Multi-Agent Abstract Argumentation Frameworks
With Incomplete Knowledge of Attacks

Andreas Herzig[1] and **Antonio Yuste Ginel**[2]

[1]IRIT, CNRS
[2]University of Málaga
Andreas.Herzig@irit.fr, antonioyusteginel@gmail.com

Abstract

We introduce a multi-agent, dynamic extension of abstract argumentation frameworks (AFs), strongly inspired by epistemic logic, where agents have only partial information about the conflicts between arguments. These frameworks can be used to model a variety of situations. For instance, those in which agents have bounded logical resources and therefore fail to spot some of the actual attacks, or those where some arguments are not explicitly and fully stated (enthymematic argumentation). Moreover, we include second-order knowledge and common knowledge of the attack relation in our structures (where the latter accounts for the state of the debate), so as to reason about different kinds of persuasion and about strategic features. This version of multi-agent AFs, as well as their updates with public announcements of attacks (more concretely, the effects of these updates on the acceptability of an argument) can be described using S5-PAL, a well-known dynamic-epistemic logic. We also discuss how to extend our proposal to capture arbitrary higher-order attitudes and uncertainty.

1 Introduction

Argumentation has been shown to be a powerful perspective to deal with some problems within the field of artificial intelligence [Bench-Capon and Dunne, 2007], such as reasoning on the basis of conflicting information, autonomous agents interaction, and, more recently, bridging machine and human reasoning [Modgil *et al.*, 2013]. A prominent tool in this context are *abstract argumentation frameworks* (AFs), introduced by [Dung, 1995]. AFs are just a new way to look at directed graphs, where nodes represent arguments and arrows represent some kind of conflict relation among them, usually called the *attack* relation. AFs enables formal reasoning about a fundamental question in argumentation theory, namely, given a set of (possibly) conflicting arguments: which of them are to be accepted by a rational agent?

Context. Although they have shown to be extremely fruitful, AFs are notably abstract in their original formulation and they therefore face many limitations when modelling argumentative scenarios. They in particular do not account for

multiple agents and for dynamics, both of which are essential to real-life argumentation. Many proposals attempted to overcome these two limitations. As to the former, there is an already well-established tradition of applying argumentation techniques to multi-agent systems [Maudet *et al.*, 2006; Carrera and Iglesias, 2015]. Simultaneously, dynamics has received increasing attention during the last decade [Doutre and Mailly, 2018]. Recent studies on strategic argumentation [Thimm, 2014] encompass both aspects. Here, the main research question is how agents (should) select their next movement during a dialogue in order to accomplish their goal. A recurrent idea in this area is the use of opponent models [Rienstra *et al.*, 2013; Thimm, 2014], i.e., letting agents have beliefs about other agents' views of the underlying AF, as a means for computing their strategy.

Contribution. This paper introduces a novel multi-agent and dynamic extension of AFs that is strongly inspired by epistemic logic. There are a number of features that distinguish our approach from previous ones. First, the usual focus when modelling partial information is considering awareness of arguments, rather than attacks, the customary object of knowledge [Schwarzentruber *et al.*, 2012; Thimm, 2014; Rahwan and Larson, 2008; Rienstra *et al.*, 2013]. We follow the less explored perspective, considered for instance in [Dyrkolbotn and Pedersen, 2016], that models an agent's view as a (possibly incomplete) set of attacks. This option fits well in debates where some arguments are not fully stated (enthymemes), or where agents have bounded reasoning resources, and hence fail to spot some of the conflicts among arguments. Second, we include simple opponent-models: second-order views of the attack relation, i.e., what an agent knows that another agent knows, in order to make our structures appealing for the study of strategical issues. Third, we also take into account the notion of *public view*, accounting for the state of a debate. [1]

2 Background

We recall AFs and their stable extensions, as well as their encoding in propositional logic. Our focus on the stable semantics is by no means restrictive and we may replace it by any

---

[1]Results will be stated throughout this paper without proof due to space limitations, but they can be found in Antonio Yuste Ginel's forthcoming PhD dissertation.

other semantics that can be captured in propositional logic [Besnard *et al.*, 2020]. We also briefly recall epistemic logic and public announcement logic.

2.1 Argumentation Frameworks

A *Dung argumentation framework* (AF) is a graph $\mathcal{R} = \langle A, R \rangle$ where A is a finite set (the set of arguments) and $R \subseteq A \times A$ is a relation on A (the attack relation).

A *stable extension of an AF* $\mathcal{R} = \langle A, R \rangle$ is a set of arguments $E \subseteq A$ such that (1) there are no a and b in E such that $(a, b) \in R$ (E is conflict-free), and (2) for every argument $b \in A \setminus E$ there exists $a \in E$ such that $(a, b) \in R$ (any argument outside the extension is attacked by the extension). An argument a is *sceptically accepted* (resp. *credulously accepted*) if it belongs to every (resp. at least one) stable extension.

2.2 AFs in Propositional Logic

AFs can be encoded in propositional logic, as first proposed in [Besnard and Doutre, 2004] and extensively discussed in [Gabbay, 2011; Besnard *et al.*, 2020]. We suppose that the propositional variables contain *attack variables* $r_{a,b}$ capturing that $(a, b) \in R$. Given a set of arguments A, $\text{Att}_A = \{r_{a,b} : (a, b) \in A \times A\}$ is the set of all attack variables. Then the *theory of* $\mathcal{R} = \langle A, R \rangle$ is the Boolean formula

$$\theta(\mathcal{R}) = \left(\bigwedge_{(a,b) \in R} r_{a,b} \right) \wedge \left(\bigwedge_{(a,b) \in (A \times A) \setminus R} \neg r_{a,b} \right).$$

The theory of $\mathcal{R}$ correctly captures the attack relation: $(a, b) \in R$ if and only if $\theta(\mathcal{R}) \models r_{a,b}$. Moreover, it is complete w.r.t. attack variables: for every $r_{a,b} \in \text{Att}_A$, either $\theta(\mathcal{R}) \models r_{a,b}$ or $\theta(\mathcal{R}) \models \neg r_{a,b}$.

Extensions can also be characterised in propositional logic if we add furthermore acceptance variables in_a, to be read "a is accepted" or "a is in". For a fixed set of arguments A, the set of propositional variables of our language is therefore $\text{Prp}_A = \text{Att}_A \cup \{\text{in}_a : a \in A\}$. Valuations (noted v, v', etc.) are subsets of Prp_A, and satisfaction of Boolean formulas φ built from Prp_A is defined in the standard way and noted $v \models \varphi$ as usual. As as shown in [Besnard and Doutre, 2004], the stable semantics of an AF $\mathcal{R} = \langle A, R \rangle$ can then be characterised by the Boolean formula

$$\text{Stable} = \bigwedge_{a \in A} \left(\text{in}_a \leftrightarrow \bigwedge_{b \in A} (r_{b,a} \rightarrow \neg \text{in}_b) \right).$$

The models of that formula correspond to the extensions under the stable semantics: for every valuation $v \subseteq \text{Prp}_A$, $v \models \theta(\mathcal{R}) \wedge \text{Stable}$ if and only if the set $E_v = \{a \in A : \text{in}_a \in v\}$ is a stable extension of $\mathcal{R}$.

2.3 Public Announcement Logic

The language of public announcement logic (PAL) extends that of propositional logic by epistemic and public announcement operators. We consider the S5 version of PAL with common knowledge, in brief, S5-PAL [Ditmarsch *et al.*, 2007, Chapter 4]. Its language is defined by the following grammar:

$$\varphi ::= p \mid \neg \varphi \mid \varphi \wedge \varphi \mid \mathbf{K}_i \varphi \mid \mathbf{C}_{\text{Agt}} \varphi \mid [\varphi!] \varphi$$

where p ranges over a countable set of propositional variables Prp and i ranges over a finite set of agents Agt. In order to avoid tedious exceptions in what follows, we assume that $|\text{Agt}| \geq 2$. The formula $\mathbf{K}_i \varphi$ reads "i knows that φ" and $\mathbf{C}_{\text{Agt}} \varphi$ reads "it is common knowledge among all agents that φ". The latter enables reasoning about the common ground in a conversation [Clark and Schaefer, 1989]. The program $\varphi!$ is the public announcement of φ, where we suppose that announcements are truthful: φ can be announced only if φ is true. The formula $[\psi!] \varphi$ reads "if ψ can be announced then φ is true afterwards". The dual $\langle \psi! \rangle \varphi$ abbreviates $\neg [\psi!] \neg \varphi$ and therefore reads "ψ can be announced and φ is true afterwards". The "everybody knows" operator $\mathbf{E}_{\text{Agt}} \varphi$ abbreviates $\bigwedge_{i \in \text{Agt}} \mathbf{K}_i \varphi$. The dual of $\mathbf{K}_i \varphi$ is noted $\hat{\mathbf{K}}_i \varphi$ and defined as $\neg \mathbf{K}_i \neg \varphi$.

An *S5-model* is a triple $M = \langle W, \{\sim_i\}_{i \in \text{Agt}}, V \rangle$ such that W is a non-empty set, every $\sim_i \subseteq W \times W$ is an equivalence relation, and $V : W \longrightarrow \wp(\text{Prp})$ is a valuation. A *pointed S5-model* is a couple $\langle M, w \rangle$ such that $M = \langle W, \{\sim_i\}_{i \in \text{Agt}}, V \rangle$ is an S5-model and $w \in W$. The truth conditions for the modal operators are:

$M, w \models \mathbf{K}_i \varphi$ if $M, v \models \varphi$ for every v s.th. $(w, v) \in \sim_i$

$M, w \models \mathbf{C}_{\text{Agt}} \varphi$ if $M, v \models \varphi$ for every v s.th.

$$(w, v) \in \left(\bigcup_{i \in \text{Agt}} \sim_i \right)^*$$

$M, w \models [\psi!] \varphi$ if $M, w \models \psi$ implies $M^{\psi!}, w \models \varphi$

where $\left(\bigcup_{i \in \text{Agt}} \sim_i \right)^*$ is the reflexive and transitive closure of the union of the equivalence relations $\sim_i$ and where $M^{\psi!} = \langle W^{\psi!}, \{\sim_i^{\psi!}\}_{i \in \text{Agt}}, V^{\psi!} \rangle$ is the update of M by $\psi!$, defined by: $W^{\psi!} = \{v \in W : M, v \models \psi\}$; $\sim_i^{\psi!} = \sim_i \cap (W^{\psi!} \times W^{\psi!})$; and $V^{\psi!} = V|_{W^{\psi!}}$ (the restriction of V to $W^{\psi!}$).

Axiomatically, every $\mathbf{K}_i$ obeys the axioms $\mathbf{K}_i \top$, $\mathbf{K}_i \varphi \rightarrow \varphi$, $\mathbf{K}_i (\varphi \wedge \psi) \leftrightarrow (\mathbf{K}_i \varphi \wedge \mathbf{K}_i \psi)$, $\mathbf{K}_i \varphi \rightarrow \mathbf{K}_i \mathbf{K}_i \varphi$, and $\neg \mathbf{K}_i \varphi \rightarrow \mathbf{K}_i \neg \mathbf{K}_i \varphi$. The common knowledge operator $\mathbf{C}_{\text{Agt}}$ is also an S5 operator. Moreover, common knowledge implies individual knowledge $\mathbf{C}_{\text{Agt}} \varphi \rightarrow \mathbf{E}_{\text{Agt}} \varphi$ and satisfies the induction axiom $\mathbf{C}_{\text{Agt}} (\mathbf{E}_{\text{Agt}} \varphi \vee \mathbf{E}_{\text{Agt}} \neg \varphi) \rightarrow (\mathbf{C}_{\text{Agt}} \varphi \vee \mathbf{C}_{\text{Agt}} \neg \varphi)$ [Herzig and Perrotin, 2020].

In the sequel we are going to use a fragment of the language of PAL where announcements are restricted to epistemic formulas of the form $\mathbf{K}_i \varphi$.[2] Such public announcements capture, at least partly, public announcements made by one of the agents in Agt. In contrast, in standard PAL it is considered that announcements are made by an agent who is not in Agt and is therefore 'outside the system' [Herzig, 2017].

3 Multi-Agent AFs

Given an AF $\mathcal{R} = \langle A, R \rangle$, the subset of the set of all attacks an agent knows is the agent's *view* of $\mathcal{R}$. For example, suppose agent 1 knows that b and b' both attack a and that c (resp. c') attacks b (resp. b'), and suppose agent 2 only knows

---

[2] We could have introduced the fragment only, but preferred to introduce the full PAL-language in order to follow its standard presentation.
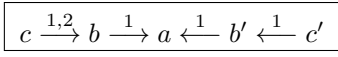
$$c \xrightarrow{1,2} b \xrightarrow{1} a \xleftarrow{1} b' \xleftarrow{1} c'$$

Figure 1: An agent-annotated AF where the attacks (b, a), (b', a) and (c', b') are known by 1 but not by 2

that c attacks b. We depict this situation in Figure 1 by an agent-annotated AF.

Agents argue by informing each other about attacks. We suppose that they do so in a truthful and sincere manner: when 1 announces that b attacks a then b indeed attacks a, and moreover 1 knows that. In Figure 1, both 1 and 2 can announce that c attacks b because they know this; and only 1 can announce that b attacks a while 2 cannot. These inform acts are *public announcements*: when i has put forward a piece of information then it becomes part of the common ground of the dialogue. In epistemic logic terms: the piece of information becomes common knowledge [Clark and Schaefer, 1989].

In the above example we have only considered first-order knowledge: knowledge about the world, namely about the attack relation. We may also consider second-order knowledge: knowledge about other agents' knowledge. Suppose e.g. that 1 knows that 2 knows that c attacks b. Then 1 knows that 2 is able to announce that c attacks b. Therefore 1 knows that her announcement that b attacks a can be countered by 2 and will not contribute to get a publicly rejected.

We now define the agents' first-order views, their second-order views, and the common ground of the conversation.

3.1 First-Order Views

Given a background AF $\mathcal{R} = \langle A, R \rangle$ and an agent $i \in \text{Agt}$, i's view of $\mathcal{R}$ is an AF $\mathcal{R}^i = \langle A, R^i \rangle$ that is a subgraph of $\mathcal{R}$, i.e., such that

$$R^i \subseteq R. \tag{C1}$$

The agent-annotated AF of Figure 1 depicts 1's and 2's views. We have $R^1 = \{(c, b), (b, a), (c', b'), (b', a)\}$ and $R^2 = \{(c, b)\}$.

3.2 Second-Order Views

Second-order views account for what agents know about other agent's knowledge. In strategic argumentation terms, they are one-depth opponent models without uncertainty [Rienstra *et al.*, 2013]. $(\mathcal{R}^i)^j = \langle A, (R^i)^j \rangle$ is agent j's positive view of i's positive view.[3] When (a, b) is in j's view of i's view then j knows that i knows that a attacks b.

Just as for first-order knowledge, we require that j's knowledge about i's knowledge and ignorance is correct, i.e., that $(R^i)^j \subseteq R^i$. Moreover, what j knows about i's knowledge must also be known by j herself, i.e., we should also have $(R^i)^j \subseteq R^j$. Put together, we require that

$$(R^i)^j \subseteq R^i \cap R^j. \tag{C2}$$

---

[3] So we abstract away from, e.g., the attacks of which j knows that i knows they are false.

We furthermore suppose that the agents are introspective, i.e., the inclusion also holds in the other sense when j equals i:

$$(R^i)^i = R^i. \tag{C3}$$

We could go beyond second-order views and define arbitrary higher-order views of AFs. We do not do so here because second-order views suffice to account for strategic dialogues of the above kind where an agent knows she can win a debate thanks to her knowledge of other agents' knowledge. It would also pose some technical difficulties because higher-order views would a priori make the representation of multi-agent AFs infinite. We comment more on this in Section 6.1.

3.3 Public Views

A *public view of* $\mathcal{R}$ is yet another AF $\mathcal{R}^{pub} = \langle A, R^{pub} \rangle$. We require $\mathcal{R}^{pub}$ to be a subgraph of all individual first-order and second-order views:

$$R^{pub} \subseteq \bigcap_{i,j \in \text{Agt}} (R^i)^j. \tag{C4}$$

Then $R^{pub} \subseteq \left(\bigcap_{i \in \text{Agt}} R^i \right) \cap \left(\bigcap_{i,j \in \text{Agt}} (R^i)^j \right)$ follows with condition (C2). Let us compare our public views to the notion of *state of a debate* in [Dyrkolbotn and Pedersen, 2016]. The latter is a pair $\langle A, R^d \rangle$ where $\bigcap_{i \in \text{Agt}} R^i \subseteq R^d \subseteq \bigcup_{i \in \text{Agt}} R^i$. Hence, the fact that a pair (a, b) is known by everyone is a sufficient condition to be part of the state of a debate, while the same fact is a necessary condition for (a, b) to be part of the public view.

3.4 Multi-Agent AFs

Putting all these ingredients together we define *multi-agent AFs* as quintuplets of the form

$$\mathcal{M} = \langle A, R, \{R^i\}_{i \in \text{Agt}}, \{(R^i)^j\}_{i,j \in \text{Agt}}, R^{pub} \rangle$$

such that the above conditions (C1), (C2), (C3), (C4) hold. Every $\langle A, R^i \rangle$ is i's view of $\langle A, R \rangle$, every $\langle A, (R^i)^j \rangle$ is j's view of i's view of $\langle A, R \rangle$, and $\langle A, R^{pub} \rangle$ is a public view of $\langle A, R \rangle$.

Interpretation. As mentioned, our choice to model partial knowledge of an AF by partial knowledge of the attack relation is rather non-standard, the more usual choice being to model incomplete information through partial awareness of arguments [Schwarzentruber *et al.*, 2012; Thimm, 2014; Rahwan and Larson, 2008; Rienstra *et al.*, 2013]. Both alternatives seem to be orthogonal, and our choice is mainly for the sake of exploration. Nevertheless, we believe it is natural in the context of enthymematic arguments where the hypotheses together with background knowledge entail the conclusion: the agents' unawareness of some of the background hypotheses (enthymeme) explains their incomplete knowledge of the attack relation. Besides, even when arguments are explicitly and fully stated by participants in a debate, non-omniscient agents can fail to see part of the attack relation due to their limited logical skills. One may for example consider resource-bounded agents that are good (sound) reasoners equipped with an incomplete set of inference rules, as proposed in [Konolige, 1984].

3.5 Extensions of a Multi-Agent AF

What is accepted in the current state of the debate is determined by the stable extensions of the public view: a *public stable extension* of the multi-agent AF

$$\mathcal{M} = \langle A, R, \{R^i\}_{i \in \mathtt{Agt}}, \{(R^i)^j\}_{i,j \in \mathtt{Agt}}, R^{pub} \rangle$$

is a set of arguments E^{pub} that is a stable extension of $\langle A, R^{pub} \rangle$. Then, an argument is *publicly-sceptically* (resp. *publicly-credulously*) *accepted* iff it is sceptically (resp. credulously) accepted w.r.t. $\langle A, R^{pub} \rangle$. This definition can be connected with the distinction between public persuasion and private persuasion that is made in the introduction of [Dupin de Saint-Cyr *et al.*, 2016]. In this sense, public persuasion amounts to public acceptance (or rejection) while private persuasion is related to second-order views.

3.6 Updating a Multi-Agent AF

The agents' incomplete knowledge of the attack relation evolves when one of them announces an attack known by her. That attack then becomes part of what is public (and consequently also part of the agents' first- and second-order views).

An *announcement* is a triple consisting of an agent and two arguments, noted $i{:}(a,b)$ and read "i announces that a attacks b". Given a multi-agent AF $\mathcal{M} = \langle A, R, \{R^j\}_{j \in \mathtt{Agt}}, \{(R^j)^k\}_{j,k \in \mathtt{Agt}}, R^{pub} \rangle$ and an announcement $i{:}(a,b)$, the *update of* $\mathcal{M}$ by $i{:}(a,b)$ incorporates the attack (a,b) into all views, resulting in a new multi-agent AF $\mathcal{M} \dot{+} i{:}(a,b)$. The update function has to be partial because agents can only make truthful and sincere announcements. It is defined on $\mathcal{M}$ if $(a,b) \in R^i$ and is undefined otherwise. When it is defined then the result of the update is

$$\mathcal{M} \dot{+} i{:}(a,b) = \langle A, R, \{R^j \cup \{(a,b)\}\}_{j \in \mathtt{Agt}},$$
$$\{(R^j)^k \cup \{(a,b)\}\}_{j,k \in \mathtt{Agt}},$$
$$R^{pub} \cup \{(a,b)\} \rangle.$$

It can be easily checked that if $\mathcal{M}$ is a multi-agent AF satisfying the above conditions (C1), (C2), (C3), (C4) then $\mathcal{M} \dot{+} i{:}(a,b)$ is either undefined or is again a multi-agent AF satisfying (C1), (C2), (C3), (C4).

4 Characterising Multi-Agent AFs and their Updates in Epistemic Logic

Just as an AF can be captured in propositional logic, a multi-agent AF can be captured in epistemic logic: the formula $\mathbf{K}_i \mathsf{r}_{a,b}$ expresses that $(a,b) \in R^i$; likewise, $\neg \mathbf{K}_i \mathsf{r}_{a,b}$ expresses that $(a,b) \notin R^i$. We call formulas of the form $\mathbf{K}_i \mathsf{r}_{a,b}$ *first-order attack variables*. Formulas of the form either $\mathbf{K}_i \mathsf{r}_{a,b}$ or $\neg \mathbf{K}_i \mathsf{r}_{a,b}$ are *first-order attack literals*. Similarly, formulas of the form $\mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$ are *second-order attack variables*. Note that we should not encode $(a,b) \notin R^i$ as $\mathbf{K}_i \neg \mathsf{r}_{a,b}$ because in her current view, i may not know that a attacks b, but may learn this later on.

4.1 Characterisation of Multi-Agent AFs

Given an AF $\mathcal{R} = \langle A, R \rangle$, *agent i's theory of $\mathcal{R}$* is the conjunction of first-order attack variables

$$\theta^i(\mathcal{R}) = \left(\bigwedge_{(a,b) \in R} \mathbf{K}_i \mathsf{r}_{a,b} \right) \wedge \left(\bigwedge_{(a,b) \in (A \times A) \setminus R} \neg \mathbf{K}_i \mathsf{r}_{a,b} \right).$$

The agents' theories are complete w.r.t. first-order attack variables: either $\theta^i(\mathcal{R}) \models \mathbf{K}_i \mathsf{r}_{a,b}$ or $\theta^i(\mathcal{R}) \models \neg \mathbf{K}_i \mathsf{r}_{a,b}$. Moreover, $\theta^i(\mathcal{R})$ is an epistemic characterisation of the attack relation: a attacks b if and only if $\theta^i(\mathcal{R}) \models \mathbf{K}_i \mathsf{r}_{a,b}$. The other way round, completeness of $\theta^i(\mathcal{R})$ implies that a does not attack b if and only if $\theta^i(\mathcal{R}) \models \neg \mathbf{K}_i \mathsf{r}_{a,b}$. The formulas $\theta^i(.)$ are useful to capture i's view of the AF in epistemic logic: we will typically apply $\theta^i(.)$ to $\mathcal{R}^i$ and write $\theta^i(\langle A, R^i \rangle)$.

Second-order views account for what agents know about other agents' knowledge and can nicely be captured in epistemic logic by nesting epistemic operators. For example, $\mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$ expresses that $(a,b) \in (R^i)^j$, and $\mathbf{K}_j \mathsf{r}_{a,b} \wedge \neg \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$ expresses that $(a,b) \in R^j$ and $(a,b) \notin (R^i)^j$. Given an AF $\mathcal{R} = \langle A, R \rangle$, *agent j's theory about agent i* is the following conjunction of second-order attack variables:

$$\theta^{i;j}(\mathcal{R}) =$$
$$\left(\bigwedge_{(a,b) \in R} \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b} \right) \wedge \left(\bigwedge_{(a,b) \in (A \times A) \setminus R} \neg \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b} \right).$$

The theory $\theta^{i;j}(\mathcal{R})$ is complete w.r.t. second-order attack variables: either $\theta^{i;j}(\mathcal{R}) \models \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$ or $\theta^{i;j}(\mathcal{R}) \models \neg \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$. It characterises the attack relation: $(a,b) \in R$ if and only if $\theta^{i;j}(\mathcal{R}) \models \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$. We will use second-order knowledge expressed in $\theta^{i;j}(.)$ to reason about j's view of i's view, typically writing $\theta^{i;j}(\langle A, (R^i)^j \rangle)$.

Finally, the *public theory of $\mathcal{R}$* is captured by means of the common knowledge operator:

$$\theta^{pub}(\mathcal{R}) =$$
$$\left(\bigwedge_{(a,b) \in R} \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{a,b} \right) \wedge \left(\bigwedge_{(a,b) \in (A \times A) \setminus R} \neg \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{a,b} \right).$$

Again, the public theory is complete w.r.t. common knowledge of the attack relation and also characterises the attack relation. Again, we typically write $\theta^{pub}(\langle A, R^{pub} \rangle)$.

Let us put all this together: let

$$\mathcal{M} = \langle A, R, \{R^i\}_{i \in \mathtt{Agt}}, \{(R^i)^j\}_{i,j \in \mathtt{Agt}}, R^{pub} \rangle$$

be a multi-agent AF. Then the *epistemic theory of $\mathcal{M}$* is the conjunction of the above theories:

$$\theta(\mathcal{M}) = \theta(\langle A, R \rangle) \wedge \left(\bigwedge_{i \in \mathtt{Agt}} \theta^i(\langle A, R^i \rangle) \right) \wedge$$
$$\left(\bigwedge_{i,j \in \mathtt{Agt}} \theta^{i;j}(\langle A, (R^i)^j \rangle) \right) \wedge \theta^{pub}(\langle A, R^{pub} \rangle).$$

Proposition 1. *For every multi-agent AF* $\mathcal{M} = \langle A, R, \{R^i\}_{i \in \mathtt{Agt}}, \{(R^i)^j\}_{i,j \in \mathtt{Agt}}, R^{pub} \rangle$ *we have:*

- $(a,b) \in R$ *iff* $\theta(\mathcal{M}) \to \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \notin R$ *iff* $\theta(\mathcal{M}) \to \neg \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \in R^i$ *iff* $\theta(\mathcal{M}) \to \mathbf{K}_i \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \notin R^i$ *iff* $\theta(\mathcal{M}) \to \neg \mathbf{K}_i \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \in (R^i)^j$ *iff* $\theta(\mathcal{M}) \to \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \notin (R^i)^j$ *iff* $\theta(\mathcal{M}) \to \neg \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \in R^{pub}$ *iff* $\theta(\mathcal{M}) \to \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{a,b}$ *is S5 valid;*
- $(a,b) \notin R^{pub}$ *iff* $\theta(\mathcal{M}) \to \neg \mathbf{C}_{\mathtt{Agt}} \mathsf{r}_{a,b}$ *is S5 valid.*

4.2 Characterisation of Extensions

We characterise the public stable extensions of a multi-agent AF in epistemic logic as follows:

$$\mathsf{Stable}^{pub} = \bigwedge_{a \in A} \left(\mathsf{in}_a \leftrightarrow \bigwedge_{b \in A} (\mathbf{C}_{\mathtt{Agt}} r_{b,a} \rightarrow \neg \mathsf{in}_b) \right).$$

Our definition requires some explanations: why didn't we insert the common knowledge operator into the definition of stable extensions of Section 2 in a different way? For example, we could have defined "a is in" as "every b that is in is also known to not attack a". However, this is too strong a requirement in debates where the agents learn the attack relation step by step: then no argument a can ever be accepted because the agent may learn later on in the debate about some attack of a that she currently does not know.

Proposition 2. *Let $\mathcal{M}$ be a multi-agent AF. Then:*

- E^{pub} *is a public stable extension of $\mathcal{M}$ iff*

  $$\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \left(\bigwedge_{a \in E^{pub}} \mathsf{in}_a \right) \wedge \left(\bigwedge_{a \in A \setminus E^{pub}} \neg \mathsf{in}_a \right)$$

  is satisfiable in S5.

- *If $R^{pub} = R$, then $\theta(\mathcal{M}) \rightarrow (\mathsf{Stable}^{pub} \leftrightarrow \mathsf{Stable})$ is S5 valid.*

- $a \in A$ *is publicly-sceptically accepted (resp. publicly-credulously accepted) iff $(\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub}) \rightarrow \mathsf{in}_a$ is S5-valid (resp. $\theta(\mathcal{M}) \wedge \mathsf{Stable}^{pub} \wedge \mathsf{in}_a$ is S5-satisfiable).*

4.3 Characterisation of Updates

Let us show how updates can be characterised in S5-PAL.

Proposition 3. *Let $\mathcal{M}$ be a multi-agent AF. Then $(a, b) \in R^i$ iff $\theta(\mathcal{M}) \rightarrow \langle \mathbf{K}_i r_{a,b}! \rangle \top$ is S5-PAL valid.*

Proposition 4. *Let $\mathcal{M}$ be a multi-agent AF such that $(a, b) \in R^i$. Let $\mathcal{M} \dotplus i{:}(a, b)$ be its update by $i{:}(a, b)$. For every $(c, d) \in A \times A$:*

- $(c, d) \in R^j \cup \{(a, b)\}$ *iff*
  $\theta(\mathcal{M}) \rightarrow [\mathbf{K}_i r_{a,b}!] \mathbf{K}_j r_{c,d}$ *is S5-PAL valid;*

- $(c, d) \in (R^j)^k \cup \{(a, b)\}$ *iff*
  $\theta(\mathcal{M}) \rightarrow [\mathbf{K}_i r_{a,b}!] \mathbf{K}_k \mathbf{K}_j r_{c,d}$ *is S5-PAL valid;*

- $(c, d) \in R^{pub} \cup \{(a, b)\}$ *iff*
  $\theta(\mathcal{M}) \rightarrow [\mathbf{K}_i r_{a,b}!] \mathbf{C}_{\mathtt{Agt}} r_{c,d}$ *is S5-PAL valid.*

5 An Example

Let us take up the agent-annotated AF of Figure 1, and suppose the following informal reading for the involved arguments, representing a debate on public health policies to stop the spreading of COVID-19.

a: "everybody should wear, at least, a surgical mask."
b: "there are not enough surgical masks."
c: "home-made masks can replace surgical masks."
b': "surgical masks don't prevent people from getting infected."
c': "surgical masks prevent people from infecting others."

Let $\mathcal{M}$ be the multi-agent AF where the first-order knowledge is the one depicted in Figure 1 and where moreover 1 has some second-order knowledge about 2's knowledge: we suppose that 1 knows that 2 knows that c attacks b. Suppose that the public view is initially empty and that 1's goal is that a is publicly-sceptically rejected (meaning that a must be attacked by a publicly-sceptically accepted argument). Observe that this goal is dishonest, since according to 1's view a should be accepted.

Agent 1 can start the debate in two different ways: either by announcing that b attacks a, or by announcing that b' attacks a. We will formally show below that 1 knows that the first announcement will not contribute to the achievement of her goal. Therefore 1 should better announce that b' attacks a: even if 1 is not sure that she will attain her goal, this is at least possible for her. In terms of the planning literature, the announcement $1{:}(b', a)$ is therefore a strong plan to achieve 1's goal.

Alternatively, $\mathcal{M}$ can be updated by first $1{:}(b, a)$ and then $2{:}(c, b)$, resulting in $(\mathcal{M} \dotplus 1{:}(b, a)) \dotplus 2{:}(c, b)$, for which the public view is $(R^{pub} \cup \{(b, a)\}) \cup \{(c, b)\} = \{(b, a), (c, b)\}$.

In dynamic epistemic logic terms, we have that

$$\theta(\mathcal{M}) \rightarrow \langle \mathbf{K}_1 r_{b,a}! \rangle \langle \mathbf{K}_2 r_{c,b}! \rangle \Big(\mathbf{C}_{\mathtt{Agt}} r_{b,a} \wedge \mathbf{C}_{\mathtt{Agt}} r_{c,b} \wedge$$
$$\bigwedge_{(x,y) \notin \{(b,a),(c,b)\}} \neg \mathbf{C}_{\mathtt{Agt}} r_{x,y} \Big)$$

is S5-PAL valid. Moreover, 1 knows this thanks to her second-order knowledge:

$$\theta(\mathcal{M}) \rightarrow \mathbf{K}_1 \langle \mathbf{K}_1 r_{b,a}! \rangle \langle \mathbf{K}_2 r_{c,b}! \rangle \Big(\mathbf{C}_{\mathtt{Agt}} r_{b,a} \wedge \mathbf{C}_{\mathtt{Agt}} r_{c,b} \wedge$$
$$\bigwedge_{(x,y) \notin \{(b,a),(c,b)\}} \neg \mathbf{C}_{\mathtt{Agt}} r_{x,y} \Big)$$

is also S5-PAL valid. Therefore

$$\theta(\mathcal{M}) \rightarrow \mathbf{K}_1 \langle \mathbf{K}_1 r_{b,a}! \rangle \langle \mathbf{K}_2 r_{c,b}! \rangle \big(\mathsf{Stable}^{pub} \rightarrow (\mathsf{in}_a \wedge \neg \mathsf{in}_b) \big)$$

is S5-PAL valid.

6 Discussion

Let us now discuss how multi-agent AFs can be extended in two different directions, so as to include, respectively, third- and higher-order epistemic information and uncertainty.

6.1 Higher-Order Views

We have only considered the agents' first- and second-order views of an AF. As we have mentioned in Section 3.2, one may wish to go beyond this and use the full power of epistemic logic where epistemic operators in formulas can be arbitrarily nested. Let us sketch here how this can be achieved by using S5-models directly in order to represent multi-agent AFs with arbitrary higher-order views.

Define the *characteristic valuation* of an AF $\mathcal{R} = \langle A, R \rangle$ as $v_{\mathcal{R}} = \{r_{a,b} : (a, b) \in R\}$. A *pointed S5-model for $\mathcal{R}$* is a pointed S5-model $\langle M, w \rangle$ such that (i) $\langle M, w \rangle \models \theta(\mathcal{R})$ and (ii) for every $u \in W$ and every $a \in A$ either $\mathsf{in}_a \in V(u)$ or $\mathsf{in}_a \notin V(u)$ (that is, the valuation of in_a-variables

is uniform throughout W). We can rewrite condition (i) more semantically as $V(w) \cap \text{Att}_A = v_{\mathcal{R}}$. Now, all the information contained in a multi-agent AF can be extracted from $\langle M, w \rangle$. More in detail, agent i's view of $\mathcal{R}$ can be defined as

$$R^i = \{(a,b) \in A \times A \ : \ M, w \models \mathbf{K}_i \mathsf{r}_{a,b}\}.$$

Again, this can be expressed without making any reference to the epistemic language as: $R^i = (\bigcap_{u \in \sim_i[w]} V(u)) \cap \text{Att}_A$, where $\sim_i[w] = \{u \in W \ : \ w \sim_i u\}$. Analogous formulations can be made for $(R^i)^j$ and R^{pub}. Condition (C1) is guaranteed by reflexivity of $\sim_i$. Second-order views are then defined as

$$(R^i)^j = \{(a,b) \in A \times A \ : \ M, w \models \mathbf{K}_j \mathbf{K}_i \mathsf{r}_{a,b}\}.$$

Condition (C2) is guaranteed by reflexivity of $\sim_j$ and $\sim_i$. Condition (C3) is guaranteed by reflexivity and transitivity of $\sim_i$. The public view is defined as

$$R^{pub} = \{(a,b) \in A \times A \ : \ M, w \models \mathbf{C}_{\text{Agt}} \mathsf{r}_{a,b}\}.$$

Condition (C4) is guaranteed by the definition of common knowledge. Agents' n-order views of $\mathcal{R}$ could be defined in general as

$$(((R^{i_0})^{\cdots})^{i_n}) = \{(a,b) \in A \times A \ :$$
$$M, w \models \mathbf{K}_{i_n} \mathbf{K}_{i_{n-1}} \ldots \mathbf{K}_{i_0} \mathsf{r}_{a,b}\}$$

where $((R^{i_0})^{\cdots})^{i_n}$ represents what agent i_n knows that agent i_{n-1} knows... that agent i_0 knows about R. The updated multi-agent AF $\mathcal{M} \dot{+} i{:}(a,b)$ would be then the one induced by $M^{\mathbf{K}_i \mathsf{r}_{a,b}!}$. Analogous results to those of Proposition 2 can be formulated and proved in this setting.

6.2 Uncertainty

AFs assume perfect knowledge about the attack relation and the existence of arguments. Recently, some works have relaxed this assumption, *incomplete argumentation frameworks* being one of the most popular approaches [Baumeister *et al.*, 2018]. Our multi-agent AFs are strongly linked to a well-known subclass of incomplete AFs, namely, attack-incomplete AFs, previously studied under the denomination *partial AFs* [Cayrol *et al.*, 2007]. An (attack-)*incomplete AF* (IAF) is a tuple $\langle A, R, R_? \rangle$ such that $R, R_? \subseteq A \times A$ and $R \cap R_? = \emptyset$. Intuitively, R captures the view of a single agent while $R_?$ captures the set of attacks that the agent sees as uncertain (she does not know whether they hold). The fundamental notion when reasoning about IAFs is that of a *completion*: an AF $\langle A, R^* \rangle$ such that $R \subseteq R^* \subseteq R \cup R_?$. Semantic notions (extensions and argument acceptability) are then combined in different ways with that of completion (see [Baumeister *et al.*, 2018; Fazzinga *et al.*, 2020]). There are some obvious differences between our approach and IAFs. First, IAFs lack a multi-agent component. Second, IAFs fit better a doxastic interpretation (rather than an epistemic one), since there is no direct way to express that an attack actually holds, and thus to distinguish between true and false beliefs. Third, our multi-agent AFs lack any mechanism for modelling uncertainty.

Combining both approaches seems a promising line of research. By mixing multi-agency with uncertainty we obtain, besides the already discussed attack relations, four new different kinds: $R_?^i$ the set of attacks that i is uncertain about; $(R_?^i)^j$ the attacks that j knows that i is uncertain about; $(R^i)_?^j$ the attacks that j does not know whether i knows; and $(R_?^i)_?^j$ the attacks that j does not know whether i knows whether. Again, epistemic logic can help us to throw some light on these notions, by providing a list of intuitive constraints that they must satisfy. Let us illustrate this idea. The fact that $(a,b) \in R_?^i$ is naturally encoded in epistemic logic as $\hat{\mathbf{K}}_i \mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_i \neg \mathsf{r}_{a,b}$. Hence $(a,b) \in R_?^i \cup R^i$ is encoded as $(\hat{\mathbf{K}}_i \mathsf{r}_{a,b} \wedge \hat{\mathbf{K}}_i \neg \mathsf{r}_{a,b}) \vee \mathbf{K}_i \mathsf{r}_{a,b}$ which is S5-equivalent to $\hat{\mathbf{K}}_i \mathsf{r}_{a,b}$. Following a similar reasoning, we can deduce that $(a,b) \in (R^j)_?^i \cup (R^j)^i$ is encoded as $\hat{\mathbf{K}}_i \mathbf{K}_j \mathsf{r}_{a,b}$. Since $\hat{\mathbf{K}}_i \mathbf{K}_j \mathsf{r}_{a,b} \to \hat{\mathbf{K}}_i \mathsf{r}_{a,b}$ is S5-valid, it seems reasonable to require that $(R^j)_?^i \cup (R^j)^i \subseteq R^i \cup R_?^i$.

7 Related Work and Future Goals

Interactions among formal argumentation and epistemic logic are receiving increasing attention. These interactions can be classified into two groups: (i) the study of the notion of *justified* or *argument-based belief* [Shi *et al.*, 2018; Grossi and van der Hoek, 2014] and (ii) epistemic reasoning about AFs [Schwarzentruber *et al.*, 2012; Sakama and Son, 2019; Proietti and Yuste-Ginel, 2021], in which the current work is framed. The main differences with [Sakama and Son, 2019] is that we use a simpler—and thus less expressive—logical encoding and that we do take into account several aspects—such as multi-agent scenarios and an explicit modelling of the dynamics of argumentation—that were left aside in [Sakama and Son, 2019]. As for [Schwarzentruber *et al.*, 2012; Proietti and Yuste-Ginel, 2021], we change the object of knowledge (from awareness of arguments to attacks), and we show that, at least to some extent, there is no need for incorporating the full Kripke semantics of epistemic logic in order to capture some of the strategic features of argumentation.

We have studied the relation between multi-agent AFs and multiagent S5, a standard logic of knowledge. The question of whether the current approach can be extended to belief in a relatively simple manner remains open. Such an extension requires the integration of a mechanism for belief revision in order to model argumentative/doxastic dynamics. There are other paths for future research. Besides developing in detail the ideas sketched in Section 6, we intend to define dialogues based on our multi-agent AFs and apply our current results to several reasoning tasks as, for instance, computing the next movement within a dialogue.

Acknowledgments

References

[Baumeister *et al.*, 2018] Dorothea Baumeister, Daniel Neugebauer, Jörg Rothe, and Hilmar Schadrack. Verification in incomplete argumentation frameworks. *Artificial Intelligence*, 264:1–26, 2018.

[Bench-Capon and Dunne, 2007] Trevor J.M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artificial intelligence*, 171(10-15):619–641, 2007.

[Besnard and Doutre, 2004] Philippe Besnard and Sylvie Doutre. Checking the acceptability of a set of arguments. In *Proceedings of NMR'2004*, pages 59–64, 2004.

[Besnard *et al.*, 2020] Philippe Besnard, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. Logical theories and abstract argumentation: A survey of existing works. *Argument & Computation*, 11(1-2):41–102, 2020.

[Carrera and Iglesias, 2015] Álvaro Carrera and Carlos A. Iglesias. A systematic review of argumentation techniques for multi-agent systems research. *Artificial Intelligence Review*, 44(4):509–535, 2015.

[Cayrol *et al.*, 2007] Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Handling ignorance in argumentation: Semantics of partial argumentation frameworks. In Khaled Mellouli, editor, *ECSQARU*, pages 259–270. Springer, 2007.

[Clark and Schaefer, 1989] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13(2):259–294, 1989.

[Ditmarsch *et al.*, 2007] Hans P. van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer, 2007.

[Doutre and Mailly, 2018] Sylvie Doutre and Jean-Guy Mailly. Constraints and changes: a survey of abstract argumentation dynamics. *Argument and Computation*, 9:223–248, 2018.

[Dung, 1995] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–357, 1995.

[Dupin de Saint-Cyr *et al.*, 2016] Florence Dupin de Saint-Cyr, Pierre Bisquert, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. Argumentation update in YALLA (yet another logic language for argumentation). *Int. J. Approx. Reasoning*, 75:57–92, 2016.

[Dyrkolbotn and Pedersen, 2016] Sjur K. Dyrkolbotn and Truls Pedersen. Arguably argumentative: A formal approach to the argumentative theory of reason. In Vincent C. Müller, editor, *Fundamental Issues of Artificial Intelligence*, pages 317–339. Springer, 2016.

[Fazzinga *et al.*, 2020] Bettina Fazzinga, Sergio Flesca, and Filippo Furfaro. Revisiting the notion of extension over incomplete abstract argumentation frameworks. In *Proceedings of IJCAI-20*, pages 1712–1718. IJCAIO, 7 2020.

[Gabbay, 2011] Dov M. Gabbay. Dung's argumentation is essentially equivalent to classical propositional logic with the Peirce–Quine dagger. *Logica Universalis*, 5(2):255–318, 2011.

[Grossi and van der Hoek, 2014] Davide Grossi and Wiebe van der Hoek. Justified beliefs by justified arguments. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Proceedings of KR*. AAAI Press, 2014.

[Herzig and Perrotin, 2020] Andreas Herzig and Elise Perrotin. On the axiomatisation of common knowledge. In Nicola Olivetti, Rineke Verbrugge, Sara Negri, and Gabriel Sandu, editors, *13th Conference on Advances in Modal Logic, AiML 2020, Helsinki, Finland, August 24-28, 2020*, pages 309–328. College Publications, 2020.

[Herzig, 2017] Andreas Herzig. Dynamic epistemic logics: promises, problems, shortcomings, and perspectives. *J. Appl. Non Class. Logics*, 27(3-4):328–341, 2017.

[Konolige, 1984] Kurt Konolige. *A Deduction Model of Belief and its Logics*. PhD thesis, Computer Science Department, Stanford University, 1984.

[Maudet *et al.*, 2006] Nicolas Maudet, Simon Parsons, and Iyad Rahwan. Argumentation in multi-agent systems: Context and recent developments. In *Proceedings of ArgMAS*, pages 1–16. Springer, 2006.

[Modgil *et al.*, 2013] Sanjay Modgil, Francesca Toni, Floris Bex, Ivan Bratko, Carlos I. Chesñevar, Wolfgang Dvořák, Marcelo A. Falappa, Xiuyi Fan, Sarah Alice Gaggl, Alejandro J. García, María P. González, Thomas F. Gordon, João Leite, Martin Možina, Chris Reed, Guillermo R. Simari, Stefan Szeider, Paolo Torroni, and Stefan Woltran. *The Added Value of Argumentation*, pages 357–403. Springer Netherlands, Dordrecht, 2013.

[Proietti and Yuste-Ginel, 2021] Carlo Proietti and Antonio Yuste-Ginel. Dynamic epistemic logics for abstract argumentation. *Synthese*, 2021. 10.1007/s11229-021-03178-5.

[Rahwan and Larson, 2008] Iyad Rahwan and Kate Larson. Mechanism design for abstract argumentation. In *Proceedings of AAMAS'08*, pages 1031–1038. IFAAMAS, 2008.

[Rienstra *et al.*, 2013] Tjitze Rienstra, Matthias Thimm, and Nir Oren. Opponent models with uncertainty for strategic argumentation. In *Proceedings of IJCAI*. AAAI Press, 2013.

[Sakama and Son, 2019] Chiaki Sakama and Tran Cao Son. Epistemic argumentation framework. In *Proceedings of PRICAI 2019*, pages 718–732. Springer, 2019.

[Schwarzentruber *et al.*, 2012] François Schwarzentruber, Srdjan Vesic, and Tjitze Rienstra. Building an epistemic logic for argumentation. In *JELIA*, pages 359–371. Springer, 2012.

[Shi *et al.*, 2018] Chenwei Shi, Sonja Smets, and Fernando R. Velázquez-Quesada. Beliefs supported by binary arguments. *Journal of Applied Non-Classical Logics*, 28(2-3):165–188, 2018.

[Thimm, 2014] Matthias Thimm. Strategic argumentation in multi-agent systems. *KI-Künstliche Intelligenz*, 28(3):159–168, 2014.