On the Epistemic Logic of Incomplete Argumentation Frameworks
Andreas Herzig, Antonio Yuste-Ginel

To cite this version:

HAL Id: hal-03450084
https://ut3-toulouseinp.hal.science/hal-03450084
Submitted on 25 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
On the Epistemic Logic of Incomplete Argumentation Frameworks

Andreas Herzig\textsuperscript{1}, Antonio Yuste-Ginel\textsuperscript{2}
\textsuperscript{1}IRIT, CNRS, Toulouse
\textsuperscript{2}Departamento de Filosofía, Universidad de Málaga (UMA)
herzig@irit.fr, antonioyusteginel@gmail.com

Abstract

We study the relation between two existing formalisms: incomplete argumentation frameworks (IAFs) and epistemic logic of visibility (ELV). We show that the set of completions of a given IAF naturally corresponds to a specific equivalence class of possible worlds within the model of visibility. This connection is further strengthened in two directions. First, we show how to reduce argument acceptance problems of IAFs to ELV model-checking problems. Second, we highlight the epistemic assumptions that underlie IAFs by providing a minimal epistemic logic for IAFs.

1 Introduction

Argumentation is a well-established research area within the field of Artificial Intelligence. It has been proved useful for a variety of purposes (Bench-Capon and Dunne 2007; Atkinson et al. 2017), including (but not limited to): providing an operational semantics for diverse non-monotonic logics, establishing a bridge between human and machine reasoning, and improving artificial agents’ interaction. At the abstract level, that is, without analysing the structure and nature of arguments, Dung’s argumentation frameworks (AFs) are by far the most studied formal construct (Dung 1995). Conceptually, AFs respond to one fundamental question of (general) argumentation theory: given a set of possibly conflicting arguments, which subsets of arguments are to be accepted by a rational agent?

AFs are, however, too abstract for some purposes. Consequently, they have been extended in many different directions, including the handling of preferences (Amgoud and Vesic 2011), and more refined forms of interaction between arguments (Cayrol and Lagasquie-Schiex 2005). One of the main assumptions that AFs come equipped with is perfect knowledge about the structure of the modelled discussion (about the structure of the argumentation framework). The formal argumentation community has recently made efforts to relax this assumption. This has been done using quantitative methods, mainly probabilistic (Li, Oren, and Norman 2011), and qualitative ones. Among the second type, incomplete argumentation frameworks (IAFs) are prominent (Coste-Marquis et al. 2007; Baumeister, Neugebauer, and Rothe 2018; Baumeister et al. 2018b; Fazzinga, Flesca, and Furfaro 2020). An IAF partitions both the set of arguments $A$ and the set of attacks $R$ of an AF into two sets, distinguishing arguments (attacks) known with certainty from those perceived as uncertain by the agent.

In this paper, we pursue a simple research question: what is the epistemic logic underlying IAFs? Our answer, presented in Section 3, consists in establishing a strong connection between IAFs and the logic of visibility (ELV) (Herzig, Lorini, and Maffre 2018). We further exploit this connection in two different senses. First, we show how it can be used to reduce acceptance reasoning tasks of IAFs to ELV model-checking problems (Section 4). Second, we provide a minimal epistemic logic for IAFs that makes the epistemic assumptions underlying them explicit (Section 5). For space reasons, results are stated without proof; they can be found in Antonio Yuste-Ginel’s forthcoming PhD dissertation.

2 Preliminaries

2.1 Abstract Argumentation Frameworks

An abstract argumentation framework (AF) is a directed graph, i.e., a pair $(A, R)$ where $A$ is a finite set whose elements stand for arguments and $R \subseteq A \times A$ represents some kind of conflict-based relation among arguments, typically, an attack relation. AFs were first studied by (Dung 1995), where the author proposed to capture the informal notion of argument acceptability through different argumentation semantics. For the sake of brevity, we limit ourselves to stable semantics, but our analysis applies to all semantics that can be captured in propositional logic (see (Besnard, Cayrol, and Lagasque-Schiex 2020)). A stable extension of an AF $(A, R)$ is a set of arguments $E \subseteq A$ s.th.: (i) there are no $x, y \in E$ s.th. $(x, y) \in R$ (‘$E$ is conflict-free’); and (ii) for every $x \in A \setminus E$, there is a $y \in E$ such that $(y, x) \in R$ (‘$E$ attacks every argument outside itself’).

2.2 Incomplete AFs

An IAF is a tuple $\text{IAF} = (A, A^\ast, R, R^\ast)$ s.th. $A \cap A^\ast = \emptyset$; $R, R^\ast \subseteq (A \cup A^\ast) \times (A \cup A^\ast)$; and $R \cap R^\ast = \emptyset$. $(A, R)$ is the definite part of IAF while $(A^\ast, R^\ast)$ is the uncertain part of IAF. We assume that $A \cup A^\ast$ is finite. A completion of IAF is any pair $(A^\ast, R^\ast)$ s.th.:

- $A \subseteq A^\ast \subseteq A \cup A^\ast$; and
- $R\mid_{A^\ast} \subseteq R^\ast \subseteq (R \cup R^\ast)\mid_{A^\ast}$.
where we note \( R_Y = R \cap (Y \times Y) \) the restriction of a relation \( R \subseteq X \times X \) to \( Y \subseteq X \).

All the classic reasoning problems regarding AFs were generalized to IAFs and their complexity was characterised (Baumeister et al. 2018b; Baumeister et al. 2021). We here focus on argument acceptance problems. Classic acceptance problems for standard AFs can be generalized to IAFs as follows. The set of all stable extensions of \((A, R)\) is noted \( \text{st}(A, R) \).

### 2.3 Epistemic Logic of Visibility

Epistemic logic of visibility or observability was born as a lightweight alternative to standard epistemic logic (van der Hoek, Troquard, and Wooldridge 2011; Herzig, Lorini, and Maffre 2018; Cooper et al. 2021). We consider its single-agent version and adapt it to our purposes. Let \( \Sigma \) be a finite set of arguments, called the signature. We define the set of atoms for signature \( \Sigma \) in two steps:

\[
\text{Prop}^\Sigma = \{ aw_x \mid x \in \Sigma \} \cup \{ t_{x,y} \mid (x, y) \in \Sigma \times \Sigma \} \cup \{ in_x \mid x \in \Sigma \};
\]

\[
\text{ATM}^\Sigma = \text{Prop}^\Sigma \cup \{ Sp \mid p \in \text{Prop}^\Sigma \}.
\]

Intuitively, \( aw_x \) means that the agent is aware of argument \( x \); \( t_{x,y} \) means that \( x \) attacks \( y \); \( in_x \) means that argument \( x \) is in the extension; and \( Sp \) means that the agent sees the value of \( p \) (equivalently, she knows whether \( p \)). We use the term awareness in a rather loose sense here. Among different possible interpretations, it can be seen as the arguments the agent’s opponent is aware of. This explains why an agent can be aware of an argument without knowing whether she is aware of it. Moreover, note that \( \Sigma \) does not need to coincide with \( A \cup A^2 \); there can be arguments whose awareness the agent knows not to hold.

The epistemic language for signature \( \Sigma \) is generated by the following BNF:

\[
\varphi ::= \alpha \mid \neg \varphi \mid (\varphi \land \varphi) \mid K \varphi \quad (\alpha \in \text{ATM}^\Sigma)
\]

where \( K \varphi \) reads “the agent knows that \( \varphi \)”\(^1\). We also consider the dual \( \neg K \neg \varphi \), read “the agent considers possible that...”.

The **single-agent visibility model** for \( \text{ATM}^\Sigma \) is the pair \( M_{vis} = (2^\text{ATM}^\Sigma, \sim) \), where the accessibility relation \( \sim \subseteq 2^\text{ATM}^\Sigma \times 2^\text{ATM}^\Sigma \) is such that \( v \sim v' \) iff for every \( p \in \text{Prop}^\Sigma \):

1. \( Sp = v \) then \( (p \in v \iff p \in v') \);
2. \( Sp \in v \iff Sp \in v' \).

Formulas are interpreted at states of \( M_{vis} \), i.e., at propositional valuations. We only state the clause for the epistemic operator:

\[
M_{vis}, v \models K \varphi \iff v \sim v' \text{ implies } M_{vis}, v' \models \varphi.
\]

Note that the second constraint in the definition of the visibility model guarantees that \( K \) is a fully introspective operator \((K \varphi \rightarrow KK \varphi) \land (\neg K \varphi \rightarrow K \neg K \varphi) \) are both valid in the visibility model). Moreover, \( Kp \leftrightarrow (Sp \land p) \) is valid too.

### 3 Incomplete AFs and Visibility Models

Relations between IAFs and possible worlds semantics were spotted in the literature (Baumeister et al. 2018; Baumeister, Neugebauer, and Rothe 2018; Baumeister et al. 2021), and just recently explored in more depth by reducing IAFs to general epistemic logic (Proietti and Yuste-Ginel 2021). Nevertheless, a basic research question about the relation between these two formalisms remains still open: what is the epistemic logic underlying IAFs? We provide an answer by connecting IAFs to the single-agent logic of visibility. Formally, we show a one-to-one correspondence between completions of a given IAF and a specific equivalence class of the visibility model. For the sake of simplicity, we dispense with acceptance variables in this section, that is, we assume that \( \text{Prop}^\Sigma = \{ aw_x \mid x \in \Sigma \} \cup \{ t_{x,y} \mid (x, y) \in \Sigma \times \Sigma \} \).

Let \( \text{IAF} = (A, A^2, R, R^2) \) be an IAF with \( A \cup A^2 \subseteq \Sigma \). We associate to IAF its single-agent propositional valuation \( v_{\text{IAF}} = \{ Saw_x \mid x \in \Sigma \setminus A^2 \} \cup \{ aw_x \mid x \in A \} \cup \{ S_{t_{x,y}} \mid (x, y) \in (\Sigma \times \Sigma) \setminus R^2 \} \cup \{ S_{in_x} \mid x \in \Sigma \} \).

The other way round, for every propositional valuation \( v \subseteq \text{ATM}^\Sigma \) we define its induced AF \((A_v, R_v)\) by:

\[
A_v = \{ x \in \Sigma \mid aw_x \in v \};
\]

\[
R_v = \{ (x, y) \in A_v \times A_v \mid t_{x,y} \in v \}.
\]

### 3.1 From IAFs to the Visibility Model

Using the notions we have just defined, we can provide the first clear connection between IAFs and the visibility model.

**Proposition 1.** Let \( \text{IAF} = (A, A^2, R, R^2) \) be an IAF s.th. \( A \cup A^2 \subseteq \Sigma \), let \( v_{\text{IAF}} \) be its propositional valuation, and let \( M_{vis} = (2^\text{ATM}^\Sigma, \sim) \) be the single-agent visibility model for the set of atoms \( \text{ATM}^\Sigma \), then:

- For each completion \((A^*, R^*)\) of IAF there is a \( u \in \sim [v_{\text{IAF}}] \) s.th. \((A^*, R^*) = (A_u, R_u)\).

- For each \( u \in \sim [v_{\text{IAF}}] \) there is a completion \((A^*, R^*)\) of IAF s.th. \((A^*, R^*) = (A_u, R_u)\).

Hence completions can be understood as possible worlds of the visibility model (i.e., as propositional valuations over \( \text{ATM}^\Sigma \)) that are indistinguishable for the formalized agent.
3.2 From the Visibility Model to IAFs

The visibility model for a signature Σ contains all IAFs that can be built from Σ. Formally, we say that \((A, A', R, R')\) is built over Σ whenever \(A \cup A' \subseteq \Sigma\).

**Proposition 2.** Let \(Σ\) be a signature, let \(M_{vis} = (2^{\text{ATM}_\Sigma}, \sim)\) be the single-agent visibility model for \(\text{ATM}_\Sigma\). For every IAF \(\text{IAF} = (A, A', R, R')\) built over \(Σ\) there is an \(u \in 2^{\text{ATM}^\Sigma}\) s.th. \(u \models \text{IAF}\).

In the other direction things get more subtle. Given a valuation \(v \in \text{ATM}_\Sigma\), we define the IAF represented by \(v\) as \(\text{IAF}(v) = (A(v), A'(v), R(v), R'(v))\), where:

\[A(v) = \{ x \in Σ | v \models x \} \]
\[A'(v) = \{ x \in Σ | v \models \neg x \} \]
\[R(v) = \{ (x, y) \in Σ \times Σ | v \models x \land y \} \]
\[R'(v) = \{ (x, y) \in Σ \times Σ | v \models x \land \neg y \} \]

Informally, \(\text{IAF}(v)\) is the IAF where certain arguments are those whose awareness the agent is not aware of; uncertain arguments are those whose awareness the agent is not sure about; and analogously for attacks (conditionally on the awareness of the involved arguments). It is easy to check that, for each valuation \(v\), (i) \(\text{IAF}(v)\) is indeed an IAF; and (ii) the set of all completions of \(\text{IAF}(v)\) equals the set of IAFs \((A_u, R_u)\) such that \(u \in v\).

Moreover, given IAF, it holds that
\[\text{IAF}(v) = \text{IAF}(\text{IAF}(v)).\]

In words, we can go from IAFs to states of the visibility model and back without losing any information. However, this is not true if we start from the set of all valuations: it does not hold that for any \(u\), \(v \models \text{IAF}(u) = u\).

The reason is that there are some valuations containing “defective information”: there are attacks that are true in all indistinguishable valuations, and hence known, even if the agent knows that she is not aware of one of the involved arguments. These pieces of defective information are cleaned out by applying the function \(\text{IAF}\). Alternatively we could impose that valuations should satisfy the property \(\text{Kr}_{x,y} \rightarrow (\text{Kaw}_x \land \text{Kaw}_y)\).

4 Checking Acceptance via Model-Checking

As an application of the correspondence between IAFs and the logic of visibility, we can reduce acceptance problems regarding the former to model-checking problems of the latter. This can be done modulo some propositional encoding of standard argument semantics. We illustrate this idea for the case of stable semantics. Let us first recall how stable extensions of (awareness-relativised) AFs can be captured in propositional logic. We make use of the shorthand \(r_{x,y} = r_{x,y} \land \text{aw}_x \land \text{aw}_y\), and define the formula

\[\text{Stable} = \bigwedge_{x \in Σ} \left( (\text{in}_x \rightarrow \text{aw}_x) \land (\text{aw}_x \rightarrow (\text{in}_x \leftrightarrow \neg \bigvee_{y \in Σ} (\text{in}_y \land r_{y,x}))) \right).\]

As shown in (Doutre, Mafric, and McBurney 2017, Proposition 1), we have that \(v \models \text{Stable iff } E_v = \{ x \in Σ | \text{in}_x \in v \}\) is a stable extension of \((A_v, R_v)\). Using this encoding and the correspondence between completions and accessible states in the visibility model of Proposition 1 we obtain:

**Proposition 3.** Let \(\text{IAF} = (A, A', R, R')\) be an IAF built over \(Σ\), let \(v \models \text{IAF}\) be its propositional valuation, let \(u \in A\) and let \(M_{vis} = (2^{\text{ATM}_\Sigma}, \sim)\) be the single-agent visibility model for the set of atoms \(\text{ATM}_\Sigma\), then:

- The answer to the st-NSA (stable-Necessary-Sceptical-Acceptance) problem with input IAF and \(a \in Σ\) is yes iff \(M_{vis}, v \models \text{IAF} \models \text{K} (\text{Stable} \rightarrow \text{in}_a)\).
- The answer to the st-PCA (stable-Possible-Credulous-Acceptance) problem with input IAF and \(a \in Σ\) is yes iff \(M_{vis}, v \models \text{IAF} \models \text{K} (\text{Stable} \land \text{in}_a)\).

Curiously enough, capturing st-PSA and st-NCA in the logic of visibility requires augmenting the language, jumping from the single-agent version to the two-agent version. The underlying reason is that the definitions of st-PCA and st-NSA use a \(\exists \forall \) (resp. \(\forall \exists \)) pattern for quantifying over completions and extensions, the definitions of the other two reasoning tasks alternate quantifiers. Our solution is to modularize the formalized agent: the agent consists of a part for reasoning about completions (named 1 or the epistemic part) and another one for reasoning about extensions (named 2 or the argumentative part).

First of all, we generalize the set of logical atoms so as to include a finite, non-empty set of agents \(\text{Agt}\):

\[\text{ATM}_\text{Agt} = \text{Prop}^\Sigma \cup \{ S_i p | i \in \text{Agt}, p \in \text{Prop}^\Sigma \},\]

where \(\text{Prop}^\Sigma\) is the set of propositional variables as defined in Section 2.3.

Note that we do not need to parametrize awareness variables for each agent. As mentioned, we are going to modularize the agent regarding his epistemic and argumentative reasoning, but not her awareness. In the remaining of this section, we shall restrict our attention to \(\text{Agt} = \{ 1, 2 \}\). The multi-agent epistemic language extends the epistemic language used before by including a different epistemic operator \(\text{K}_{i,j}\) for each agent in \(\text{Agt}\). Nevertheless, since agent 2 represents the “argumentative part” of the formalized agent, we use \(\square_2\) instead of \(\text{K}_2\), for conceptual clarity. We also use \(\diamond_2\) as a shorthand for \(\neg \square_2 \neg \).

The multi-agent visibility model is the pair \(M_{vis} = (2^{\text{ATM}_\text{Agt}}, \{ \sim_i \}_{i \in \text{Agt}})\) where each \(\sim_i \subseteq 2^{\text{ATM}_\text{Agt}} \times 2^{\text{ATM}_\text{Agt}}\) is such that \(v \sim_i v'\) iff for every \(p \in \text{Prop}^\Sigma\),

1. if \(S_i p \in v\) then \((p \in v')\);
2. \(S_i^p \in v\) iff \(S_i p \in v'\) for every \(j \in \text{Agt}\).

We adapt the notion of propositional valuation associated to an IAF to the two-agent language as follows. Let IAF \(= (A, A', R, R')\) be an IAF s.th. \(A \subseteq A' \subseteq \Sigma\). Define

\[v_{IAF} = \{ S_{aw} x | x \in Σ \setminus A' \} \cup \{ S_{r_{x,y}} (x, y) | (x, y) \in (Σ \times Σ) \setminus R' \} \cup \{ S_{aw} x | x \in Σ \} \cup \{ S_{aw} x | x \in Σ \} \cup \{ S_{aw} x | x \in A \} \cup \{ r_{x,y} | (x, y) \in R' \}.\]
This extends the associated valuation of the single-agent visibility language so as to include visibility atoms indexed with agent 2 for all awareness and attack variables. Conceptually, 2 is a part of the agent that always reasons under the assumption that the underlying AF is completed.

**Proposition 4.** Let \( \text{IAF} = (A, A', R, R') \) be an IAF built over \( \Sigma \), let \( \nu_{\text{IAF}} \) be its propositional valuation, let \( M_{\text{vis}} \) be the visibility model for \( \text{ATM}_{\text{vis}}^{\Sigma_{1,2}} \), and let \( a \in A \), then:

1. The answer to the st-nca (stable-Necessary-Credulous-Acceptance) problem with input IAF and \( a \in A \) is yes iff \( M_{\text{vis}}, \nu_{\text{IAF}} \models K_1 \square 2 (\text{Stable} \land \in a) \).
2. The answer to the st-pca (stable-Possible-Sceptical-Acceptance) problem with input IAF and \( a \in A \) is yes iff \( M_{\text{vis}}, \nu_{\text{IAF}} \models K_1 \square 2 (\text{Stable} \rightarrow \in a) \).

5 A Minimal Epistemic Logic for IAFs

As it is well-known, modal operators interpreted over equivalence relations (as the one we have just used in the logic of visibility) validate S5 axioms, which formally correspond to a factive, fully introspective kind of knowledge; see e.g. (Fagin et al. 2004, Chapter 2). However, one may want to use epistemic logic to reason about AFs without embracing such strong principles. Interestingly, the connection we have established in Section 3 can be used to unravel a minimal epistemic logic for IAFs. By doing so, we clarify the main epistemic assumptions underlying IAFs.

We start defining the logic semantically. Let \( \Sigma \) be an argument signature, \( \text{Prop}^\Sigma \) its associated set of propositions, and \( \text{agt} \) a finite, non-empty set of agents. A **Kripke model** is a tuple \((W, R, V)\) where \( W \neq \emptyset \) is a set of possible worlds, \( R : \text{agt} \rightarrow \varphi(W \times W) \) assigns to each agent an epistemic accessibility relation \( R_i \), and \( V : \text{Prop}^\Sigma \rightarrow \varphi(W) \) is a valuation function.

Since the current discussion is orthogonal to acceptance problems, we dispense again with acceptance variables and assume \( \text{Prop}^\Sigma = \{aw_x, (x, y) \in \Sigma \times \Sigma\} \cup \{x, y\} \in \Sigma \times \Sigma\}. Given a Kripke model \((W, R, V)\) and a world \( u \in W \), we define the \text{AF associated to} \( u \) as the tuple \((A_u, R_u)\), where \( A_u = \{x \in \Sigma \mid u \in V(aw_x)\} \) and \( R_u = \{(x, y) \in \Sigma \times \Sigma \mid u \in V(x, y) \} \cap (A_u \times A_u)\}. Given the epistemic multi-agent language, the concepts of truth \((M, w \models \varphi)\), validity in a model \((M \models \varphi)\), and validity \((\models \varphi)\) are the usual ones in epistemic modal logic, see e.g. (Fagin et al. 2004, Chapter 2). Moreover, an **IAF-Friendly Kripke model** is a Kripke model where for every \( w \in W \), every \( x, y \in \Sigma \), and every \( i \in \text{agt} \):

(AWAR) if \( R_{\text{vis}}[w] \cap V(x, y) \neq \emptyset \), then \( R_{\text{vis}}[w] \cap V(aw_x) \neq \emptyset \) and \( R_{\text{vis}}[w] \cap V(aw_y) \neq \emptyset \); and

(COMP) there is an incomplete AF, IAF, built over \( \Sigma \) s.t.h.: the set of completions of IAF is equal to \( \{(A_v, R_v) \mid v \in R_{\text{vis}}[w]\}\). Informally, an IAF-friendly Kripke model is just a Kripke model where (i) if an agent considers possible that \( x \) attacks \( y \) then she considers possible being aware of both \( x \) and \( y \); and (ii) the set of \( i \)-successors of each world represents the set of completions of an incomplete AF. Now, consider the following two schemas:

3We recall that axiom \( D \) is the modal schema axiomatizing seriality of each relation \( R_i \), and conceptually corresponding to consistency of each agents’ belief. It is needed to axiomatise the logic of IAF-friendly Kripke models because, by definition, every IAF has a non-empty set of completions.

**Acknowledgements**

Andreas Herzig is partially supported by the EU ICT-48 2020 project TAILOR (No. 952215). Antonio Yuste-Ginel gratefully acknowledges funding received from the PhD grant No. MECDFPU 2016/04113.
References


