA)$, $E \in \text{emS4}$, and $M \otimes E$ is defined, then $M \otimes E \in S4(AoA)$.
- (iv) If $M \in KD45(AoA)$, and $E \in \text{pure}$, then $M \otimes E \in KD45(AoA)$.

Proof See “Appendix A3”. □

²⁹ The name pure , standing for “purely argumentative events”, is due to the fact that its elements involve neither epistemic nor factual preconditions.

Table 3 Reduction axioms for dynamic systems

<i>Axioms</i>	
$[E, s]p \leftrightarrow (\text{pre}(s) \rightarrow \text{pos}(s)(p))$	(Atm)
$[E, s]\neg\varphi \leftrightarrow (\text{pre}(s) \rightarrow \neg[E, s]\varphi)$	(Neg)
$[E, s](\varphi \wedge \psi) \leftrightarrow ([E, s]\varphi \wedge [E, s]\psi)$	(Con)
$[E, s]\Box_i\varphi \leftrightarrow (\text{pre}(s) \rightarrow \bigwedge_s \mathcal{T}_{it} \Box_i[E, t]\varphi)$	(Box)
<i>Rules</i>	
From $\varphi \leftrightarrow \psi$, infer $\delta \leftrightarrow \delta[\varphi/\psi]$	SE

In rule SE, $\delta[\varphi/\psi]$ is the result of substituting one or more occurrences of φ in δ by ψ . Furthermore, it is assumed for simplicity that substitutions do not affect the occurrences of formulas inside dynamic modalities, i.e. $([E, s]\delta)[\varphi/\psi] := [E, s]\delta[\varphi/\psi]$

Remark 5 (The general value of EM_1 and EM_2) If we dispense with the argumentation/awareness interpretation of the current formalism, Lemma 1(ii) tells us that we can look at EM_1 and EM_2 as general, sufficient conditions that guarantee the preservation of certain constraints over propositional valuations after product update. Therefore, they can be reused in any framework including event models and indexed operators (awareness operators, in our case) ranging over atomic entities (arguments, in our case). As suggested before, $PIAw$ and $GNIAw$ characterize a *de re* reading of operators ranging over atomic entities. Therefore, EM_1 and EM_2 are structural event constraints that, taken together, work as a sufficient condition to preserve these *de re* operators.

Lemma 1 allows to prove the following general preservation result:

Lemma 2 (Validity preservation) *All axioms instances of Table 3 written in \mathcal{L}^{ea} (resp. \mathcal{L}^{em12} , \mathcal{L}^{emS4} , \mathcal{L}^{pure}), are valid in \mathcal{EA} (resp. AoA , $S4(AoA)$, $KD45(AoA)$) and all applications of SE in \mathcal{L}^{ea} (resp. \mathcal{L}^{em12} , \mathcal{L}^{emS4} , \mathcal{L}^{pure}) preserves validity in \mathcal{EA} (resp. AoA , $S4(AoA)$, $KD45(AoA)$).*

Proof See “Appendix A3”. □

General completeness results follow from Lemma 2. Let us first define the targeted axiom systems:

Definition 17 (Dynamic axiom systems)

- EA^{ea} extends EA with all axioms schemes and rules of Table 3 that can be written in \mathcal{L}^{ea} (see Definitions 15 and 16).
- AoA^{em12} extends AoA with all axioms schemes and rules of Table 3 that can be written in \mathcal{L}^{em12} .
- $S4(AoA)^{emS4}$ extends $S4(AoA)$ with all axioms schemes and rules of Table 3 that can be written in \mathcal{L}^{emS4} .
- $KD45(AoA)^{pure}$ extends $KD45(AoA)$ with all axioms schemes and rules of Table 3 that can be written in \mathcal{L}^{pure} .

Theorem 2 *The proof system EA^{ea} (resp. AoA^{em12} , $S4(AoA)^{emS4}$, $KD45(AoA)^{pure}$) is sound and strongly complete w.r.t. \mathcal{EA} (resp. AoA , $S4(AoA)$, $KD45(AoA)$).*

Proof See “Appendix A3”. □

Let us remark that in the case of KD45(AoA) our completeness result is restricted to event models belonging to *pure*. Although *pure* is a rather simple class of event models, all the actions used in our analysis of Example 7, as well as the one that will be used in the next section, fall into it. Unfortunately, modelling certain complex scenarios requires mixing purely argumentative actions with other types—for instance, public and private announcement of formulas, where preconditions are not trivial. One last axiomatisation, inspired by the works of Balbiani et al. (2012) and Aucher (2008), aims to fill this gap. The interested reader can find it in the “Appendix A4” (Theorem 3). The axiomatisation is based on an modal language with a global modality. Interestingly, this more expressive language also allows to provide necessary and sufficient conditions for the preservation of PIAw and GNIAw under product update.

7 Modelling persuasion, sceptic updates and conditional trust

In Example 7 of the previous section we unfolded the dynamics of our running example by assuming that Mom was open to accept whatever Charlie says at face value. In a more likely scenario this does not happen: Mom will filter the information received from Charlie, precisely because she does not trust him in such circumstances. Yet Charlie would still be confident, as kids often are, to be fooling her. Although Mom is not immediately aware of the counter-argument against the *Pscience* publication, she can obtain it after a quick (private) search on *Pscience*'s website. It is important to stress that Mom does not discard argument c , she rather accepts it, but eventually finds out the counterargument f . This is a rather common mechanism of epistemic vigilance, one of the kind we mentioned in Sect. 2. One possible way of capturing this epistemic action in our framework is what we call a *sceptic update* Scp_j^x , where the recipient j of an argument x privately and non-deterministically learns an attacker of x (if any). When our language contains Pub and Pri, it is possible to define a modality $[\text{Scp}_j^x]\varphi$, expressing that φ holds after j performs a sceptic update upon receiving argument x ³⁰:

$$[\text{Scp}_j^x]\varphi := [\text{Pub}^x, \Delta] \left(\bigwedge_{y \in A} (y \rightsquigarrow x \rightarrow [\text{Pri}_j^y, \bullet]\varphi) \right).$$

As an example, the bottom part of Fig. 6 represents the outcome of Mom's sceptic update as the result of two consecutive actions—a public addition of c followed by Mom privately learning f —on the initial $\mathcal{A}\mathcal{O}\mathcal{A}$ -model M_1 (Fig. 6 (top part)). In our model, Charlie thinks that he has succeeded $M_1, w_0 \models [\text{Scp}_2^c]\Box_1 \text{stracc}_2(a)$, while actually he has not $M_1, w_0 \not\models [\text{Scp}_2^c]\neg \text{stracc}_2(a)$ and, moreover, agent 2 (Mom) believes all this $M_1, w_0 \models [\text{Scp}_2^c]\Box_2(\Box_1 \text{stracc}_2(a) \wedge \neg \text{stracc}_2(a))$.

³⁰ From a semantic perspective, Scp_j^x is an action encoded as the multi-pointed event model $\cup_{y: xRy} ((\text{Pub}^x, \Delta); (\text{Pri}_j^y, \bullet))$, where \cup and $;$ respectively stand for non-deterministic choice and for sequential composition. See van Ditmarsch and Kooi (2008) for the precise meaning of both operators.

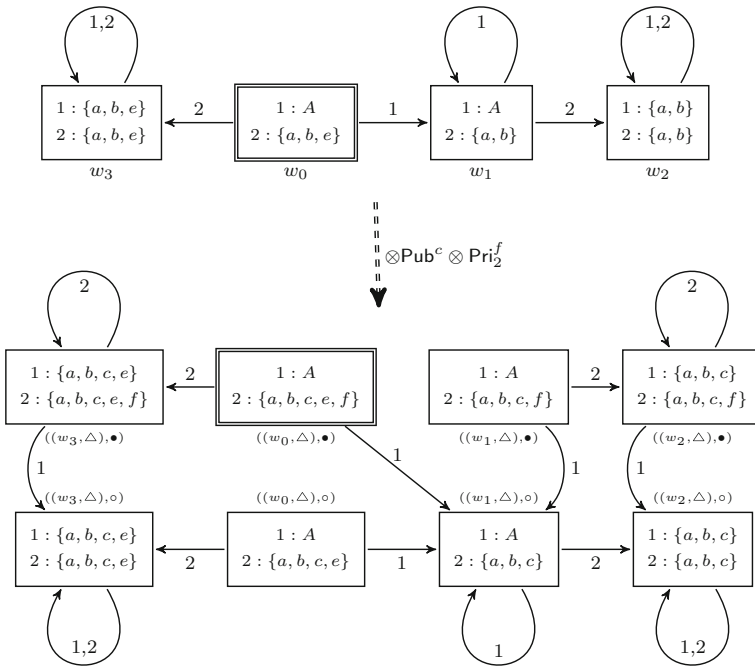


Fig. 6 Model M (top part) and model $(M \otimes \text{Pub}^c) \otimes \text{Pri}_2^f$ (bottom part)

We now have two scenarios with substantially different outcomes. The question is how close they reflect the typical behaviour of players in a more or less adversarial exchange. In both cases, we assumed that Charlie is confident that Mom will accept everything he says without further inquiry. Is Charlie the prototype of a skilled debater? Clearly not: He still lacks some mindreading and the subtlety of anticipating easy counterobjections, those skills that kids typically learn to use at an advanced stage of their cognitive development, and after a lot of trial and error. We further assumed that Mom has full trust in the first scenario and absolute distrust in the second. Distrust is driven by epistemic vigilance, but circumstances are not always black or white. After all, there are cases in which she has the right—or even the educational duty—to trust her kid.

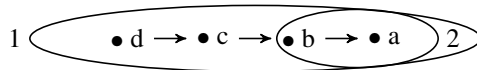
We want to put our finger on the fact that trust is most of the time mixed, and it is such in a relevant sense. Not only it varies with the source of information—Mom may trust Charlie and not Dad—or the type of information we get from the source—she may trust Charlie more or less depending on the matter at stake. Trust is often also conditional on the epistemic circumstances we find ourselves in, all other things being equal. In order to see this clearly, we introduce a different example, which we borrow from Kagemusha, a famous film directed by Akira Kurosawa.

Example 8 (Kagemusha) The warlord of the clan Takeda has been killed unbeknownst to everybody except for the members of his clan and his political decoy (kagemusha). It is vital that the warlord’s death stay secret and that his double keeps playing his

role. Therefore, everybody outside the clan must be persuaded that a : “the warlord is alive”. The warlord’s funeral is then performed anonymously and in a peculiar way: a jar with the ashes is launched into the lake Suwa on a raft.

Unfortunately, spies from rival clans are around and, by snooping on this strange ritual, they start suspecting the truth, that is they are provided with an evidential argument b that rebuts a . Now, by accident, the spies are spied, in turn, by the kagemusha, who reports this to the rest of the clan. The clan then decides to bake up an alternative (false) explanation of the ritual—an offering of sake to the god of the lake—and to tell it around. This alternative explanation c undercuts the second and reinstates the first. This has the effect of persuading the spies that they were wrong, that the ritual was indeed not a funeral and the warlord is still alive.

As things stand, argument c is *de facto* undermined by a decisive argument d , to the effect that c does not hold water. But the spies do not find access to d , and the clan’s strategy has its effect of persuading them that a is reinstated and therefore acceptable. The following MAF captures the situation right after the spies have observed the funeral, where agent 1 represents Takeda’s clan and agent 2 represents the spies:



What is clear from the story is that the spies would never have accepted the fake explanation c at face value had they only suspected that the clan was aware of being spied. Instead, they would have easily resorted to d by performing a sceptic update. The only difference between success and failure lies in the initial epistemic state of the agents involved, as the modelling in Fig. 7 shows. In the first scenario (captured in model M_2) 2 believes that 1 believes that her goal is already achieved (2 is only aware of a at w_2), i.e. $M_2, w_0 \models \Box_2 \Box_1 \text{stracc}_2(a)$, while in the second case (captured in model M'_2) 2 believes that 1 believes that it is not (2 is aware of a and b at w_2), i.e., $M'_2, w_0 \not\models \Box_2 \Box_1 \text{stracc}_2(a)$.³¹

What is also clear is that the different attitude displayed by the spies in the alternative situations does not depend on the source—which is the same—nor on the subject matter—again, the same. It is fair to say that the trust they put in the information received is part of one and the same *conditional plan* for updating information. They sceptically process the information received if they *believe that the clan believes that its goal is not achieved yet but will be after communicating c* ; otherwise they uncritically accept it. This condition can be generally defined as

$$CSU_{i,j}^x := \Box_j \Box_i (\neg \text{goal}_i \wedge [\text{Pub}^x, \Delta] \text{goal}_i),$$

where i is the speaker, $\text{goal}_i \in \mathcal{L}(\mathcal{V}_{Ag}^A, \Box)$ is its goal, j is the hearer, and x is the communicated argument. The clan, on its side, is well aware of this: they know they can get away with a fake only because, given the circumstances, epistemic vigilance is defused.

³¹ Again, valuations V_2 and V'_2 are assumed to respect some enumeration of $\wp(\{a, b, c, d\})$. This information is left implicit.

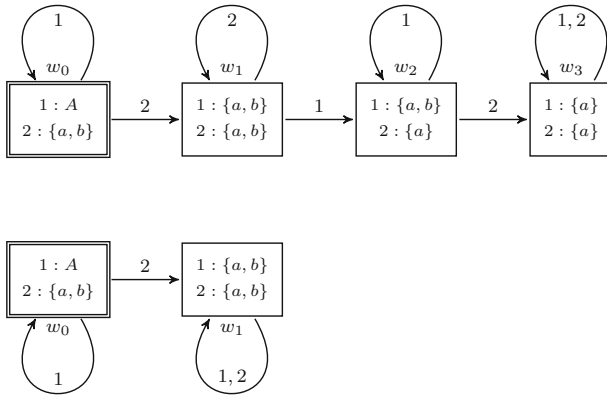


Fig. 7 Kagemusha, scenarios M_2 and M'_2 . See Example 4 for an explanation of the graphical representation of $\mathcal{A}o\mathcal{A}$ -models

It is possible to reason about the effects of conditional plans like the above in our language by defining more complex modalities like the following one that captures the effects of this kind of strategic update:

$$[\text{Str}_j^x]\varphi := (\neg\text{CSU}_{i,j}^x \rightarrow [\text{Pub}^x, \Delta]\varphi) \wedge (\text{CSU}_{i,j}^x \rightarrow [\text{Scp}_j^x]\varphi),$$

where i is the speaker, j is the hearer, and x is the communicated argument. Note that $M_2, w_0 \models [\text{Str}_2^a]\text{stracc}_2(a)$ but $M'_2, w_0 \models \neg[\text{Str}_2^a]\text{stracc}_2(a)$. This kind of operations has received attention in semantically oriented belief revision (see e.g. Rodenhäuser 2014, §2.6.1 and the definition of *mixed doxastic attitude*). A throughout analysis of the subtleties of strategic communication seems to require powerful tools of analysis akin to those currently developed in the area of epistemic planning (Andersen et al. 2012). This investigation goes beyond the scope of our paper and we leave it for future research.

8 Relation to other formalisms

Recently, uncertainty about AFs has been modelled through quantitative methods (Li et al. 2011) and qualitative ones within the formal argumentation community. Among the qualitative approaches the use of *incomplete argumentation frameworks* (Coste-Marquis et al. 2007; Baumeister et al. 2018a,b) and *control argumentation frameworks* (Dimopoulos et al. 2018) has been prominent. Also, opponent modelling in strategic argumentation (Oren and Norman 2009) has been endowed with higher order uncertainty about adversaries (Rienstra et al. 2013). Our logic can be naturally connected to these three lines of research.

8.1 Incomplete AFs

General models of incompleteness in abstract argumentation (Baumeister et al. 2018b) capture uncertainty by extending standard AFs with *uncertain arguments* $A^?$ and *uncertain attacks* $R^?$. Their formal definition is as follows.

Definition 18 (*Incomplete AF and completions* Baumeister et al. 2018b) An incomplete AF is a tuple IAF = $(A, A^?, R, R^?)$ s.t. $R, R^? \subseteq (A \cup A^?) \times (A \cup A^?)$, $A \cap A^? = \emptyset$ and $R \cap R^? = \emptyset$. (A, R) is called the *definite part* of IAF while $(A^?, R^?)$ is called the *uncertain part* of IAF.

A *completion* of IAF is any pair (A^*, R^*) s.t.:

- $A \subseteq A^* \subseteq (A \cup A^?)$; and
- $R_{|A^*} \subseteq R^* \subseteq (R \cup R^?)_{|A^*}$ where $R_{|A^*} := R \cap (A^* \times A^*)$.

Completions can be seen as possible ways of removing uncertainty by making some arguments and attacks definite. Here, the constraint on R^* entails that definite attacks between a and b must be present in all completions where both a and b are present.

Classic computational problems for AFs, such as *sceptical* or *credulous acceptance* are easily generalized to incomplete AFs. As an example, consider two generalizations of the classic preferred reasoning tasks as given in Baumeister et al. (2018a):

Pr-Possible–Sceptical–Acceptance (Pr-PSA)	
Given:	An incomplete argumentation framework IAF = $(A, A^?, R, R^?)$ and an argument $a \in A$
Question:	Is it true that there is a completion $F^* = (A^*, R^*)$ of $(A, A^?, R, R^?)$ s.t. for all $E \in \text{Pr}(F^*)$, $a \in E$?
Pr-Necessary–Credulous–Acceptance (Pr-NCA)	
Given:	An incomplete argumentation framework IAF = $(A, A^?, R, R^?)$ and an argument $a \in A$.
Question:	Is it true that for each completion $F^* = (A^*, R^*)$ of $(A, A^?, R, R^?)$, there is a $E \in \text{Pr}(F^*) : a \in E$?

The Pr-Necessary–Sceptical–Acceptance and Pr-Possible–Credulous–Acceptance problems are obtained by changing quantifiers of the definitions above in an obvious way. Similarly, Pr can be replaced by any other solution concept. It is not difficult to show that the set of completions of an IAF is a single-agent $S5(\mathcal{EA})$ -model in disguise, where $A \cup A^?$ is the underlying pool of arguments. This has the effect that the above computational problems can be regarded as *model-checking* problems in our framework. Let us make this claim more precise.

8.1.1 From incomplete AFs to \mathcal{EA} -models

Given an incomplete argumentation framework IAF = $(A, A^?, R, R^?)$, we can build a single-agent \mathcal{EA} -model to reason about IAF using our object language. First, we

fix some enumeration of $\wp(A \cup A^?) = \{E_1, \dots, E_n\}$. Then, we define the set of propositional variables associated to IAF as $\mathcal{V}^{\text{IAF}} = \mathcal{V}_{\{1\}}^{A \cup A^?}$. Since we have only one agent, we remove subindices from awareness and epistemic operators. We can then provide the following definition:

Definition 19 Let $\text{IAF} = (A, A^?, R, R^?)$ be given, the $\mathcal{S5}(\mathcal{EA})$ -model associated to IAF (for the enumeration of $\wp(A \cup A^?) = \{E_1, \dots, E_n\}$) is the model

$$M^{\text{IAF}} := (W^{\text{IAF}}, \mathcal{R}^{\text{IAF}}, V^{\text{IAF}})$$

where

- $W^{\text{IAF}} := \{w^{(A^*, R^*)} \mid (A^*, R^*) \text{ is a completion of IAF}\}$,³²
- $\mathcal{R}^{\text{IAF}} := W^{\text{IAF}} \times W^{\text{IAF}}$;
- V^{IAF} is defined for each kind of variables as follows:
 - $V^{\text{IAF}}(\text{aw}(x)) = \{w^{(A^*, R^*)} \mid x \in A^*\}$,
 - $V^{\text{IAF}}(x \rightsquigarrow y) = \{w^{(A^*, R^*)} \mid (x, y) \in R^*\}$,
 - for every k such that $1 \leq k \leq n$: $V^{\text{IAF}}(x \in E_k) = W$ if $x \in E_k$, and
 - for every k such that $1 \leq k \leq n$: $V^{\text{IAF}}(x \in E_k) = \emptyset$ if $x \notin E_k$.

The above reduction allows to obtain the following result:

Proposition 2 Let $\text{IAF} = (A, A^?, R, R^?)$, $\wp(A \cup A^?) = \{E_1, \dots, E_n\}$, M^{IAF} be the $\mathcal{S5}(\mathcal{EA})$ -model associated to IAF (for the enumeration of $\wp(A \cup A^?) = \{E_1, \dots, E_n\}$), and let $w \in M^{\text{IAF}}[W]$. We have that:

- The answer to Pr-PSA with input IAF and $a \in A$ is yes iff $M^{\text{IAF}}, w \models \Diamond \text{stracc}(a)$.³³
- The answer to Pr-NCA with input IAF and $a \in A$ is yes iff $M^{\text{IAF}}, w \models \Box \bigvee_{1 \leq k \leq n} (\text{preferred}(E_k) \wedge a \in E_k)$.

Proof See ‘‘Appendix A5’’. □

In other words, the main reasoning problems about incomplete AFs can be reduced to model-checking problems in our framework.³⁴

8.1.2 From \mathcal{EA} -models to incomplete AFs

In the opposite direction, we can easily transform members of a specific class of \mathcal{EA} -models into incomplete AFs, with a sound and systematic way to associate states to completions. This is provided by the following definition.

Definition 20 Let $M = (W, \mathcal{R}, V)$ be a total $\mathcal{S5}(\mathcal{EA})$ -model, that is an $\mathcal{S5}(\mathcal{EA})$ -model s.t. $\mathcal{R} = W \times W$, for $\mathcal{V}_{\{1\}}^C$ where C is any finite, non-empty set of arguments,

³² See Definition 18 for the notion of completion.

³³ See p. 12 for the meaning of stracc.

³⁴ The above result is just formulated for two reasoning problems and preferred solutions for the sake of brevity, but it can be easily generalised to all other acceptance problems concerning any solution concept.

and such that V represents the enumeration $\wp(C) = \{E_1, \dots, E_n\}$. We define the *incomplete argumentation framework associated to M* as the tuple

$$\text{IAF}_M = (A_M, A_M^?, R_M, R_M^?)$$

where

- $A_M = \{x \in C \mid V(\text{aw}(x)) = W\}$;
- $A_M^? = \{x \in C \mid V(\text{aw}(x)) \neq W, V(\text{aw}(x)) \neq \emptyset\}$;
- $R_M = \{(x, y) \in C \times C \mid V(\text{aw}(x)) \cap V(\text{aw}(y)) \subseteq V(x \rightsquigarrow y)\}$; and
- $R_M^? = \{(x, y) \in C \times C \mid V(\text{aw}(x)) \cap V(\text{aw}(y)) \cap V(x \rightsquigarrow y) \neq \emptyset\} \setminus R_M$.

By definition, we have that $A_M \cap A_M^? = \emptyset$ and $R_M \cap R_M^? = \emptyset$, therefore IAF_M is an incomplete AF.³⁵ Moreover, we can associate a directed graph (A_w^*, R_w^*) to each state $w \in M[W]$, where $A_w^* := A(w)$ and $R_w^* := R(w)|_{A(w)}$. It is almost immediate to check that each (A_w^*, R_w^*) is a completion of IAF_M .

Given a total $S5(\mathcal{EA})$ -model $M = (W, \mathcal{R}, V)$ and its associated IAF_M , we say that the valuation V exhausts the completions of IAF_M iff for each completion (A^*, R^*) of IAF_M , there is a state $u \in M[W]$ s.t. $(A^*, R^*) = (A_u^*, R_u^*)$.³⁶ Under this restriction, we can prove the following correspondence result analogous to Proposition 2.

Proposition 3 *Let M be a total $S5(\mathcal{EA})$ -model for $\mathcal{V}_{\{1\}}^C$ whose valuation represents the enumeration $\wp(C) = \{E_1, \dots, E_n\}$ and exhausts the completions of IAF_M , and let $w \in M[W]$, then:*

- $M, w \models \Diamond \text{stracc}(a)$ iff the answer to Pr-PSA with input IAF_M and $a \in A_M$ is yes.
- $M, w \models \Box \bigvee_{1 \leq k \leq n} (\text{preferred}(E_k) \wedge a \in E_k)$ iff the answer to Pr-NCA with input IAF_M and $a \in A_M$ is yes.

Proof See ‘‘Appendix A5’’. □

Remark 6 (*AF spaces*) Interestingly, if we drop the exhaustive valuation requirement, we obtain a one-to-one association from total $S5(\mathcal{EA})$ -models to a more general class of structures, that we call *AF spaces* and which is worth of interest. An *AF space* is a pair $(\text{IAF}, \mathcal{X})$ where \mathcal{X} is any set of completions of IAF . Incomplete AFs can be seen as a special case of AF spaces (those for which \mathcal{X} is maximal w.r.t. set inclusion). The converse, however, does not hold. As an example, consider the set of completions associated to the worlds of the $S5(\mathcal{EA})$ -model depicted in Fig. 2, i.e. $\{\{b, c, d\}, \{(c, b)\}, \{(b, c, d), \{(b, d)\}\}\}$. It is easy to show that there is no IAF with such a set of completions. We can obviously redefine the main acceptance problems for AF spaces. As an example, the following is the variant of the Pr-PSA:

³⁵ According to Definition 20, it is possible that $A_M \cup A_M^? = \emptyset$, but this is also not excluded by Baumeister et al. (2018b, Definition 22) or Baumeister et al. (2018a, Definition 4). If we require $A_M \neq \emptyset$, as done by Baumeister et al. (2018b, Definition 4) then the underlying \mathcal{EA} -model must satisfy that $V(\text{aw}(x)) = W$ for some $x \in C$.

³⁶ As the familiar reader may notice, there is a tight connection between this restriction on Kripke models and the logic of visibility as presented, for instance, in Herzig et al. (2018). We leave a more detailed analysis for future work.

Pr- \mathcal{X} -Possible-Sceptical-Acceptance (Pr- \mathcal{X} PSA)	
Given:	An AF space $((A, A^?, R, R^?), \mathcal{X})$ and an argument $a \in A$.
Question:	Is it true that there is a completion $F^* \in \mathcal{X}$ s.t. for all $E \in \text{Pr}(F^*), a \in E$?

Intuitively, AF spaces drop the assumption that the agent perceives $(A^?, R^?)$ as *completely uncertain*, meaning that all combinations of its elements are possible (as far as they are completions). This can be seen as an inconvenience for some modelling processes, since uncertainty does not need to have that level of homogeneity in many real-life argumentative scenarios.

Interestingly, within the class of total $S5(\mathcal{EA})$ -models we can isolate two subclasses corresponding to the specific types of IAFs most discussed in the literature, and this by simply applying some of the restrictions we have axiomatised.

- If M satisfies PIAw and NIAw, then $A_M^? = \emptyset$. In other words, IAF_M is an *attack-incomplete* AF, also called *partial* AF, firstly studied by Cayrol et al. (2007).
- If M satisfies AU, then $R_M^? = \emptyset$. In other words, IAF_M is an *argument-incomplete* AF introduced by Coste-Marquis et al. (2007).

As it should be now evident, \mathcal{EA} -models are much more general than incomplete AFs. More concretely, there are three types of information that we can model with \mathcal{EA} -models but fall out of the scope of incomplete AFs: nested beliefs, multi-agent information and non-total uncertainty about the elements of $A^?$ and $R^?$. Moreover, the basic modal language can be used to answer the queries of the main reasoning tasks regarding argument acceptability in incomplete AFs.

8.2 Control AFs

Control argumentation frameworks (Dimopoulos et al. 2018) are a more complex kind of structure for representing qualitative uncertainty. They enrich incomplete AFs in two different senses. First, they augment incomplete AFs with an additional uncertain attack relation \rightleftharpoons , whose precise meaning will be clarified later on. Second, they include a dynamic component by considering yet another partition of the underlying AF (the *control part*), which is intuitively assumed to be modifiable by the agent. In this subsection, we provide a natural epistemic multi-agent interpretation of control argumentation frameworks (CAFs) using our logic. The intuitive picture behind this interpretation is that of an agent, PRO (the proponent), reasoning about how to convince another agent, OPP (the opponent). Here, the uncertain part of CAFs captures the lack of total knowledge of PRO about OPP's knowledge of the underlying AF. Moreover, the so-called *control part* of a CAF represents the private knowledge of PRO. We also provide a reduction of the main reasoning tasks regarding CAFs to our logic. Let us start with the main definitions concerning CAFs and their semantics (Dimopoulos et al. 2018).

Definition 21 (*Control argumentation framework*) A control argumentation framework is a triple $\text{CAF} = (F, C, U)$ where:

- $F = (A, R)$ is called the *fixed part*, where $R \subseteq (A \times A^?) \cup (A \times A^?)$ and A and $A^?$ are two finite sets of arguments;
- $U = (A^?, (R^? \cup \Leftarrow))$ is called the *uncertain part*, where $R^?, \Leftarrow \subseteq (A \times A^?) \cup (A \times A^?)$ and \Leftarrow is symmetric and irreflexive;
- $C = (A_C, R_C)$ is called the *control part* where A_C is yet another finite set of arguments and $R_C \subseteq (A_C \times (A \cup A^? \cup A_C)) \cup ((A \cup A^? \cup A_C) \times A_C)$;
- $A, A^?$, and A_C are pairwise disjoint; and
- $R, R^?, \Leftarrow$, and R_C are pairwise disjoint.

We sometimes call $A \cup A_C \cup A^?$ the *domain of CAF* and denote it as Δ^{CAF} . Intuitively, the new components can be thought as follows. \Leftarrow is an attack relation s.t. the existence of its elements is known by the agent, but the direction is unknown. So, whenever $(x, y) \in \Leftarrow$, it intuitively means that the agent knows that there is an attack among x and y but it does not know who attacks who. As for $C = (A_C, R_C)$, it is supposed to be the part of the framework that depends on the actions of the agent. These intuitions are formally specified in the following definitions:

Definition 22 (*Completion*) A completion of $\text{CAF} = (F, C, U)$ is any AF (A^*, R^*) s.t.:

- $(A \cup A_C) \subseteq A^* \subseteq (A \cup A_C \cup A^?)$;
- $(R \cup R_C)_{|A^*} \subseteq R^* \subseteq (R \cup R_C \cup R^? \cup \Leftarrow)_{|A^*}$; and
- for every $x, y: (x, y) \in \Leftarrow$ and $x, y \in A^*$ implies $(x, y) \in R^*$ or $(y, x) \in R^*$.

From an epistemic perspective, completions can be understood as possible knowledge bases that PRO attributes to OPP. Note that control arguments A_C are always a subset of every completion. Something similar happens with control attacks (conditionally on the domain A^* of each completion). The intuition here is that (F, C, U) provides the picture of a *finished* debate seen from PRO's point of view, where she has communicated all her available arguments A_C . The spectrum of debate states that are between the initial one (where nothing has been said) and (F, C, U) are captured by the notion of control configuration:

Definition 23 (*Control configuration*) Given $\text{CAF} = (F, C, U)$, a *control configuration* is a subset of control arguments $\text{CFG} \subseteq A_C$ and its *associated CAF* is $\text{CAF}_{\text{CFG}} := (F, C_{\text{CFG}}, U)$ where $C_{\text{CFG}} := (\text{CFG}, R_C \upharpoonright_{A \cup A^? \cup A_{\text{CFG}}})$.

One more time, classical reasoning tasks regarding AFs can be naturally generalised to CAFs. As an example, consider the following one (Dimopoulos et al. 2018):

Pr-Necessary-Sceptical-Controllability (Pr-NSCon)	
Given:	A control argumentation framework CAF = (F, C, U) and an argument a ∈ A.
Question:	Is it true that there is a configuration CFG ⊆ A _C s.t. for every completion F* = (A*, R*) of CAF _{CFG} and for all E ∈ Pr(F*), a ∈ E?

We now show how to build a two-agent \mathcal{EA} -model to reason about a given CAF. First, let CAF = (F, C, U), we define the set of variables of CAF as $\mathcal{V}^{\text{CAF}} := \mathcal{V}_{\{\text{PRO}, \text{OPP}\}}^{\text{AUACUA}^?}$.

Definition 24 (Associated model) Let CAF = (F, C, U), let $\wp(A \cup A_C \cup A^?) = \{E_1, \dots, E_n\}$, we define the \mathcal{EA} -model associated to CAF as $M^{\text{CAF}} := (W^{\text{CAF}}, \mathcal{R}^{\text{CAF}}, V^{\text{CAF}})$ where:

- $W^{\text{CAF}} := \{w^{(A^*, R^*)} \mid (A^*, R^*) \text{ is a completion of } \text{CAF}_{\emptyset}\}$.
- $\mathcal{R}_{\text{PRO}}^{\text{CAF}} := W^{\text{IAF}} \times W^{\text{IAF}}$ and $\mathcal{R}_{\text{OPP}}^{\text{CAF}} := \emptyset$.³⁷
- V^{CAF} is defined for each kind of variables as follows:
 $V^{\text{CAF}}(\text{aw}_{\text{PRO}}(x)) = W^{\text{CAF}}$;
 $V^{\text{CAF}}(\text{aw}_{\text{OPP}}(x)) = \{w^{(A^*, R^*)} \mid x \in A^*\}$;
 $V^{\text{CAF}}(x \rightsquigarrow y) = \{w^{(A^*, R^*)} \mid (x, y) \in R^* \text{ or } (x, y) \in R_C\}$;
 for every k such that $1 \leq k \leq n$: $V^{\text{IAF}}(x \in E_k) = W$ if $x \in E_k$; and
 for every k such that $1 \leq k \leq n$: $V^{\text{IAF}}(x \in E_k) = \emptyset$ if $x \notin E_k$.
 Moreover, for any $\text{CFG} \subseteq A_C$ we define $M^{\text{CFG}} := M^{\text{CAF}} \otimes \text{Pub}^{\text{CFG}}$.

Remark 7 Note that the set of completions of CAF_{\emptyset} is equal to $\{(A_w^*, R_w^*) \mid w \in M^{\text{CAF}}[W]\}$ where $A_w^* := A_{\text{OPP}}^M(w)$ and $R_w^* := R^M(w)_{|A_w^*}$.³⁸ Moreover, for any $\text{CFG} \subseteq A_C$, it can be shown that the set of completions of CAF_{CFG} is equal to $\{(A_w^*, R_w^*) \mid w \in M^{\text{CFG}}[W]\}$.

The following proposition digs into this multi-agent epistemic interpretation of CAFs:

Proposition 4 Let CAF = (F, C, U) be a CAF, let M^{CAF} be its associated model, and let $w \in M^{\text{CAF}}[W]$. We have that:

- $A = \{x \in \Delta^{\text{CAF}} \mid M^{\text{CAF}}, w \models \Box_{\text{PRO}} \text{aw}_{\text{OPP}}(x)\}$, i.e. the set of fixed arguments is the set of arguments that the proponent knows that the opponent is aware of.
- $A^? = \{x \in \Delta^{\text{CAF}} \mid M^{\text{CAF}}, w \models \Diamond_{\text{PRO}} \text{aw}_{\text{OPP}}(x) \wedge \Diamond_{\text{PRO}} \neg \text{aw}_{\text{OPP}}(x)\}$, i.e. uncertain arguments are those that PRO considers possible both that OPP is aware of them and that OPP is not.

³⁷ Actually, the accessibility relation for OPP is irrelevant.

³⁸ Notice that the definitions of $A_i(\cdot)$ and $R(\cdot)$ (Definition 8) did not include the parameter M , since we were reasoning about different states in the same model. Nevertheless, for this subsection, it will be useful to recover the parameter, since we need to reason about awareness sets and attacks that holds at states throughout different models.

- $A_C = \{x \in \Delta^{CAF} \mid M^{CAF}, w \models \Box_{PRO} \neg \text{aw}_{OPP}(x)\}$, i.e. control arguments are the arguments that PRO knows that OPP is not aware of.
- $R = \{(x, y) \in \Delta^{CAF} \times \Delta^{CAF} \mid M^{CAF}, w \models \Box_{PRO}((\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y)) \rightarrow x \rightsquigarrow y) \wedge \Diamond_{PRO}(\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y)))\}$, i.e. fixed attacks are those that the proponent knows that the opponent is aware of (conditionally on the awareness of the involved arguments). Moreover, the second condition serves to distinguish R from R_C .
- $\Leftarrow = \{(x, y) \in \Delta^{CAF} \times \Delta^{CAF} \mid M^{CAF}, w \models \varphi_1 \wedge \varphi_2 \wedge \varphi_3\}$ where:
 - $\varphi_1 = \Box_{PRO}((\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y)) \rightarrow (x \rightsquigarrow y \vee y \rightsquigarrow x))$,
 - $\varphi_2 = \Diamond_{PRO}(\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y) \wedge \neg x \rightsquigarrow y)$, and
 - $\varphi_3 = \Diamond_{PRO}(\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y) \wedge \neg y \rightsquigarrow x)$. So if $(x, y) \in \Leftarrow$, then PRO knows (conditionally on OPP’s awareness of x and y) that either x attacks y or vice-versa. Moreover, the meaning of \Leftarrow (provided by Definition 22) forces PRO to consider as epistemically possible situations where OPP is aware of both arguments but where (x, y) (resp. (y, x)) does not hold.
- $R^? = \{(x, y) \in \Delta^{CAF} \times \Delta^{CAF} \mid M^{CAF}, w \models \varphi_1 \wedge \varphi_2 \wedge (\varphi_3 \vee \varphi_4)\}$ where:
 - $\varphi_1 = \Diamond_{PRO}(\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y) \wedge x \rightsquigarrow y)$,
 - $\varphi_2 = \Diamond_{PRO}(\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y) \wedge \neg x \rightsquigarrow y)$,
 - $\varphi_3 = \Diamond_{PRO}(\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y) \wedge \neg x \rightsquigarrow y \wedge \neg y \rightsquigarrow x)$, and
 - $\varphi_4 = \Box_{PRO}((\text{aw}_{OPP}(x) \wedge \text{aw}_{OPP}(y)) \rightarrow y \rightsquigarrow x)$. In words, uncertain attacks are those that (i) PRO considers possible that OPP is aware of them and that OPP is unaware (first two conjuncts), and (ii) they are not members of \Leftarrow (third conjunct).
- $R_C = \{(x, y) \in \Delta^{CAF} \times \Delta^{CAF} \mid M^{CAF}, w \models \Box_{PRO}(x \rightsquigarrow y) \wedge (\Box_{PRO} \neg \text{aw}_{OPP}(x) \vee \Box_{PRO} \neg \text{aw}_{OPP}(y))\}$, i.e. control attacks are those such that: (i) they are private for PRO (meaning that it knows that OPP is a unaware of some of the involved arguments), and (ii) PRO is sure that they hold.

Finally, we can reduce controllability of a given CAF to a model-checking problem in the associated \mathcal{EA} -model. For doing so, we will use the following shorthand, informally expressing that E_k is part of PRO’s private knowledge (i.e. that E_k is a set of control arguments in the associated \mathcal{EA} -model).

$$\text{private}_{PRO}(E_k) := \bigwedge_{x \in \Delta^{CAF}} (x \in E_k \rightarrow \Box_{PRO} \neg \text{aw}_{OPP}(x)).$$

Proposition 5 *Let $CAF = (F, C, U)$ be a CAF, let $\wp(A \cup A^? \cup A_C) = \{E_1, \dots, E_n\}$, let $M^{CAF} = (W^{CAF}, \mathcal{R}^{CAF}, V^{CAF})$ be its associated model, and let $w \in W^{CAF}$. We have that:*

- *The answer to Pr-NSCon with input CAF and $a \in A$ is yes iff*

$$M^{CAF}, w \models \bigvee_{1 \leq l \leq n} (\text{private}_{PRO}(E_l) \wedge [\text{Pub}^{E_l}, \Delta] \Box_{PRO} \text{stracc}_{OPP}(a)).$$

Proof See “Appendix A5”. □

Again, the proposition can be easily adapted to other controllability problems. Besides, the fact that the control part of a CAF is representable through public additions reveals that the very notion of control configuration assumes that the speaker (proponent) is sure about the effects of the communication. More refined forms of communication (like the ones we have studied in Sect. 7) seem to deserve future attention so as to develop variants of CAFs.

8.3 Reasoning about opponent models

Strategic argumentation (Thimm 2014) studies how agents should interact in adversarial dialogues in order to maximize their expected utility. A useful tool in this context is *opponent modelling* (Oren and Norman 2009; Rienstra et al. 2013), a well-known technique among AI researchers that deals with more general adversarial situations (Carmel and Markovitch 1996a, b). Opponent modelling for abstract argumentation assumes, as we do for MAFs, that there is an underlying UAF (A, R) which contains all arguments relevant to a particular discourse (Oren and Norman 2009; Rienstra et al. 2013; Thimm 2014; Black et al. 2017). Based on this, it provides a model of a proponent in a strategic dialogue. The central notion is that of a *belief state* of the proponent, which is defined in general as a couple (B, E) where $B \subseteq A$ is the set of arguments the proponent is aware of and $E \subseteq \wp(A)$ is the set of belief states the agent considers possible for its opponent.³⁹ The belief state can be more or less refined: at level 0 of refinement it only includes the arguments the proponent is aware of. At level 1 it also contains her beliefs about her opponent's awareness, at level 2 it includes her beliefs about her opponent's beliefs about her own awareness and so on, up to an arbitrary level n of nesting.

In our semantics, any pointed \mathcal{EA} -model (M, w) for two agents contains all information available to define a belief state of any level n of refinement for agent i . To make this more precise we introduce some notation.

$$\text{View}_i(w) := \bigcap \{A_i(w') \mid w' \in \mathcal{R}_i[\{w\}]\}.$$

where $\mathcal{R}_i[W'] := \{u \in W \mid \exists u' \in W' (u' \mathcal{R}_i u)\}$ for any $W' \subseteq W[M]$. Intuitively, $\text{View}_i(w)$ consists of the arguments that agent i believes (knows) to be aware of at state w . Based on this, we can define the *belief states of agent i at state w for an arbitrary level n* as follows

$$\begin{aligned} \text{BS}_i^0(w) &:= (\text{View}_i(w), \emptyset). \\ \text{BS}_i^{n+1}(w) &:= (\text{View}_i(w), \{\text{BS}_j^n(z) \mid z \in \mathcal{R}_i[\{w\}], j \neq i\}). \end{aligned}$$

It is interesting to show that the actual definitions of a belief state provided by Oren and Norman (2009) and Rienstra et al. (2013) are a particular case of our definition modulo the restriction to specific classes of pointed \mathcal{EA} -models.⁴⁰

³⁹ In the actual formulation of Oren and Norman (2009) and the later extension of Rienstra et al. (2013) a belief state contains other components, e.g. a utility function u which provides a numerical utility to possible dialogue traces. We leave these elements aside in our present discussion.

⁴⁰ To be more precise, the definition provided by Rienstra et al. (2013) forces non well-founded belief states containing either cycles or infinite chains. This definition can be captured by simply setting $\text{BS}_i(w) := (\text{View}_i(w), \{\text{BS}_j(z) \mid z \in \mathcal{R}_i[\{w\}], j \neq i\})$ and taking a serial \mathcal{EA} -model as the input. However, we leave this aside for presentational purposes.

In the case of the *simple agent* models (Oren and Norman 2009, Definition 5; Rienstra et al. 2013, Definition 8), a belief state of level n has the form $(B^0, (B^1, \dots (B^n, \emptyset) \dots))$, where each B^i is an awareness set (of the proponent if i is even and of the opponent if i is odd), and where $B^{i+1} \subseteq B^i$. Here, B_0 contains the awareness set of the proponent, B_1 the awareness set the proponent attributes to the opponent, B_2 the awareness set the proponent thinks the opponent attributes to him, and so forth. From our model-theoretic perspective, this tacitly assumes that we are in a \mathcal{AoA} -model where each \mathcal{R}_i is functional. Indeed, functionality forces each $\mathcal{R}_i[\{w\}]$ to consist of a singleton set. This implies that each $\text{BS}_i^n(w)$ has a singleton set as its second element E . Moreover, combined positive and negative introspection guarantee that $\text{View}_i(w) = \bigcap \{A_i(w') \mid w' \in \mathcal{R}_i[\{w\}]\} = A_i(w)$, as presupposed in simple agent models. Furthermore, GNIAw forces $B^{i+1} \subseteq B^i$ as desired.

In the more general case of *uncertain agent models* (Rienstra et al. 2013, Definition 10), a belief state (B, E) instead consists of an awareness set B for agent i and a set of belief states E of the opponent, each one of the form (B', E') such that $B' \subseteq B$. Again, the latter condition assumes that GNIAw holds. The fact that B is the awareness set of the actual state tacitly assumes PIAw as before, but functionality does not need to hold any more, and therefore we are in the more general class of (serial) \mathcal{AoA} -models.

Yet a more general class of models, *extended agent models* is defined by Rienstra et al. (2013, Definition 11). Here *virtual arguments* are added as arguments the agent is not aware of but consider possible other agents are. From our point of view this corresponds to the failure of GNIAw (while PIAw and NIAw still hold).

Applied to this approach to strategic argumentation, our logics and semantics provide a systematic way to reason about the effects of different kinds of argumentative events on the belief states of agents. This can be useful, in turn, to compute the best move for an agent at a given moment of a dialogue. Furthermore, an important part of the work in strategic argumentation using opponent models consists in finding *appropriate* ways to update belief states. More formally, given a class of belief states \mathcal{B} and a universal set of arguments C , the challenge consists in finding functions of the form $\text{upd}: \mathcal{B} \times \wp(C) \rightarrow \mathcal{B}$. From this perspective, our Lemma 1 provides sufficient conditions for accomplishing this task given different constraints on \mathcal{B} .

9 Discussion, open problems and future research

As mentioned in Sect. 3, there are many alternative design choices for multi-agent argumentation frameworks, which are worth discussing. A first choice concerns the finiteness of the argumentative pool, i.e. (a) of p. 8. Indeed, the set A of potentially available arguments may well be infinite. In principle, this option is viable for a propositional language with a countable set of variables. However, a propositional language allows to encode the standard solution concepts only in the finite case.⁴¹ As many other works in this field, we restrict ourselves to finite AFs, which is enough for modelling most real-life debates.

⁴¹ Previous works by Grossi (2010a, b) shows that, in general, some solution concepts require expressive power beyond modal μ -calculus and are likely to be full second-order.

A second branching option for design concerns (b), the fact that A is fixed in advance. One can instead assume that it is evolving through updates, as in Doutre and Mailly (2018, Sect. 1.3). Our option has been shared by Sakama (2012), Doutre et al. (2014), de Saint-Cyr et al. (2016) and Caminada and Sakama (2017), among others. The rationale behind it is that this imposes no limitation for modelling acquisition of new arguments by an agent and other relevant dynamics of information change, at least when the propositional language is rich enough to encode subjective *awareness* of arguments (Sect. 4).

Another option is not to assume (c), the existence of an *objective* attack relation R between members of A . Proposals like Dyrkolbotn and Pedersen (2016), Baumeister et al. (2018b) avoid (c). This goes in hand with the very minimal assumption that agents only share a “pool” of arguments A , but no constraint on how these arguments interact with each other. This amounts to eliminating the R component of our structures, and may be adequate in contexts where conflicts between arguments cannot be assessed even from a third person perspective. We should however stress that this is just a special case of a MAF, one where $R = \emptyset$. In line with others—for instance Schwarzenruber et al. (2012)—we decided to build assumption (c) into our design, since our Kripke semantics still allows, in the general case, to model radical disagreement about attacks at the epistemic level. Besides, this assumption is acceptable in many applications and provides a straightforward way to define more complex notions we are after. We note however, that it is possible to perform the same constructions without assuming (c) by a slightly different language and semantics.

Regarding the nature of the subjective awareness of arguments (A_i) and attacks (R_i), there are multiple choices to be made, which consist in accepting or rejecting the following constraints:

- (d) $A_i \subseteq A$ (agents are only aware of “real” arguments).
- (e) $R_i \subseteq A_i \times A_i$ (agents are only aware of attacks among arguments they are aware of).
- (f) $R_i \subseteq R$ (sound awareness of attacks).
- (g) $R \cap (A_i \times A_i) \subseteq R_i$ (complete awareness of attacks).
- (h) $A \subseteq A_i$ (agents are aware of all “real” arguments).

Recall that our choice for design (Definition 2) integrates (d), (e), (f), and (g), but all of them are open to discussion. Although strongly intuitive, (d) and (e) are questioned by Schwarzenruber et al. (2012), which defines a logic for reasoning about “non-existent” or “virtual” arguments $\{?_0, ?_1, \dots\}$. We do not integrate constraint (h) as it discards the natural intuition that different agents are aware of different sets of arguments.⁴² Under this assumption the agents’ view can only differ with respect to the attack relations, as in Dyrkolbotn and Pedersen (2016), Cayrol et al. (2007). Again, this condition isolates a specific subclass of our MAFs, those for which, $A_i = A$, which can be captured axiomatically by imposing all awareness atoms as axioms. Assuming both (f) and (g), i.e. SCAA, is common in the literature on multi-agent abstract argumentation (Caminada 2006; Sakama 2012; Schwarzenruber et al. 2012; Doutre et al. 2017; Rahwan and Larson 2009). However, SCAA may seem too idealized

⁴² For example when discussing about climate change, a climatologist is likely to be aware of many more arguments pro or contra than the layman.

in many contexts, since it brings the notion of awareness of arguments closer to the one of *knowledge of arguments*.⁴³ Agents may indeed have different abilities to spot conflicts between arguments⁴⁴ or, even more, they may be entitled to radically different views about the nature of the attacks.⁴⁵ Here again, the just mentioned differences in awareness of attacks can still be modelled in our Kripke semantics. Indeed, what matters here is the distinction between simple SCAA and common knowledge (belief) of SCAA. The latter is a much stronger assumption and their difference becomes transparent in the language and semantics of epistemic logic (Sect. 5).

The aim of this paper has been to introduce a new DEL framework for reasoning about multi-agent abstract argumentation. This involves the setup of a three-layer logic: propositional, epistemic and dynamic. Our first goal was to encode the key argumentation-theoretic notions in the language of propositional logic, and we showed that this is possible in the finite case. Concerning the epistemic layer, we provided complete axiomatizations for a number of intuitive constraints on awareness of arguments and attacks. Moreover, specific constraints isolate different classes of structures already used in abstract argumentation to model qualitative uncertainty about AFs, and our logic is comprehensive enough to reason about them (Sect. 8.1). As for the third layer, its language and semantics allow modelling subtle forms of information change (Sect. 7) and reasoning about other formalisms for uncertainty and dynamics (Sects. 8.2, 8.3).

Although event models for DEL are apt to describe the effects of complex information updates, its language describes only indirectly the agential component of a debate. More in detail, the language allows to reason about what happens after some combination of communicative act and information update has been performed, but it does not allow to reason about what agents “see to it that” in a debate. This is likely to require additional tools from logics of agency and epistemic planning, which suggest promising venues for future work.

Author Contributions Both authors contributed equally to this paper.

Funding This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 748421. Carlo Proietti would also like to thank the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond) for funding received during the preparatory phase of this research (P16-0596:1). The research activity of A. Yuste-Ginel is supported by the Spanish Ministry of Universities through the predoctoral Grant MECD-FPU 2016/04113.

Declaration

⁴³ In this sense, knowing an argument would mean being aware of all an only the right conflict relations in which it is involved.

⁴⁴ I may be aware of some argument against anthropogenic forcing on climate change, and a climatologist may present me the result of some scientific study that undermines such argument. Yet I may not be able to see the attack from the second argument to the first.

⁴⁵ Given two potentially conflicting arguments a and b , it may be the case that two individuals disagree on their relative weights and therefore on whether a attacks b or vice versa. See Dyrkolbotn and Pedersen (2016) for examples and discussion. A milder assumption is that awareness of attacks is shared, i.e. uniform among individuals. Under this assumption the possibility is still open that all agents may be wrong with respect to the objective attacks.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

A1. Encodings

Proposition 1

Proof All results are obtained via a chain of equivalences from the left-hand side of each item to its right-hand side.

[1]

$$\begin{aligned}
 & \text{Suppose that } v_{\text{MAF}} \models E_k \sqsubseteq E_l. \\
 & \iff \\
 & v_{\text{MAF}} \models \bigwedge_{a \in A} (a \in E_k \rightarrow a \in E_l) \text{ (definition of } \sqsubseteq \text{)}. \\
 & \iff \\
 & \text{For all } a \in A: v_{\text{MAF}} \models a \in E_k \text{ implies } v_{\text{MAF}} \models a \in E_l \text{ (semantics of } \wedge \text{ and } \rightarrow \text{)}. \\
 & \iff \\
 & \text{For all } a \in A: a \in E_k \text{ implies } a \in E_l \text{ (definition of } v_{\text{MAF}} \text{)}. \\
 & \iff \\
 & E_k \subseteq E_l \text{ (definition of } \sqsubseteq \text{)}.
 \end{aligned}$$

For the other claim

$$\begin{aligned}
 & \text{Suppose that } v_{\text{MAF}} \models E_k \sqsubset E_l. \\
 & \iff \\
 & v_{\text{MAF}} \models E_k \sqsubseteq E_l \wedge \bigvee_{a \in A} (a \in E_l \wedge \neg a \in E_k) \text{ (definition of } \sqsubset \text{)}. \\
 & \iff \\
 & E_k \subseteq E_l \text{ and for some } a \in A: v_{\text{MAF}} \models a \in E_l \text{ and } v_{\text{MAF}} \models \neg a \in E_k \text{ (by the semantics of} \\
 & \quad \text{Boolean connectives and previous claim)}. \\
 & \iff \\
 & E_k \subseteq E_l \text{ and for some } a \in A: a \in E_l \text{ and } a \notin E_k \text{ (definition of } v_{\text{MAF}} \text{)}. \\
 & \iff \\
 & E_k \subset E_l \text{ (definition of } \sqsubset \text{)}.
 \end{aligned}$$

For the equivalent formulation, note that Th_{MAF} is uniquely satisfied by v_{MAF} .

[2]

$$\begin{aligned}
 & \text{Suppose that } v_{\text{MAF}} \models \text{conf_free}_i(E_k). \\
 & \iff
 \end{aligned}$$

$v_{MAF} \models \bigwedge_{a \in A} \left(a \in E_k \rightarrow \left(\text{aw}_i(a) \wedge \neg \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a) \right) \right)$ (definition of $\text{conf_free}_i(E_k)$).

\iff

For every $a \in A$: $v_{MAF} \models a \in E_k$ implies $\left(v_{MAF} \models \text{aw}_i(a) \text{ and } v_{MAF} \models \neg \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a) \right)$ (semantics of \wedge and \rightarrow).

\iff

For every $a \in A$: $a \in E_k$ implies $\left(a \in A_i \text{ and } v_{MAF} \models \neg \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a) \right)$ (definition of v_{MAF}).

\iff

For every $a \in A$: $a \in E_k$ implies $\left(a \in A_i \text{ and } v_{MAF} \models \bigwedge_{b \in A} \neg (b \in E_k \wedge b \rightsquigarrow a) \right)$ (De Morgan).

\iff

For every $a \in A$: $a \in E_k$ implies $\left(a \in A_i \text{ and for all } b \in A$: either $v_{MAF} \not\models b \in E_k$ or $v_{MAF} \not\models b \rightsquigarrow a$ (semantics of \wedge and \neg).

\iff

For every $a \in A$: $a \in E_k$ implies $\left(a \in A_i \text{ and for all } b \in A$: either $b \notin E_k$ or it is not the case that bRa (definition of v_{MAF}).

\iff

$E_k \subseteq A_i$ and E_k is conflict-free (definition of \subseteq and conflict-free).

[3]

Suppose that $v_{MAF} \models \text{complete}_i(E_k)$.

\iff

$v_{MAF} \models \text{conf_free}_i(E_k) \wedge \bigwedge_{a \in A} \left(a \in E_k \leftrightarrow \bigwedge_{b \in A} \left((\text{aw}_i(b) \wedge b \rightsquigarrow a) \rightarrow \bigvee_{c \in A} (c \in E_k \wedge c \rightsquigarrow b) \right) \right)$ (definition of complete_i).

\iff

$v_{MAF} \models \text{conf_free}_i(E_k)$ and

$v_{MAF} \models \bigwedge_{a \in A} \left(a \in E_k \leftrightarrow \bigwedge_{b \in A} \left((\text{aw}_i(b) \wedge b \rightsquigarrow a) \rightarrow \bigvee_{c \in A} (c \in E_k \wedge c \rightsquigarrow b) \right) \right)$ (semantics of \wedge).

\iff

$(E_k \subseteq A_i \text{ and } E_k \text{ is conflict-free})$ and

$v_{MAF} \models \bigwedge_{a \in A} \left(a \in E_k \leftrightarrow \bigwedge_{b \in A} \left((\text{aw}_i(b) \wedge b \rightsquigarrow a) \rightarrow \bigvee_{c \in A} (c \in E_k \wedge c \rightsquigarrow b) \right) \right)$ (item 2).

\iff

$(E_k \subseteq A_i \text{ and } E_k \text{ is conflict-free})$ and for all $a \in A$: $(v_{MAF} \models a \in E_k$ iff for all $b \in A$: $(v_{MAF} \models \text{aw}_i(b) \text{ and } v_{MAF} \models b \rightsquigarrow a)$ implies that for some $c \in A$: $(v_{MAF} \models c \in E_k \text{ and } v_{MAF} \models c \rightsquigarrow b))$ (semantics of Boolean connectives).

\iff

$(E_k \subseteq A_i \text{ and } E_k \text{ is conflict-free})$ and for all $a \in A$: $(a \in E_k$ iff for all $b \in A$: $(b \in A_i \text{ and } bRa)$ implies that for some $c \in A$: $(c \in E_k \text{ and } cRb))$ (definition of v_{MAF}).

\iff

E_k is complete w.r.t. (A_i, R_i) (Definition 3, since $c \in E_k \subseteq A_i$).

[4]

Suppose that $v_{\text{MAF}} \models \text{preferred}_i(E_k)$.

\iff

$v_{\text{MAF}} \models \text{complete}_i(E_k) \wedge \neg \bigvee_{1 \leq l \leq n} (\text{complete}_i(E_l) \wedge (E_k \sqsubset E_l))$ (definition of preferred_i).

\iff

$v_{\text{MAF}} \models \text{complete}_i(E_k)$ and $v_{\text{MAF}} \models \neg \bigvee_{1 \leq l \leq n} (\text{complete}_i(E_l) \wedge (E_k \sqsubset E_l))$ (semantics of \wedge).

\iff

(P) $\overbrace{E_k \text{ is complete w.r.t. } (A_i, R_i) \text{ and } v_{\text{MAF}} \models \neg \bigvee_{1 \leq l \leq n} (\text{complete}_i(E_l) \wedge (E_k \sqsubset E_l))}$ (item 3).

\iff

(P) and $v_{\text{MAF}} \models \bigwedge_{1 \leq l \leq n} \neg(\text{complete}_i(E_l) \wedge (E_k \sqsubset E_l))$ (DeMorgan).

\iff

(P) and for every $1 \leq l \leq n$: $v_{\text{MAF}} \models \text{complete}_i(E_l)$ implies $v_{\text{MAF}} \not\models (E_k \sqsubset E_l)$ (semantics of \wedge, \rightarrow and \neg and propositional reasoning).

\iff

(P) and for every $1 \leq l \leq n$: E_l is complete w.r.t. (A_i, R_i) implies $v_{\text{MAF}} \not\models (E_k \sqsubset E_l)$ (item 3).

\iff

E_k is complete w.r.t. (A_i, R_i) and for every $1 \leq l \leq n$: E_l is complete w.r.t. (A_i, R_i) implies that it is not the case that $E_k \sqsubset E_l$ (item 1, second claim).

\iff

E_k is preferred w.r.t. (A_i, R_i) (Definition 3).

[5]
[strong acceptance]

Suppose that $v_{\text{MAF}} \models \text{stracc}_i(a)$.

\iff

$v_{\text{MAF}} \models \bigwedge_{1 \leq k \leq n} (\text{preferred}_i(E_k) \rightarrow a \in E_k)$ (definition of stracc_i).

\iff

For every $1 \leq k \leq n$: $v_{\text{MAF}} \models \text{preferred}_i(E_k) \rightarrow a \in E_k$ (semantics of \wedge).

\iff

For every $1 \leq k \leq n$: $v_{\text{MAF}} \models \text{preferred}_i(E_k)$ implies $v_{\text{MAF}} \models a \in E_k$ (semantics of \rightarrow).

\iff

For every $1 \leq k \leq n$: E_k is preferred w.r.t. (A_i, R_i) implies $v_{\text{MAF}} \models a \in E_k$ (item 4).

\iff

For every $1 \leq k \leq n$: E_k is preferred w.r.t. (A_i, R_i) implies $a \in E_k$ (definition of v_{MAF}).

\iff

a is strongly accepted for i (Definition 4).

[weak acceptance]

$$\begin{aligned}
 & \text{Suppose that } v_{\text{MAF}} \models \text{wekacc}_i(a). \\
 & \iff \\
 v_{\text{MAF}} \models & \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge a \in E_k) \wedge \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge \neg a \in E_k) \wedge \\
 & \neg \bigvee_{1 \leq k \leq n} (\text{preferred}_i(E_k) \wedge \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a)) \text{ (definition of wekacc}_i). \\
 & \iff \\
 & (\text{for some } 1 \leq k \leq n: v_{\text{MAF}} \models \text{preferred}_i(E_k) \text{ and } v_{\text{MAF}} \models a \in E_k) \text{ and } (\text{for some } \\
 & \quad 1 \leq k \leq n: v_{\text{MAF}} \models \text{preferred}_i(E_k) \text{ and } v_{\text{MAF}} \models \neg a \in E_k) \text{ and} \\
 v_{\text{MAF}} \models & \bigwedge_{1 \leq k \leq n} \neg (\text{preferred}_i(E_k) \wedge \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a)) \text{ (semantics of Boolean} \\
 & \quad \text{connectives and DeMorgan).} \\
 & \iff \\
 & \text{(P}_1\text{)} \\
 & \quad \overbrace{(\text{for some } 1 \leq k \leq n: E_k \in \text{Pr}(A_i, R_i) \text{ and } a \in E_k) \text{ and}} \\
 & \quad \text{(P}_2\text{)} \\
 & \quad \overbrace{(\text{for some } 1 \leq k \leq n: E_k \in \text{Pr}(A_i, R_i) \text{ and } a \notin E_k) \text{ and}} \\
 v_{\text{MAF}} \models & \bigwedge_{1 \leq k \leq n} \neg (\text{preferred}_i(E_k) \wedge \bigvee_{b \in A} (b \in E_k \wedge b \rightsquigarrow a)) \text{ (item 4, definition of} \\
 & \quad \text{Pr}(\cdot) \text{ and definition of } v_{\text{MAF}}). \\
 & \iff \\
 & \text{(P}_1\text{) and (P}_2\text{) and (for every } 1 \leq k \leq n: v_{\text{MAF}} \models \text{preferred}_i(E_k) \text{ implies there is no} \\
 & \quad b \in A \text{ s.t. } v_{\text{MAF}} \models b \in E_k \text{ and } v_{\text{MAF}} \models b \rightsquigarrow a) \text{ (semantics of Boolean connectives and} \\
 & \quad \text{propositional reasoning).} \\
 & \iff \\
 & \text{(P}_1\text{) and (P}_2\text{) and (for every } 1 \leq k \leq n: v_{\text{MAF}} \models \text{preferred}_i(E_k) \text{ implies there is no} \\
 & \quad b \in A \text{ s.t. } b \in E_k \text{ and } bRa) \text{ (definition of } v_{\text{MAF}}). \\
 & \iff \\
 & \text{(P}_1\text{) and (P}_2\text{) and (for every } 1 \leq k \leq n: E_k \text{ is preferred w.r.t. } (A_i, R_i) \text{ implies there} \\
 & \quad \text{is no } b \in A \text{ s.t. } b \in E_k \text{ and } bRa) \text{ (item 4).} \\
 & \iff \\
 & \text{(P}_1\text{) and (P}_2\text{) and } (\forall E_k \in \text{Pr}(A_i, R_i) a \notin E_k^+) \text{ (definition of Pr}(\cdot) \text{ and } (\cdot)^+). \\
 & \iff \\
 & a \text{ is weakly accepted by } i \text{ (Definition 4).}
 \end{aligned}$$

□

Proofs for other cases (weak rejection, strong rejection and borderline) run very similar and are left to the reader.

A2. Proof of Theorem 1

To prove completeness, we need a few preliminaries. First, we need to define the class of structures that will fit the canonical characterization of EA (and its extensions).

Definition 25 (*Quasi-model*) A *quasi- \mathcal{EA} -model* (or *quasi-model*, for short) for $\mathcal{V}_{\text{Ag}}^A$ is a tuple (W, \mathcal{R}, V) where $W \neq \emptyset$, $\mathcal{R}: \text{Ag} \rightarrow \wp(W \times W)$ and $V: \mathcal{V}_{\text{Ag}}^A \rightarrow \wp(W)$ s.t. for every $a \in A$, $E_k \in \wp(A)$, $w \in W$:

1. $w \in V(a \in E_k)$ implies that for every $v \in W$: $(w, v) \in \bigcup_{i \in \text{Ag}} \mathcal{R}_i$ implies $v \in V(a \in E_k)$ (positive introspection of subsets).
2. $w \notin V(a \in E_k)$ implies that for every $v \in W$: $(w, v) \in \bigcup_{i \in \text{Ag}} \mathcal{R}_i$ implies $v \notin V(a \in E_k)$ (negative introspection of subsets).
3. For every $w \in W$, $\hat{V}(w)$ represents an enumeration of $\wp(A)$.

For the sake of readability, we shorten $\bigcup_{i \in \text{Ag}} \mathcal{R}_i$ as \mathcal{R}_{Ag} . A quasi-*AoA*-model is a quasi-model satisfying *positive* and *negative introspection of attacks*,⁴⁶ *PIAw* and *GNIAw*. In general, a quasi- \mathcal{C}^L -model is one that satisfies the constraints of \mathcal{C}^L . For instance, a quasi-*S4(AoA)*-model is a quasi-*AoA*-model where every \mathcal{R}_i is a preorder. The notion of truth in pointed quasi-models is the same as in pointed models (Definition 9). Given a pointed quasi-model $(M, w) = ((W, \mathcal{R}, V), w)$, the *sub(quasi-)model of M generated by w using \mathcal{R}_{Ag}* is the tuple $M' = (W', \mathcal{R}', V')$ s.t.: W' is the smallest set containing w and closed under \mathcal{R}_{Ag} ; $\mathcal{R}'_i = \mathcal{R}_i \cap (W' \times W')$ for each $i \in \text{Ag}$; and $V'(p) = V(p) \cap W'$ for each $p \in \mathcal{V}$. Given a pointed quasi-model (M, w) , we denote by M^w the sub(quasi-)model of M generated by w using \mathcal{R}_{Ag} .

Lemma 3 *Let $(M, w) = ((W, \mathcal{R}, V), w)$ be a pointed quasi- \mathcal{EA} -model and let $M' = (W', \mathcal{R}', V')$ be the submodel generated by w using \mathcal{R}_{Ag} , then:*

1. W' is also closed under \mathcal{R}_i for each $i \in \text{Ag}$.
2. If \mathcal{R}_i is a preorder (resp. an equivalence relation; a serial, transitive and euclidean relation), then so it is \mathcal{R}'_i .
3. M' is an \mathcal{EA} -model.
4. If M is a quasi-*AoA*-model, then M' is an *AoA*-model.
5. For every $\varphi \in \mathcal{L}(\mathcal{V}_{\text{Ag}}^A, \square)$ and every $v \in W'$: $M, v \models \varphi$ iff $M', v' \models \varphi$.

Proof Points 1 and 2 are direct consequences of $\mathcal{R}_{\text{Ag}} = \bigcup_{i \in \text{Ag}} \mathcal{R}_i$.

For point 3, we have to prove that *SU* and *ER* are satisfied by M' . For *SU*, take an arbitrary $a \in E_k \in \mathcal{B}$, and let us reason by cases. Case (a): Suppose $w \in V'(a \in E_k)$ (recall w is the generating point of M'), we have that (*) $w \in V(a \in E_k)$, by definition of V' . Let $v \in W'$. By definition of generated submodel, there must be a chain w_0, \dots, w_n s.t $w_0 = w, w_n = v$ and $w_0 \mathcal{R}_{\text{Ag}} w_1, \dots, w_{n-1} \mathcal{R}_{\text{Ag}} w_n$. From the last assertion, (*) and condition 1 of the definition of quasi-model, it is easy to deduce that $w_i \in V(a \in E_k)$ for each $1 \leq i \leq n$, therefore in particular $v \in V(a \in E_k)$. Since $v \in W'$, we have that $v \in V'(a \in E_k)$. We can deduce then, without loss of generality, that $u \in W'$ implies $u \in V'(a \in E_k)$, which implies that $V'(a \in E_k) = W'$ (as $\wp(W')$ is the range of V'). As for the case (b) $w \notin V(a \in E_k)$. Let $v \in W'$, the argument goes as before, but using condition 2 of the definition of quasi-model. Hence, we get that each element w_i of the chain satisfies $w_i \notin V'(a \in E_k)$ and, in particular, $v \notin V'(a \in E_k)$. Generalizing over v , we get $u \in W'$ implies $u \notin V'(a \in E_k)$, which implies $V'(a \in E_k) = \emptyset$ (as $\wp(W')$ is the range of V'). Finally, note that M' satisfies *ER* as a direct consequence of the definition of generated submodel and condition 3 of the definition of quasi-model.

Point 4 reduces to show that M' satisfies *AU*, *PIAw* and *GNIAw*. Note that *PIAw* and *GNIAw* are immediate, since taking generated submodels does not alter awareness

⁴⁶ These are conditions 1 and 2 of Definition 25 reformulated for attack variables.

sets. As for AU, its proof runs completely analogous to point 3, with attacks variables instead of subset variables.

Point 5 is proven by induction on φ . The steps for propositional variables and Boolean connectives are straightforward, hence we just show the one for modal formulas. Assume, as induction hypothesis, that for every $v \in W'$: $M, v \vDash \varphi$ iff $M', v' \vDash \varphi$.

\square_i Suppose $M, v \vDash \square_i \varphi$. This is true iff for every $u \in W$: $v\mathcal{R}_i u$ implies $M, u \vDash \varphi$ (semantics of \square_i). Note that, by Lemma 3.1 for every $u \in W$ s.t. $v\mathcal{R}_i u$, we have that $u \in W'$. Using the induction hypothesis and the last observation we have that for every $u \in W$ ($v\mathcal{R}_i u$ implies $M, u \vDash \varphi$) iff for every $u \in W'$ ($v\mathcal{R}'_i u$ implies $M', u \vDash \varphi$). The last assertion is equivalent to $M', v \vDash \square_i \varphi$ (by the semantics of \square_i). \square

Let L be any of the proof systems under consideration. We say that φ is an L-theorem, in symbols $\vdash_L \varphi$, iff there is a sequence $\varphi_1, \dots, \varphi_n$ s.t. each φ_i is either an instance of an L-axiom scheme or it has been obtained from precedent formulas of the sequence by the application of an L-inference rule. We say that φ is L-derivable from a set of formulas Γ ($\Gamma \vdash_L \varphi$) iff there are $\psi_1, \dots, \psi_n \in \Gamma$ s.t. $\vdash_L (\psi_1 \wedge \dots \wedge \psi_n) \rightarrow \varphi$. A set Γ is L-consistent iff $\Gamma \not\vdash_L \perp$. A set Γ is a L-maximally consistent set (abbreviated L-MCs) iff (i) it is consistent and (ii) there is no consistent Γ' s.t. $\Gamma \subset \Gamma'$. Let us denote by \mathfrak{MC}^L the set of all L-MCs. Proofs for the two following lemmas are standard (see e.g. Blackburn et al. 2002, Section 4.1):

Lemma 4 (Lindenbaum) *Let $\Gamma \subseteq \mathcal{L}(\mathcal{V}_{Ag}^A, \square)$. Suppose Γ is L-consistent, then $\Gamma \subseteq \Gamma^*$ for some $\Gamma^* \in \mathfrak{MC}^L$.*

Lemma 5 (Properties of MCs) *Let $\Gamma \in \mathfrak{MC}$ and let $\varphi \in \mathcal{L}(\mathcal{V}_{Ag}^A, \square)$ then: (i) if $\Gamma \vdash_L \varphi$, then $\varphi \in \Gamma$, (ii) $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$.*

The canonical model for L is defined as $M^L := (W^L, \mathcal{R}^L, V^L)$ where:

- $W^L := \mathfrak{MC}^L$;
- $\Gamma \mathcal{R}_i^L \Delta$ iff $\{\varphi \mid \square_i \varphi \in \Gamma\} \subseteq \Delta$; and
- $\Gamma \in V^L(p)$ iff $p \in \Gamma$.

Remark 8 Note that no M^L is a \mathcal{C}^L -model. To see this, note that both $\{a \in E_k\}$ and $\{\neg a \in E_k\}$ are L-consistent, and by the Lindenbaum Lemma they are members of different states of M^L . Hence $V^L(a \in E_k) \neq W^L$ and $V^L(a \in E_k) \neq \emptyset$, i.e. SU is violated by M^L .

Lemma 6 *M^L is a quasi- \mathcal{C}^L -model.*

Proof Suppose that $L = EA$. We have to show that M^{EA} satisfies the three conditions of Definition 25. For condition 1, suppose that $\Gamma \in V^{EA}(a \in E_k)$ and $\Gamma \mathcal{R}_{Ag}^{EA} \Delta$. From the first part of the hypothesis we have that $a \in E_k \in \Gamma$ (definition of V^{EA}). This implies that $\Gamma \vdash a \in E_k$. Since $\vdash a \in E_k \rightarrow \square_i a \in E_k$ for every $i \in Ag$ (it is derivable from (PIS)), we have that $\Gamma \vdash a \in E_k \rightarrow \square_i a \in E_k$, because \vdash is monotonic. Using MP and Lemma 5(i) we have that (*) $\square_i a \in E_k \in \Gamma$ for every $i \in Ag$. From the second part

of the hypothesis, we have that $\Gamma \mathcal{R}_i^{EA} \Delta$ for some $i \in \text{Ag}$ (definition of \mathcal{R}_{Ag}), which is equivalent by definition of \mathcal{R}^{EA} to $(**) \{ \varphi \mid \Box_i \varphi \in \Gamma \} \subseteq \Delta$ for some $i \in \text{Ag}$. From $(*)$ and $(**)$ we obtain $a \in E_k \in \Delta$ which implies by definition of V^{EA} that $\Delta \in V^{EA}(a \in E_k)$.

As for condition 2, suppose that $\Gamma \notin V^{EA}(a \in E_k)$ and $\Gamma \mathcal{R}_{\text{Ag}}^{EA} \Delta$. From the first part of the hypothesis we get that $a \in E_k \notin \Gamma$ and this implies, by Lemma 5(ii) that $\neg a \in E_k \in \Gamma$. Then we get that $(*) \Box_i \neg a \in E_k \in \Gamma$ for every $i \in \text{Ag}$ by using a similar argument as for condition 1, but this time with axiom (NIS). From the second part of the hypothesis we obtain that $\Gamma \mathcal{R}_i^{EA} \Delta$ for some $i \in \text{Ag}$ (by definition of $\mathcal{R}_{\text{Ag}}^{EA}$), which in turn implies $(**) \{ \varphi \mid \Box_i \varphi \in \Gamma \} \subseteq \Delta$ for some $i \in \text{Ag}$. Finally, from $(*)$ and $(**)$, it is easy to deduce $\neg a \in E_k \in \Delta$, which is equivalent to $\Delta \notin V^{EA}(a \in E_k)$, by definition of V^{EA} .

For condition 3, assume $\wp(A) = \{E_1, \dots, E_n\}$ and take $\Gamma \in W^{EA}$. We have that $\Gamma \vdash \bigwedge_{1 \leq k < m \leq n} E_k \neq E_m$ (axiom ER and monotonicity of \vdash) iff for every $1 \leq k < m \leq n$, $E_k \neq E_m \in \Gamma$ (propositional reasoning and Lemma 5.(ii)) iff for every $1 \leq k < m \leq n$, $\bigvee_{x \in A} \neg(x \in E_k \leftrightarrow x \in E_m) \in \Gamma$ (definition of \neq) iff for every $1 \leq k < m \leq n$ there is some $x \in A$ such that $\neg(x \in E_k \leftrightarrow x \in E_m) \in \Gamma$ (propositional reasoning and Lemma 5.(ii)) iff for every $1 \leq k < m \leq n$ there is some $x \in A$ such that $x \in E_k \leftrightarrow x \in E_m \notin \Gamma$ (Γ is consistent). From the last assertion, using the semantics of \leftrightarrow and Lemma 5.(ii) it is easy to deduce that for every $1 \leq k < m \leq n$ there is some $x \in A$ such that either $(x \in E_k \in \Gamma$ and $x \in E_m \notin \Gamma)$ or $(x \in E_k \notin \Gamma$ and $x \in E_m \in \Gamma)$. Applying the definitions of V^{EA} and \hat{V} , as well as Definition 6, one can deduce that $\hat{V}^{EA}(\Gamma)$ represents an enumeration of $\wp(A)$. Since Γ was an arbitrary state of M^{EA} , we have that every state of M^{EA} represents an enumeration of $\wp(A)$.

Suppose that $L = \text{AoA}$. Proving that M^{AoA} satisfies positive and negative introspection of attacks is completely analogous to proving that M^{EA} is a quasi- \mathcal{EA} -model.

As for PIAw , suppose $\Gamma \mathcal{R}_i^{\text{AoA}} \Delta$ and $a \in A_i(\Gamma)$, which implies $\Gamma \in V^{\text{AoA}}(\text{aw}_i(a))$ by definition of A_i . This entails $\{ \varphi \mid \Box_i \varphi \in \Gamma \} \subseteq \Delta$ and $\text{aw}_i(a) \in \Gamma$, by definition of $\mathcal{R}_i^{\text{AoA}}$ and V^{AoA} , and therefore $\Gamma \vdash \text{aw}_i(a)$ (by definition of \vdash). It follows that $\Box_i \text{aw}_i(a) \in \Gamma$, by axiom PIAw , monotonicity of \vdash and Lemma 5. Therefore $\text{aw}_i(a) \in \Delta$ holds, from which $\Delta \in V^{\text{AoA}}(\text{aw}_i(a))$ follows (definition of V^{AoA}). The latter entails $a \in A_i(\Delta)$ (definition of A_i).

For GNIAw , suppose $\Gamma \mathcal{R}_i^{\text{AoA}} \Delta$ and $a \in A_j(\Delta)$. The first implies $\{ \varphi \mid \Box_i \varphi \in \Gamma \} \subseteq \Delta$ (definition of \mathcal{R}^{AoA}), and the second implies $\Delta \in V^{\text{AoA}}(\text{aw}_j(a))$ (definition of A_j), and therefore $\text{aw}_j(a) \in \Delta$ (definition of V^{AoA}). Suppose, for the sake of contradiction, that $a \notin A_i(\Gamma)$. This entails $\Gamma \notin V^{\text{AoA}}(\text{aw}_i(a))$ (definition of A_i), and therefore $\text{aw}_i(a) \notin \Gamma$ (definition of V^{AoA}), and consequently $\neg \text{aw}_i(a) \in \Gamma$ (Lemma 5). From the latter $\Gamma \vdash \neg \text{aw}_i(a)$ follows (definition of \vdash), which implies $\Box_i \neg \text{aw}_i(a) \in \Gamma$ (axiom GNIAw , monotonicity of \vdash and Lemma 5). Putting both lines of reasoning together we have that $\{ \varphi \mid \Box_i \varphi \in \Gamma \} \subseteq \Delta$, $\text{aw}_j(a) \in \Delta$ and $\Box_i \neg \text{aw}_i(a) \in \Gamma$, which yields $\text{aw}_j(a), \neg \text{aw}_i(a) \in \Delta$, contradicting the consistency of Δ . Therefore $a \in A_i(\Gamma)$.

For $L \in \{S4(EA), S5(EA), KD45(EA), S4(\text{AoA}), S5(EA), KD45(\text{AoA})\}$, it is sufficient to use the previous demonstrations, combined in the obvious way, and the usual

modal arguments for showing that each \mathcal{R}_i^L satisfies the targeted properties (see e.g. Fagin et al. 2004, Theorem 3.1.5). \square

The proof of the Truth Lemma ($\forall \Gamma \in \mathfrak{M}\mathcal{C}^L, \varphi \in \Gamma$ iff $M^L, \Gamma \models \varphi$) is exactly as in the basic modal case. We refer to Blackburn et al. (2002, Lemma 4.21) or Fagin et al. (2004, Theorem 3.1.3) for details.

Finally, we are able to prove Theorem 1 by contraposition. Let us show the case for $L = EA$. Suppose $\Gamma \not\vdash_{EA} \varphi$. By Lindenbaum and the Truth Lemma we obtain $M^{EA}, \Gamma^* \models \Gamma \cup \{\neg\varphi\}$. Since truth is preserved by generated submodels (Lemma 3.5), we have that $M^{EA'}, \Gamma^* \models \Gamma \cup \{\neg\varphi\}$ where $M^{EA'}$ is the submodel of M^{EA} generated by Γ^* using \mathcal{R}_{Ag}^{EA} . Moreover, by Lemma 3.3, $M^{EA'}$ is an \mathcal{EA} -model, and therefore $\Gamma \not\vdash_{\mathcal{EA}} \varphi$. The remaining cases ($L = AoA, L = S4(EA)$, etc) follow along the same lines, combining Lemma 3.1, 3.2 and/or 3.3 in the corresponding way.

A3. Proofs of Sect. 6

Lemma 1 (Closure)

Proof For (i), suppose that $M \otimes E$ is defined. It suffices to show that $M \otimes E$ satisfies ER and SU, and this is an almost direct consequence of the definition of EA-substitutions and the operation \otimes .

For item (ii), assume that $E \in em12$ and $M \otimes E$ is defined. Showing that $M \otimes E$ satisfies ER, SU, and AU is straightforward by the definition of AoA-substitutions and the operation \otimes . Thus, let $M \otimes E = (W', \mathcal{R}', V')$, we just need to show that it satisfies PIAw and GNIAw.

[PIAw] Suppose $(w, s)\mathcal{R}'_i(w', s')$ and $(w, s) \in V'(aw_i(a))$, which is equivalent to $w\mathcal{R}_i w', s\mathcal{I}_i s'$ and $M, w \models \text{pos}(s)(aw_i(a))$ (by definition of \otimes). We need to show that $(w', s') \in V'(aw_i(a))$, which is equivalent to $M, w' \models \text{pos}(s')(aw_i(a))$. We prove this for every possible value of $\text{pos}(s)(aw_i(a))$.

Case A. Suppose that $\text{pos}(s)(aw_i(a)) = \top$. Then $a \in \text{pos}_i^+(s)$ (by definition of pos_i^+). The last assertion, together with the hypothesis that E satisfies EM_1 , leads to $a \in \text{pos}_i^+(s')$, which is equivalent to $\text{pos}(s')(aw_i(a)) = \top$ (by definition of pos_i^+). Therefore, $M, w' \models \text{pos}(s')(aw_i(a))$.

Case B. Suppose that $\text{pos}(s)(aw_i(a)) = aw_i(a)$. Then $M, w \models aw_i(a)$ and, since M is an AoA -model that satisfies positive introspection of awareness and we know that $w\mathcal{R}_i w'$, we have that $M, w' \models aw_i(a)$. Now, consider the three possible values of $\text{pos}(s')(aw_i(a))$. If $\text{pos}(s')(aw_i(a)) = aw_i(a)$, then we already know that $M, w' \models \text{pos}(s')(aw_i(a))$. The case $\text{pos}(s')(aw_i(a)) = \top$ is trivial. Finally, the case $\text{pos}(s')(aw_i(a)) = \perp$ is not possible because EM_1 would force $\text{pos}(s)(aw_i(a)) = \perp$ and we know that is not the case.

Case C. The case $\text{pos}(s)(aw_i(a)) = \perp$ is absurd since we assumed that $M, w \models \text{pos}(s)(aw_i(a))$.

[GNIAw] Suppose that $(w, s)\mathcal{R}'_i(w', s')$ and $(w', s') \in V'(aw_j(a))$ for an arbitrary $j \in Ag$. Applying the definition of \otimes we have that $w\mathcal{R}_i w', s\mathcal{I}_i s'$ and $M, w' \models \text{pos}(s')(aw_j(a))$. We have to show that $(w, s) \in V'(aw_j(a))$, or equivalently that $M, w \models \text{pos}(s)(aw_j(a))$. We prove this by cases on the value of $\text{pos}(s')(aw_j(a))$.

Case A. Suppose that $\text{pos}(s')(aw_j(a)) = \top$. By definition of pos_j^+ , the latter implies $a \in \text{pos}_j^+(s')$. This, together with $sT_i s'$ and the hypothesis that E satisfies EM_2 implies $a \in \text{pos}_i^+(s)$. If this is true, then $\text{pos}(s)(aw_i(a)) = \top$ which clearly implies $M, w \models \text{pos}(s)(aw_i(a))$.

Case B. The case is analogous to case B. of item ii). Suppose that $\text{pos}(s')(aw_j(a)) = aw_j(a)$ and reason by cases on $\text{pos}(s)(aw_i(a))$. Note that $\text{pos}(s)(aw_i(a)) = \top$ makes things trivial. $\text{pos}(s)(aw_i(a)) = \perp$ is incompatible with the assumption of case B and the hypothesis that E satisfies EM_2 . Finally, for $\text{pos}(s)(aw_i(a)) = aw_i(a)$, note that $w\mathcal{R}'_i w', w' \in V(aw_j(a))$ and the fact that M is an $\mathcal{A}o\mathcal{A}$ -model implies that $M, w \models aw_i(a)$, i.e. that $M, w \models \text{pos}(s)(aw_i(a))$.

Case C. The case $\text{pos}(s')(aw_j(a)) = \perp$ is absurd since we assumed that $M, w' \models \text{pos}(s')(aw_j(a))$.

Items (iii) and (iv) are corollaries of Baltag and Renne (2016, *Action Model Closure Theorem*). We sketch the proofs for illustration. For item (iii), suppose $M \in S4(\mathcal{A}o\mathcal{A})$, $E \in \text{em}S4$ and $M \otimes E$ is defined. Let $M \otimes E = (W', \mathcal{R}', V')$. Conditions ER, SU, AU, PIAw, and GNIAw are satisfied by $M \otimes E$ by item (ii) of this lemma, and we just need to show that each \mathcal{R}'_i is both reflexive and transitive. For reflexivity, take $(w, s) \in W'$; we know that $w\mathcal{R}_i w$ and $sT_i s$ (because both relations are assumed to be reflexive). By definition of \mathcal{R}'_i we have that $(w, s)\mathcal{R}'_i(w, s)$. For transitivity, suppose $(w_0, s_0)\mathcal{R}'_i(w_1, s_1)$ and $(w_1, s_1)\mathcal{R}'_i(w_2, s_2)$. This implies $w_0\mathcal{R}_i w_1, w_1\mathcal{R}_i w_2, s_0T_i s_1$, and $s_1T_i s_2$ (by definition of \mathcal{R}'_i), which implies $w_0\mathcal{R}_i w_2$ and $s_0T_i s_2$ (by transitivity of both \mathcal{R}_i and T_i), which implies $(w_0, s_0)\mathcal{R}'_i(w_2, s_2)$ (by definition of \mathcal{R}'_i).

For item iv), transitivity is proved exactly as in the previous item and euclideanness is analogously proved. For seriality, take $(w, s) \in W'$ and note that $w\mathcal{R}_i w'$ for some $w' \in W$ and $sT_i s'$ for some $s' \in E$ (by seriality of \mathcal{R}_i and T_i respectively). Note that $(w', s') \in W'$, because $\text{pre}(s') = \top$ (recall that $E \in \text{pure}$). These two claims jointly imply, by definition of \mathcal{R}'_i , that $(w, s)\mathcal{R}'_i(w', s')$ for some $(w', s') \in W'$, i.e. \mathcal{R}'_i is serial. □

Lemma 2

Proof The proof of the validity of axioms is standard. The validity-preserving character of SE within each class of models, given its corresponding restricted dynamic language, can be proved by induction on the construction of δ . We prove it for $\mathcal{A}o\mathcal{A}$ and $\mathcal{L}^{\text{em}12}$ as an illustration. Item ii) of Lemma 1 plays a fundamental role. Let us just show the inductive step where δ has the form $[E, s]\delta_1$ with $E \in \text{em}12$ and $s \in E[S]$. Assume, as the induction hypothesis that $\models_{\mathcal{A}o\mathcal{A}} \delta_1 \leftrightarrow \delta_1[\varphi/\psi]$. Besides, suppose that $\models_{\mathcal{A}o\mathcal{A}} \varphi \leftrightarrow \psi$. Now, we have a trivial case in which $[E, s]\delta_1 = \varphi$, and then $([E, s]\delta_1)[\varphi/\psi] = \psi$, then showing that $\models_{\mathcal{A}o\mathcal{A}} \delta \leftrightarrow \delta[\varphi/\psi]$ reduces to show $\models_{\mathcal{A}o\mathcal{A}} \varphi \leftrightarrow \psi$, which was already assumed. As for the non trivial case, suppose that $\varphi \neq [E, s]\delta_1$ and note that by the assumption that substitutions do not affect formulas inside dynamic modalities (see Table 3), we have that $([E, s]\delta_1)[\varphi/\psi] = [E, s]\delta_1[\varphi/\psi]$. Therefore, we want to show $\models_{\mathcal{A}o\mathcal{A}} [E, s]\delta_1 \leftrightarrow [E, s]\delta_1[\varphi/\psi]$. For the sake of contradiction, suppose that the latter is not the case. Then, by definition of validity, we have that there are $M \in \mathcal{A}o\mathcal{A}$ and $w \in M[W]$ s.t. $M, w \not\models [E, s]\delta_1 \leftrightarrow [E, s]\delta_1[\varphi/\psi]$. Now, by the meaning of \leftrightarrow we have

two analogous cases: either $[E, s]\delta_1$ is true in M, w and $[E, s]\delta_1[\varphi/\psi]$ is false, or the other way round. Let us just see the first case (the other one is completely analogous). Suppose $M, w \models [E, s]\delta_1$ and $M, w \not\models [E, s]\delta_1[\varphi/\psi]$. Using the semantic clause for $[E, s]$ we can arrive to $M \otimes E, (w, s) \models \delta_1$ and $M \otimes E, (w, s) \not\models \delta_1[\varphi/\psi]$. Finally, by item ii) of Lemma 1 we have that $\not\models_{\mathcal{A}o\mathcal{A}} \delta_1 \leftrightarrow \delta_1[\varphi/\psi]$, which contradicts the induction hypothesis. For the validity preservation of SE within \mathcal{EA} (resp. $S4(\mathcal{A}o\mathcal{A}), \mathcal{KD}45(\mathcal{A}o\mathcal{A})$), the proof follows the same lines, but additionally using item i) (resp. iii, iv) of Lemma 1. \square

Details of the proof of Theorem 2

Soundness follows from the soundness of the static systems and Lemma 2. The proof for strong completeness works, as standard in DEL, via reduction axioms. In order to prove the key reduction lemma we combine techniques from Kooi (2007) with the intuitive idea of defining a reduction function τ with domain the dynamic language and co-domain its static fragment, as in van Ditmarsch et al. (2007), van Benthem et al. (2006), Wang and Cao (2013). This provides a simplified proof of the reduction lemma.

Definition 26 (Complexity measures and reduction function)

- [Depth] Define $d: \mathcal{L}^{ea}(\mathcal{V}_{Ag}^A, \square) \rightarrow \mathbb{N}$ as $d(p) := 0$ for every $p \in \mathcal{V}_{Ag}^A, d(\otimes\varphi) := 1 + d(\varphi)$ where $\otimes \in \{\neg, \square_i, [E, s]\}$ and $d(\varphi \wedge \psi) := 1 + \max(d(\varphi), d(\psi))$.
- [O-depth] Define $Od: \mathcal{L}^{ea}(\mathcal{V}_{Ag}^A, \square) \rightarrow \mathbb{N}$ that returns the number of nested dynamic modalities in each formula. More detailed: $Od(p) := 0, Od(\neg\varphi) := Od(\square_i\varphi) := Od(\varphi), Od(\varphi \wedge \psi) := \max(Od(\varphi), Od(\psi)),$ and $Od([E, s]\varphi) := 1 + Od(\varphi)$.

The function τ is defined for every $\varphi \in \mathcal{L}^*$ (with $\star \subseteq ea$) by the following equations:

$$\begin{aligned} \tau(p) &= p & \tau([E, s]p) &= \text{pre}(s) \rightarrow \text{pos}(s)p \\ \tau(\neg\varphi) &= \neg\tau(\varphi) & \tau([E, s]\neg\varphi) &= \text{pre}(s) \rightarrow \neg\tau([E, s]\varphi) \\ \tau(\varphi \wedge \psi) &= \tau(\varphi) \wedge \tau(\psi) & \tau([E, s](\varphi \wedge \psi)) &= \tau([E, s]\varphi) \wedge \tau([E, s]\psi) \\ \tau(\square_i\varphi) &= \square_i\tau(\varphi) & \tau([E, s]\square_i\varphi) &= \text{pre}(s) \rightarrow \bigwedge_{s\mathcal{T}_i} \square_i\tau([E, t]\varphi) \\ & & \tau([E, s][F, t]\varphi) &= \tau([E, s]\tau([F, t]\varphi)) \end{aligned}$$

We sometimes lift the domain of τ from formulas to sets of formulas $\tau(\Delta) = \{\tau(\varphi) \mid \varphi \in \Delta\}$.

Lemma 7 For every $\varphi \in \mathcal{L}^*: \tau(\varphi) \in \mathcal{L}$.

Proof (sketched) We first need to show that the Lemma holds for the special case where $Od(\varphi) = 1$. So suppose that $Od(\varphi) = 1$ and continue by induction on $d(\varphi)$. We skip details and just note that in the definition of τ , the function is applied to a formula of smaller depth in the right-hand side of the equations. The only problematic case is $\tau([E, s][F, t]\psi) = \tau([E, s]\tau([F, t]\psi))$ that we do not need to consider now because $Od([E, s][F, t]\psi) > 1$.

Now, the Lemma can be proved for every formula φ by induction on $d(\varphi)$. When we arrive at $\varphi = [E, s][F, t]\psi$, we obtain $\tau([E, s][F, t]\psi) = \tau([E, s]\tau([F, t]\psi))$. Applying the induction hypothesis we have that $\tau([F, t]\psi) \in \mathcal{L}$, but then $Od([E, s]\tau([F, t]\psi)) = 1$ which reduces to the case above. \square

Lemma 8 For every $\varphi \in \mathcal{L}^{ea}$ (resp. $\varphi \in \mathcal{L}^{em12}$, $\varphi \in \mathcal{L}^{emS4}$, $\varphi \in \mathcal{L}^{pure}$): $\vdash_{EA^{ea}} \varphi \leftrightarrow \tau(\varphi)$ (resp. $\vdash_{AoA^{em12}} \varphi \leftrightarrow \tau(\varphi)$, $\vdash_{S4(AoA)^{emS4}} \varphi \leftrightarrow \tau(\varphi)$, $\vdash_{KD45(AoA)^{pure}} \varphi \leftrightarrow \tau(\varphi)$).

Proof (sketched) We follow the same strategy as for the previous lemma, i.e. proving it first for the special case where $Od(\varphi) = 1$, by induction on $d(\varphi)$. We omit details. After that, we are able to prove it in general, again by induction on $d(\varphi)$. We just illustrate the case of $\varphi = [E, s][F, t]\psi$. Let $L \in \{EA^{ea}, AoA^{em12}, S4(AoA)^{emS4}, KD45(AoA)^{pure}\}$. Note that, by the induction hypothesis we have that

$$\vdash_L [F, t]\psi \leftrightarrow \tau([F, t]\psi).$$

Taking the propositional tautology

$$\vdash_L [E, s][F, t]\psi \leftrightarrow [E, s][F, t]\psi$$

we apply SE and obtain

$$\vdash_L [E, s][F, t]\psi \leftrightarrow [E, s]\tau([F, t]\psi).$$

Note that $\tau([F, t]\psi) \in \mathcal{L}$ (by Lemma 7). But then $Od([E, s]\tau([F, t]\psi)) = 1$ and by the previous statement we have:

$$\vdash_L [E, s]\tau([F, t]\psi) \leftrightarrow \tau([E, s]\tau([F, t]\psi)),$$

and applying SE again we have:

$$\vdash_L [E, s][F, t]\psi \leftrightarrow \tau([E, s]\tau([F, t]\psi)).$$

□

As an immediate corollary of Lemmas 2 and 8, we have that for every $\varphi \in \mathcal{L}^{ea}$ (resp. $\varphi \in \mathcal{L}^{em12}$, $\varphi \in \mathcal{L}^{emS4}$, $\varphi \in \mathcal{L}^{pure}$) $\vDash_{\mathcal{EA}} \varphi \leftrightarrow \tau(\varphi)$ (resp. $\vDash_{AoA} \varphi \leftrightarrow \tau(\varphi)$, $\vDash_{S4(AoA)} \varphi \leftrightarrow \tau(\varphi)$, $\vDash_{KD45(AoA)} \varphi \leftrightarrow \tau(\varphi)$). Finally, strong completeness follows from Theorem 1 in the usual way (Baltag et al. 2016). Let us just show the case for AoA^{em12} for illustration. Suppose $\Gamma \vDash_{AoA} \varphi$, then $\tau(\Gamma) \vDash_{AoA} \tau(\varphi)$ (by the corollary mentioned above). By Lemma 7, we know that $\tau(\Gamma) \subseteq \mathcal{L}$ and $\tau(\varphi) \in \mathcal{L}$, which implies, together with $\tau(\Gamma) \vDash_{AoA} \tau(\varphi)$ and Theorem 1 that $\tau(\Gamma) \vdash_{AoA} \tau(\varphi)$. But since AoA^{em12} is an extension of AoA , we have that $\tau(\Gamma) \vdash_{AoA^{em12}} \tau(\varphi)$. Applying the definition of deduction from assumptions (see p. 47), Lemma 8 and SE, we have that $\Gamma \vdash_{AoA^{em12}} \varphi$.

A4. Preserving doxastic relations via an enhanced truth condition

Balbiani et al. (2012) introduce an enhanced truth condition for public announcement operators, establishing as a necessary condition for $\langle \psi \rangle \varphi$ to be true that the updated model M^ψ must be in the targeted class. Their axiomatisations make use of the global

modality and apply to public announcements logic. In a different work, Aucher (2008) captures a sufficient condition, expressible in a language with a global modality, for seriality to be preserved under product update. We can therefore obtain the axiomatisation of $\mathcal{KD45}(\mathcal{A}o\mathcal{A})$ for a broader dynamic language than the one used in Theorem 2, by putting together both these results and our construction of Lemma 1. First of all, we need to augment our language with a global modality $[U]$. Formulas of the language $\mathcal{L}^*(\mathcal{V}_{Ag}^A, \square, [U])$ (for any $\star \subseteq ea$) are generated by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \square_j\varphi \mid [U]\varphi \mid [E, s]\varphi \quad p \in \mathcal{V}_{Ag}^A, j \in Ag, E \in \star, s \in E[S]$$

where $[U]\varphi$ reads “it is everywhere the case that φ ”. Now, we define the enhanced truth condition, denoted by \Vdash . The \Vdash -truth clauses for propositional variables, Boolean connectors and epistemic operators are the same as for \models . As for $[U]$, the clause is also the standard one: $M, w \Vdash [U]\varphi$ iff $\forall w' \in M[W] : M, w' \models \varphi$. However, we need to change the truth condition for the dynamic modalities, adapting the one of Balbiani et al. (2012) to our purposes:

$$M, w \Vdash [E, s]\varphi \text{ iff } (M, w \Vdash \text{pre}(s) \text{ and } M \otimes E \in \mathcal{KD45}(\mathcal{A}o\mathcal{A})) \text{ implies } M \otimes E, (w, s) \Vdash \varphi.$$

Since we want to apply again a reduction argument, a previous step is needed, that is, a sound and strongly complete axiomatisation in the extended static language $\mathcal{L}(\mathcal{V}_{Ag}^A, \square, [U])$ w.r.t. $\mathcal{KD45}(\mathcal{A}o\mathcal{A})$. More concretely, the proof system we are looking for is $\mathcal{KD45}(\mathcal{A}o\mathcal{A})_{[U]}$ which is the result of extending $\mathcal{KD45}(\mathcal{A}o\mathcal{A})$ with (i) S5 axioms schemes for $[U]$; (ii) the axiom scheme (INC) $[U]\varphi \rightarrow \square_i\varphi$ and the rule $\text{NEC}^{[U]}$ (from φ , infer $[U]\varphi$). Soundness and strong completeness of $\mathcal{KD45}(\mathcal{A}o\mathcal{A})_{[U]}$ w.r.t. $\mathcal{KD45}(\mathcal{A}o\mathcal{A})$ follows from Blackburn et al. (2002, Theorem 7.3). The same comments made in the proof of Theorem 2 apply here.⁴⁷ Note that now axioms PIAt and NIAt (resp. PIS and NIS) can be rewritten more transparently as a unique axiom $[U]a \rightsquigarrow b \vee [U]\neg a \rightsquigarrow b$ (resp. $[U]a \in E_k \vee [U]\neg a \in E_k$).

For reduction, we need to capture the new precondition for dynamic modalities (the one imposed in the definition of \Vdash) in the object language. As proved in Aucher (2008, Proposition 2), the following formula is true in a pointed model M, w iff $M \otimes E$ is defined and serial:

$$f(E) := \langle U \rangle \left(\bigvee_{s \in E[S]} \text{pre}(s) \wedge [U] \bigwedge_{s \in E[S]} (\text{pre}(s) \rightarrow \bigwedge_{i \in Ag} \diamond_i \bigvee_{s \mathcal{T}_i} \text{pre}(t)) \right).$$

Let emd45 denote the class of event models satisfying EM_1 and EM_2 and where each \mathcal{T}_i is serial, transitive and euclidean. Using this result and Lemma 1.(ii), the following can be proved:

Lemma 9 *Let M, w be a pointed model s.t. $M \in \mathcal{KD45}(\mathcal{A}o\mathcal{A})$ and let $E \in \text{emd45}$, then $M \otimes E \in \mathcal{KD45}(\mathcal{A}o\mathcal{A})$ iff $M, w \Vdash f(E)$.*

⁴⁷ Again, although $\mathcal{KD45}(\mathcal{A}o\mathcal{A})_{[U]}$ is not a normal modal logic in the sense of Blackburn et al. (2002), because the rule of uniform substitution is not sound, the proof of Blackburn et al. (2002, Theorem 7.3) can be reproduced without problems.

Table 4 Reduction axioms with $f(E)$

<i>Axioms</i>	
$[E, s]p \leftrightarrow ((\text{pre}(s) \wedge f(E)) \rightarrow \text{pos}(s)(p))$	$(\text{Atm})^f$
$[E, s]\neg\varphi \leftrightarrow ((\text{pre}(s) \wedge f(E)) \rightarrow \neg[E, s]\varphi)$	$(\text{Neg})^f$
$[E, s](\varphi \wedge \psi) \leftrightarrow ([E, s]\varphi \wedge [E, s]\psi)$	$(\text{Con})^f$
$[E, s]\Box_i\varphi \leftrightarrow ((\text{pre}(s) \wedge f(E)) \rightarrow \bigwedge_{s\mathcal{T}_it} \Box_i[E, t]\varphi)$	$(\text{Box})^f$
$[E, s][U]\varphi \leftrightarrow ((\text{pre}(s) \wedge f(E)) \rightarrow \bigwedge_{t \in E[S]} [U][E, t]\varphi)$	$(\text{Uni})^f$
<i>Rules</i>	
From $\varphi \leftrightarrow \psi$, infer $\delta \leftrightarrow \delta[\varphi/\psi]$	SE

This lemma is a key to derive the main completeness theorem.

Theorem 3 *The proof system $KD45(\text{AoA})_{[U]}^{\text{emd45}}$, written in $\mathcal{L}^{\text{emd45}}(\mathcal{V}_{\text{Ag}}^A, \Box, [U])$, that extends axioms and rules of $KD45(\text{AoA})_{[U]}$ with those of Table 4⁴⁸ is sound and complete w.r.t. the class $\mathcal{KD45}(\text{AoA})$ (if the enhanced truth condition \Vdash is assumed).*

Proof The proof is analogous to the one of Theorem 2. Lemma 9 is needed to prove the validity of the new reduction axioms. After the reduction, we make use of the completeness of $KD45(\text{AoA})_{[U]}$ mentioned above. We need however to redefine the translation function accordingly, i.e define a new $\tau' : \mathcal{L}^{\text{emd45}}(\mathcal{V}_{\text{Ag}}^A, \Box, [U]) \rightarrow \mathcal{L}(\mathcal{V}_{\text{Ag}}^A, \Box, [U])$ that extends τ with the equation $\tau'([U]\varphi) = [U]\tau'(\varphi)$, and differs from τ in the equations for $[E, s]\varphi$:

$$\begin{aligned} \tau'([E, s]p) &= (\text{pre}(s) \wedge f(E)) \rightarrow \text{pos}(s)(p), \\ \tau'([E, s]\neg\varphi) &= (\text{pre}(s) \wedge f(E)) \rightarrow \neg\tau'([E, s]\varphi), \\ \tau'([E, s](\varphi \wedge \psi)) &= \tau'([E, s]\varphi) \wedge \tau'([E, s]\psi), \\ \tau'([E, s]\Box_i\varphi) &= (\text{pre}(s) \wedge f(E)) \rightarrow \bigwedge_{s\mathcal{T}_it} \Box_i\tau'([E, t]\varphi), \\ \tau'([E, s][U]\varphi) &= (\text{pre}(s) \wedge f(E)) \rightarrow \bigwedge_{t \in E[S]} [U]\tau'([E, t]\varphi), \\ \tau'([E, s][F, t]\varphi) &= \tau'([E, s]\tau'([F, t]\varphi)). \end{aligned} \quad \square$$

We would like to mention another application of the enhanced truth condition approach. Recall that EM_1 and EM_2 are just sufficient conditions for PIAw and GNIAw to be preserved after product update. However, sufficient and necessary conditions can be provided by using the more expressive language $\mathcal{L}^*(\mathcal{V}_{\text{Ag}}^A, \Box, [U])$. Let us define the two characterizing formulas as follows

$$\begin{aligned} \text{PIAw}(E) &:= [U] \bigwedge_{s \in E[S]} \bigwedge_{i \in \text{Ag}} \bigwedge_{a \in A} \left((\text{pre}(s) \wedge \text{pos}(s)(\text{aw}_i(a))) \rightarrow \right. \\ &\quad \left. \Box_i \left(\bigwedge_{s\mathcal{T}_it} (\text{pre}(t) \rightarrow \text{pos}(t)(\text{aw}_i(a))) \right) \right). \end{aligned}$$

⁴⁸ Let us point out that, for the axiomatisation, $f(E)$ can be simplified to $f'(E) := [U] \bigwedge_{s \in E[S]} (\text{pre}(s) \rightarrow \bigwedge_{i \in \text{Ag}} \Diamond_i \bigvee_{s\mathcal{T}_it} \text{pre}(t))$; since the standard preconditions for the executability of event models ($\text{pre}(s)$) imply the first conjunct of Aucher's $f(E)$.

$$\text{GNIAw}(E) := [\text{U}] \bigwedge_{s \in E[S]} \left(\text{pre}(s) \rightarrow \bigwedge_{i \in \text{Ag}} \bigwedge_{a \in A} \left(\neg \text{pos}(s)(\text{aw}_i(a)) \rightarrow \right. \right. \\ \left. \left. \square_i \left(\bigwedge_{s \mathcal{T}_i t} (\text{pre}(t) \rightarrow \bigwedge_{j \in \text{Ag}} \neg \text{pos}(t)(\text{aw}_j(a))) \right) \right) \right).$$

The formula $\text{PIAw}(E)$ says that everywhere in the model, any point that meets the precondition of some action s and satisfies $\text{aw}_i(a)$ as a postcondition for s has only access to states that satisfy the same postcondition for any t which is related to s and executable. Similarly, $\text{GNIAw}(E)$ says that, everywhere, when $\text{aw}_i(a)$ is not satisfied as a postcondition of some executable s , then $\text{aw}_j(a)$ is not satisfied at any accessible state as a postcondition for any executable t which is related to s . It is then a matter of standard verification to prove the two following items.

Proposition 6 *Let $M \in \mathcal{EA}$ and $w \in M[W]$, then $M, w \models \text{PIAw}(E)$ iff $M \otimes E$ satisfies PIAw .*

Proposition 7 *Let $M \in \mathcal{EA}$ and $w \in M[W]$, then $M, w \models \text{GNIAw}(E)$ iff $M \otimes E$ satisfies GNIAw .*

A.5. Incomplete and control AFs

We close this technical appendix by proving the results stated in Sect. 8.

Proof of Proposition 2

Proof We just prove the first item (the other one is analogous).

Suppose that the answer to Pr-PSA with input $\text{IAF} = (A, A^?, R, R^?)$ and $a \in A$ is yes.

$$\iff$$

There is a completion (A^*, R^*) of IAF s.t. $\forall E \in \text{Pr}((A^*, R^*), a \in E)$ (def Pr-PSA).

$$\iff$$

There is a $u \in W^{\text{IAF}}$ s.t. $\forall E \in \text{Pr}((A(u), R(u)), a \in E)$ (Definition 19 and 8).

$$\iff$$

There is a $u \in W^{\text{IAF}}$ s.t. a is strongly accepted w.r.t. $(A(u), R(u))$ (Definition 4).

$$\iff$$

There is a $u \in W^{\text{IAF}}$ s.t. $\hat{V}(u) \models \text{stracc}(a)$ (Proposition 1, item 5. Note that, by Definition 19 we have that $\hat{V}(u) = v_{\text{MAF}}$ with $\text{MAF} = (A \cup A^?, (R \cup R^?)|_{A^*}, \{A^*\}, \{E_1, \dots, E_n\})$).

⟷

There is a $u \in W^{IAF}$ s.t. $M^{IAF}, u \models \text{stracc}(a)$ (Definition 9 and the fact that $\text{stracc}(a)$ does not contain modal operators, see p. 12).

⟷

$M^{IAF}, w \models \diamond \text{stracc}(a)$ (by the fact that \mathcal{R}^{IAF} is total (Definition 19) and the truth clause for \diamond). □

Proof of Proposition 3

Proof We just prove the second item, the first runs analogously.

Suppose that $M, w \models \square \bigvee_{1 \leq k \leq n} (\text{preferred}(E_k) \wedge a \in E_k)$.

⟷

For all $u \in M[W]$, $M, u \models \bigvee_{1 \leq k \leq n} (\text{preferred}(E_k) \wedge a \in E_k)$ (because \mathcal{R} is total by assumption).

⟷

For all $u \in M[W]$, $\hat{V}(u) \models \bigvee_{1 \leq k \leq n} (\text{preferred}(E_k) \wedge a \in E_k)$ (because no modal operators occurs in $\bigvee_{1 \leq k \leq n} (\text{preferred}(E_k) \wedge a \in E_k)$, see p. 12).

⟷

For all $u \in M[W]$, there is some $1 \leq k \leq n$: $\hat{V}(u) \models \text{preferred}(E_k) \wedge a \in E_k$ (truth clause for \vee).

⟷

For all $u \in M[W]$, there is a $E \in \text{Pr}((A_u^*, R_u^*))$ s.t. $a \in E$ (Proposition 1.4 and the fact that $\hat{V}(u) = v_{MAF}$ where $MAF = (C, R(u), \{A(u)\}, \{E_1, \dots, E_n\})$).

⟷

For every completion (A^*, R^*) of IAF_M , there is a $E \in \text{Pr}((A^*, R^*))$: $a \in E$ (because M exhausts the completions of IAF_M by assumption and every (A_u^*, R_u^*) is a completion of IAF_M , see Definition 20).

⟷

The answer to Pr-PSA with input IAF_M and $a \in A_M$ is yes (definition of Pr-PSA). □

Proof of Proposition 5

Proof Suppose that the answer to Pr-NSCon with input CAF and $a \in A$ is yes.

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: for every completion (A^*, R^*) of CAF_{CFG} we have that for every $E \in \text{Pr}((A^*, R^*))$, $a \in E$ (def NSCon).

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: for every $u \in M^{\text{CFG}}[W]$, for every $E \in \text{Pr}((A_u^*, R_u^*))$, $a \in E$ (Remark 7).

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: for every $u \in M^{\text{CFG}}[W]$, a is strongly accepted w.r.t. (A_u^*, R_u^*) (Definition 4).

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: for every $u \in M^{\text{CFG}}[W]$, $\hat{V}(u) \models \text{stracc}_{\text{OPP}}(a)$ (Prop. 1.5, Remark 7, and the fact that $\hat{V}(u) = v_{\text{MAF}}$ where $\text{MAF} = (\Delta^{\text{CAF}}, R^{\text{CFG}}(u), \{A_{\text{PRO}}^{\text{CFG}}(u), A_{\text{OPP}}^{\text{CFG}}(u)\}, \{E_1, \dots, E_n\})$).

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: for every $u \in M^{\text{CFG}}[W]$, $M^{\text{CFG}}, u \models \text{stracc}_{\text{OPP}}(a)$ (Definition 9 and the fact that stracc does not contain modal operators, see p. 12).

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: $M^{\text{CFG}}, (w, \Delta) \models \square_{\text{PRO}} \text{stracc}_{\text{OPP}}(a)$ (note that $\mathcal{R}_{\text{PRO}}^{\text{CAF}}$ is total in M^{CFG} because it is total in M^{CAF} (Definition 24) and the execution of Pub^{CFG} does not vary accessibility relations (see Definition 14 and Example 5)).

$$\iff$$

$\exists \text{CFG} \subseteq A_C$: $M^{\text{CAF}}, w \models [\text{Pub}^{\text{CFG}}, \Delta] \square_{\text{PRO}} \text{stracc}_{\text{OPP}}(a)$ (Definitions 24, 14).

$$\iff$$

$\exists \text{CFG} \subseteq \{x \in \Delta^{\text{CAF}} \mid M^{\text{CAF}}, w \models \square_{\text{PRO}} \neg a w_{\text{OPP}}(x)\}$: $M^{\text{CAF}}, w \models [\text{Pub}^{\text{CFG}}, \Delta] \square_{\text{PRO}} \text{stracc}_{\text{OPP}}(a)$ (Prop. 4).

$$\iff$$

$\exists \text{CFG} \subseteq \Delta^{\text{CAF}}: M^{\text{CAF}}, w \models \text{private}_{\text{PRO}}(\text{CFG})$ and $M^{\text{CAF}}, w \models [\text{Pub}^{\text{CFG}}, \Delta] \square_{\text{PRO}} \text{stracc}_{\text{OPP}}(a)$ (def $\text{private}_{\text{PRO}}$).

\iff

$M^{\text{CAF}}, w \models \bigvee_{1 \leq l \leq n} (\text{private}_{\text{PRO}}(E_l) \wedge [\text{Pub}^{E_l}, \Delta] \square_{\text{PRO}} \text{stracc}_{\text{OPP}}(a))$ (propositional reasoning, variable renaming (CFG by E_l) and the fact that $\wp(\Delta^{\text{CAF}})$ has n elements by assumption). \square

References

- Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2), 510–530.
- Andersen, M. B., Bolander, T., & Jensen, M. H. (2012). Conditional epistemic planning. In L. Fariñas del Cerro, Herzog, A., & Mengin, J. (Eds.), *Logics in Artificial Intelligence*, volume 7519 of *LNCS* (pp. 94–106). Springer.
- Aucher, G. (2008). Consistency preservation and crazy formulas in BMS. In S. Hålldobler, Lutz, C., & Wansing, H. (Eds.), *European Workshop on Logics in Artificial Intelligence*, volume 5293 of *LNCS* (pp. 21–33). Springer.
- Aucher, G., & Schwarzenrüber, F. (2013). On the complexity of dynamic epistemic logic. In B. Schipper (Ed.), *Proceedings of the 14th Theoretical Aspects of Rationality and Knowledge (TARK XIV)* (pp. 19–28). ACM.
- Balbani, P., van Ditmarsch, H., Herzig, A., & De Lima, T. (2012). Some truths are best left unsaid. In T. Bolander, Braüner, T., Ghilardi, S., & Moss, L. (Eds.), *Advances in modal logic* (Vol. 9, pp. 36–54). College Publication.
- Baltag, A., & Moss, L. S. (2004). Logics for epistemic programs. *Synthese*, 139(2), 165–224.
- Baltag, A., Moss, L. S., & Solecki, S. (2016 [1998]). The logic of public announcements, common knowledge, and private suspicions. In H. Arló-Costa, Hendricks, V. F., & van Benthem, J. (Eds.), *Readings in formal epistemology* (pp. 773–812). Springer.
- Baltag, A., & Renne, B. (2016). Dynamic epistemic logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.
- Baltag, A., & Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. In van der Hoek, W., Bonanno, G., & Wooldridge, M., editors, *Logic and the foundations of game and decision theory (LOFT 7)*, volume 3 of *Texts in Logic and Games* (pp. 9–58). Amsterdam University Press.
- Baroni, P., Caminada, M., & Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. In P. Baroni, Gabbay, D. M., Giacomin, M., & van der Torre, L. (Eds.), *Handbook of formal argumentation* (pp. 159–236). College Publications.
- Baroni, P., Giacomin, M., & Guida, G. (2005). SCC-recursiveness: A general schema for argumentation semantics. *Artificial Intelligence*, 168(1–2), 162–210.
- Baroni, P., Rago, A., & Toni, F. (2019). From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. *International Journal of Approximate Reasoning*, 105, 252–286.
- Baumeister, D., Neugebauer, D., & Rothe, J. (2018a). Credulous and skeptical acceptance in incomplete argumentation frameworks. In S. Modgil, Budzyska, K., & Lawrence, J. (Eds.), *Proceedings of the COMMA 2018*, volume 305 of *Frontiers in Artificial Intelligence and Applications*, pages 181–192. IOS Press.
- Baumeister, D., Neugebauer, D., Rothe, J., & Schadrack, H. (2018b). Verification in incomplete argumentation frameworks. *Artificial Intelligence*, 264, 1–26.
- Beirlaen, M., Heyninck, J., Pardo, P., & Straßer, C. (2018). Argument strength in formal argumentation. *IfCoLog Journal of Logics and their Applications*, 5(3), 629–675.
- Besnard, P., Cayrol, C., & Lagasque-Schiex, M.-C. (2020). Logical theories and abstract argumentation: A survey of existing works. *Argument & Computation*, 11(1–2), 41–102.
- Besnard, P., & Doutre, S. (2004). Checking the acceptability of a set of arguments. In J. P. Delgrande, & Schaub, T. (Eds.), *Proceedings of the NMR*, (pp. 59–64). AAAI Press.

- Besnard, P., Doutre, S., & Herzig, A. (2014). Encoding argument graphs in logic. In A. Laurent, Strauss, O., Bouchon-Meunier, B., & Yager, R. (Eds.), *International conference on information processing and management of uncertainty in knowledge-based systems*, volume 443 of *communications in computer and information science* (pp. 345–354). Springer.
- Black, E., Coles, A. J., & Hampson, C. (2017). Planning for persuasion. In *AAMAS 2017*, (pp. 933–942). IFAAMAS.
- Blackburn, P., De Rijke, M., & Venema, Y. (2002). *Modal logic*. Cambridge University Press.
- Booth, R., Kaci, S., Rienstra, T., & van der Torre, L. (2013). A logical theory about dynamics in abstract argumentation. In W. Liu, Subrahmanian, V. S., & Wijsen, J. (Eds.), *Scalable uncertainty management*, volume 8070 of *LNCSS* (pp. 148–161). Springer.
- Caminada, M. (2006). On the issue of reinstatement in argumentation. In M. Fisher, van der Hoek, W., Konev, B., & Lisitsa, A. (Eds.), *Logics in artificial intelligence. JELIA 2006* volume 4160 of *LNCSS* (pp. 111–123). Springer.
- Caminada, M., & Sakama, C. (2017). On the issue of argumentation and informedness. In M. Otake, Kurahashi, S., Ota, Y., Satoh, K., & Bekki, D. (Eds.), *New Frontiers in artificial intelligence. JSAI-isAI 2015. LNCSS* (Vol. 10091, pp. 317–330). Springer.
- Caminada, M. W., & Gabbay, D. M. (2009). A logical account of formal argumentation. *Studia Logica*, 93(2–3), 109–145.
- Carmel, D., & Markovitch, S. (1996a). Incorporating opponent models into adversary search. In G. Weiß & Sen, S. (Eds.), *Proceedings of the thirteenth national conference on artificial intelligence* (pp. 120–125). AAAI Press.
- Carmel, D., & Markovitch, S. (1996b). Opponent modeling in multi-agent systems. In G. Weiß, & Sen, S. (Eds.), *Adaption and learning in multi-agent systems* (pp. 40–52). Springer.
- Cayrol, C., de Saint-Cyr, F. D., & Lagasquie-Schiex, M. (2010). Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38, 49–84.
- Cayrol, C., Devred, C., & Lagasquie-Schiex, M. C. (2007). Handling ignorance in argumentation: Semantics of partial argumentation frameworks. In K. Mellouli (Ed.), *Symbolic and quantitative approaches to reasoning with uncertainty* (pp. 259–270). Springer.
- Cerutti, F., Dunne, P. E., Giacomini, M., & Vallati, M. (2013). Computing preferred extensions in abstract argumentation: A SAT-based approach. In *International workshop on theory and applications of formal argumentation* (pp. 176–193). Springer.
- Coste-Marquis, S., Devred, C., Konieczny, S., Lagasquie-Schiex, M.-C., & Marquis, P. (2007). On the merging of dung's argumentation systems. *Artificial Intelligence*, 171(10–15), 730–753.
- Cramer, M., & Guillaume, M. (2018). Empirical cognitive study on abstract argumentation semantics. In S. Modgil, Budzynska, K., & Lawrence, J. (Eds.), *Proceedings of the COMMA 2018*, volume 305 of *Frontiers in artificial intelligence and applications* (pp. 413–424). IOS Press.
- de Saint-Cyr, F. D., Bisquert, P., Cayrol, C., & Lagasquie-Schiex, M.-C. (2016). Argumentation update in YALLA (yet another logic language for argumentation). *International Journal of Approximate Reasoning*, 75, 57–92.
- Dimopoulos, Y., Maily, J.-G., & Moraitis, P. (2018). Control argumentation frameworks. In *Thirty-second AAAI conference on artificial intelligence* AAAI Press.
- Doutre, S., Herzig, A., & Perrussel, L. (2014). A dynamic logic framework for abstract argumentation. In C. Baral, De Giacomo, G., & Eiter, T. (Eds.), *Fourteenth international conference on the principles of knowledge representation and reasoning* (pp. 62–71). AAAI Press.
- Doutre, S., Maffre, F., & McBurney, P. (2017). A dynamic logic framework for abstract argumentation: adding and removing arguments. In S. Benferhat, Tabia, K., & Ali, M. (Eds.), *International conference on industrial, engineering and other applications of applied intelligent systems*, volume 10351 of *LNCSS* (pp. 295–305). Springer.
- Doutre, S., & Maily, J.-G. (2018). Constraints and changes: A survey of abstract argumentation dynamics. *Argument & Computation*, 9(3), 223–248.
- Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2), 321–357.
- Dvořák, W., Szeider, S., & Woltran, S. (2012). Abstract argumentation via monadically second order logic. In E. Hüllermeier, Link, S., Fober, T., & Seeger, B. (Eds.), *Scalable uncertainty management*, volume 7520 of *LNCSS* (pp. 85–98). Springer.

- Dyrkolbotn, S. K., & Pedersen, T. (2016). Arguably argumentative: A formal approach to the argumentative theory of reason. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 317–339). Springer.
- Fagin, R., & Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1), 39–76.
- Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. (2004). *Reasoning about knowledge*. MIT Press.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Fischer, M. J., & Ladner, R. E. (1979). Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2), 194–211.
- Gerbrandy, J., & Groeneveld, W. (1997). Reasoning about information change. *Journal of Logic, Language and Information*, 6(2), 147–169.
- Grossi, D. (2010a). Argumentation in the view of modal logic. In P. McBurney, Rahwan, I., & Parsons, S. (Eds.), *International workshop on argumentation in multi-agent systems*, volume 6614 of *LNCS* (pp. 190–208). Springer.
- Grossi, D. (2010b). On the logic of argumentation theory. In W. van der Hoek, Kaminka, G., Lesperance, Y., Luck, M., & Sen, S. (Eds.), *AAMAS 2010* (pp. 409–416). IFAAMAS.
- Hadjnikolis, C., Siantos, Y., Modgil, S., Black, E., & McBurney, P. (2013). Opponent modelling in persuasion dialogues. In F. Rossi (Ed.), *Twenty-third international joint conference on artificial intelligence*. AAAI Press.
- Hamblin, C. L. (1970). *Fallacies*. Vale Press.
- Herzig, A., Lorini, E., & Maffre, F. (2018). Possible worlds semantics based on observation and communication. In H. van Ditmarsch, & Sandu, G. (Eds.), *Jaakko Hintikka on knowledge and game-theoretical semantics* (pp. 339–362). Springer.
- Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Cornell University Press.
- Hunter, A. (2018). Towards a framework for computational persuasion. *Argument & Computation*, 9, 15–40.
- Kelly, T. (2008). Disagreement, dogmatism, and belief polarization. *Journal of Philosophy*, 105(10), 611–633.
- Kooi, B. (2007). Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17(2), 231–253.
- Li, H., Oren, N., & Norman, T. J. (2011). Probabilistic argumentation frameworks. In S. Modgil, Oren, N., & Toni, F. (Eds.), *International workshop on theory and applications of formal argumentation*, volume 7312 of *LNCS* (pp. 1–16). Springer.
- Mäs, M., & Flache, A. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PloS One*, 8(11).
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.
- Meyer, J.-J.C., & van der Hoek, W. (1995). *Epistemic logic for AI and computer science*. Cambridge University Press.
- Oren, N., & Norman, T. J. (2009). Arguing using opponent models. In McBurney, P., Rahwan, I., Parsons, S., & N., M., editors, *International workshop on argumentation in multi-agent systems*, volume 6057 of *LNCS*, pages 160–174. Springer.
- Parikh, R. (2008). Sentences, belief and logical omniscience, or what does deduction tell us? *The Review of Symbolic Logic*, 1(4), 459–476.
- Perelman, C., & Olbrechts-Tyteca, L. (1958). *Traité de l'argumentation*. La nouvelle rhétorique: Éditions de l'université de Bruxelles.
- Perkins, D., Bushey, B., & Farady, M. (1986). Learning to reason (final report for grant no. nie-g-83_0028).
- Plaza, J. (1989). Logics of public announcements. In M. Emrich, Pfeifer, M., Hadzikadic, M., & Ras, Z. (Eds.), *Proceedings 4th international symposium on methodologies for intelligent systems* (pp. 201–216). Oak Ridge National Laboratory.
- Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science*, 11(4), 481–518.
- Pollock, J. L. (1991). A theory of defeasible reasoning. *International Journal of Intelligent Systems*, 6(1), 33–54.
- Rahwan, I., & Larson, K. (2009). Argumentation and game theory. In G. Simari, & Rahwan, I. (Eds.), *Argumentation in artificial intelligence*, pages 321–339. Springer.

- Rahwan, I., Madakkatel, M. I., Bonnefon, J.-F., Awan, R. N., & Abdallah, S. (2010). Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. *Cognitive Science*, 34(8), 1483–1502.
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13(1–2), 81–132.
- Rienstra, T., Thimm, M., & Oren, N. (2013). Opponent models with uncertainty for strategic argumentation. In F. Rossi (Ed.), *Twenty-third international joint conference on artificial intelligence*. AAAI Press.
- Rodenhäuser, L. B. (2014). *A matter of trust: Dynamic attitudes in epistemic logic*. PhD thesis.
- Sakama, C. (2012). Dishonest arguments in debate games. In B. Verheij, Szeider, S., & Woltran, S. (Eds.), *Proceedings of the COMMA 2012, Frontiers in Artificial Intelligence and Applications* (pp. 177–184). IOS Press.
- Schwarzentruber, F., Vesic, S., & Rienstra, T. (2012). Building an epistemic logic for argumentation. In L. Fariñas del Cerro, Herzog, A., & Mengin, J. (Eds.), *Logics in artificial intelligence*, volume 7519 of *LNCSE* (pp. 359–371). Springer.
- Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies*, 128(1), 169–199.
- Thimm, M. (2014). Strategic argumentation in multi-agent systems. *KI-Künstliche Intelligenz*, 28(3), 159–168.
- Toulmin, S. E. (2003[1958]). *The uses of argument*. Cambridge University Press.
- van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2), 129–155.
- van Benthem, J. (2011). *Logical dynamics of information and interaction*. Cambridge University Press.
- van Benthem, J., van Eijck, J., & Kooi, B. (2006). Logics of communication and change. *Information and Computation*, 204(11), 1620–1662.
- van Benthem, J., & Velázquez-Quesada, F. R. (2010). The dynamics of awareness. *Synthese*, 177(1), 5–27.
- van Ditmarsch, H., & Kooi, B. (2008). Semantic results for ontic and epistemic change. In W. van der Hoek, Bonanno, G., & Wooldridge, M. (Eds.), *Logic and the foundations of game and decision theory (LOFT 7)*, volume 3 of *Texts in Logic and Games* (pp. 9–58). Amsterdam University Press.
- van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic*. Springer.
- van Ditmarsch, H. P., van der Hoek, W., & Kooi, B. P. (2005). Dynamic epistemic logic with assignment. In *AAMAS 2005* (pp. 141–148). ACM.
- Vinokur, A., & Burstein, E. (1974). Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach. *Journal of Personality and Social Psychology*, 29(3), 305.
- Walton, D., & Krabbe, E. C. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. State University of New York Press.
- Walton, D. N. (1984). *Logical dialogue-games and Fallacies*. University Press of America.
- Wang, Y., & Cao, Q. (2013). On axiomatizations of public announcement logic. *Synthese*, 190(1), 103–134.
- Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129–140.
- Wason, P. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151).
- Wenzel, J. W. (1992). Perspectives on argument. In W. L. Benoit, Hamble, D., & Benoit, P. J. (Eds.), *Readings in argumentation* (pp. 121–143). Foris.
- Wu, Y., & Caminada, M. (2010). A labelling-based justification status of arguments. *Studies in Logic*, 3(4), 12–29.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.