Burden of Persuasion: A Meta-argumentation Approach

Giuseppe Pisano
Alma AI – Alma Mater Research Institute for Human-Centered Artificial Intelligence, Alma Mater Studiorum—Università di Bologna, Italy
g.pisano@unibo.it

Roberta Calegari
Dipartimento di Informatica – Scienza e Ingegneria (DISI), Alma Mater Studiorum—Università di Bologna, Italy
roberta.calegari@unibo.it

Andrea Omicini
Dipartimento di Informatica – Scienza e Ingegneria (DISI), Alma Mater Studiorum—Università di Bologna, Italy
andrea.omicini@unibo.it

Giovanni Sartor
Alma AI – Alma Mater Research Institute for Human-Centered Artificial Intelligence, Alma Mater Studiorum—Università di Bologna, Italy
giovanni.sartor@unibo.it

Abstract

This work defines a burden of persuasion meta-argumentation model interpreting burden as a set of meta-arguments. Bimodal graphs are exploited to define a *meta level* (dealing with the burden) and an *object level* (dealing with standard arguments). A novel technological reification of the model supporting the burden inversion mechanism is presented and discussed.

Keywords: burdens of persuasion; argumentation; meta-argumentation

The work has been supported by the "CompuLaw" project, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 833647).

1 Introduction

In this work we discuss the model of the burden of persuasion in structured argumentation [5, 6] under a meta-argumentative approach, which leads to *(i)* a clear separation of concerns in the model, *(ii)* a simpler and more efficient implementation of the corresponding argumentation tool, *(iii)* a natural model extension for reasoning over the burden of persuasion concepts.

The work is grounded on the approaches to meta-argumentation that emphasise the inner nature of arguments and dialogues as inherently meta-logical [10, 11]. Our approach relies on those works [10, 11] that introduce only the required abstraction at the meta level. The proposed meta-argumentation framework for the burden of persuasion includes three ingredients: *(i) object-level argumentation* – to create arguments from defeasible and strict rules –, *(ii) meta-level argumentation* – to create arguments dealing with abstractions related to the burden concept using argument schemes (or meta-level rules) –, and *(iii) bimodal graphs* to define the interaction between the object level and the meta level—following the account in [10].

This work extends our previous work [13] in two main directions. First, it introduces and discusses a novel technological reification of the model supporting the burden inversion mechanism. Then, related work is discussed by positioning our contribution against the state of the art, and highlighting strengths and limitations w.r.t. other approaches—e.g., [16].

Accordingly, Section 2 introduces basic elements of the meta-argumentation framework. Section 3 formally defines the framework for the burden of persuasion introducing related argument schemes and discusses its equivalence with the model presented in [5]. Section 4 discusses a real case study in the law domain dealing with the problem of burden inversion. Finally, Section 5 presents the technological reification of the model. Related work is discussed in Section 6, whereas conclusions are drawn in Section 7.

2 Meta-argumentation framework

In this section, we introduce the meta-argumentation framework. For the sake of simplicity we choose to model our meta-argumentation framework by exploiting bimodal graphs, which are often exploited to both define meta-level concepts and understand the interactions of object-level and meta-level arguments [11, 10]. Accordingly, Subsection 2.1 presents the object-level argumentation language exploited by our model, leveraging on an ASPIC$^+$-like argumentation framework [15]. Then, Subsection 2.2 introduces bimodal argumentation graphs' main definitions. Finally,

the meta-level argumentation language based on the use of argument schemes [18] is introduced in Subsection 2.3.

2.1 Structured argumentation for object-level argumentation

Let a literal be an atomic proposition or its negation.

Notation 1. *For any literal ϕ, its complement is denoted by $\bar{\phi}$. That is, if ϕ is a proposition p, then $\bar{\phi} = \neg p$, whereas if ϕ is $\neg p$, then $\bar{\phi}$ is p.*

Let us also identify burdens of persuasion, i.e., those literals whose proof requires a convincing argument. We assume that such literals are consistent (it cannot be the case that there is a burden of persuasion on both ϕ and $\bar{\phi}$).

Definition 2.1 (Burdens of persuasion)**.** *Burdens of persuasion are represented by predicates of the form $bp(\phi)$, stating the burden is allocated on the literal ϕ.*

Literals are put in relation with *bp* predicates through defeasible rules.

Definition 2.2 (Defeasible rule)**.** *A **defeasible rule** r has the form:*

$$\rho : \quad \phi_1, ..., \phi_n, \sim\phi'_1, ..., \sim\phi'_m \Rightarrow \psi$$

with $0 \leq n, m$, and where

- *ρ is the unique identifier for r, denoted by $\mathsf{N}(r)$;*
- *each $\phi_1, \ldots, \phi_n, \phi'_1, \ldots, \phi'_m, \psi$ is a literal or a* bp *predicate;*
- *$\phi_1, \ldots \phi_n, \sim\phi'_1, ..., \sim\phi'_m$ are denoted by $Antecedent(r)$;*
- *ψ is denoted by $Consequent(r)$;*
- *$\sim\phi$ denotes the weak negation (negation by failure) of ϕ—i.e., ϕ is an exception that would block the application of the rule whose antecedent includes $\sim\phi$.*

The unique identifier of a rule can be used as a literal to specify that the named rule is applicable, and its negation to specify that the rule is inapplicable, dually [9].

A superiority relation $\succ$ is defined over rules: $s \succ r$ states that rule s prevails over rule r.

Definition 2.3 (Superiority relation)**.** *A **superiority relation** $\succ$ over a set of rules Rules is a transitive, antireflexive, and antisymmetric binary relation over Rules.*

A defeasible theory consists of a set of rules and a superiority relation over the rules.

Definition 2.4 (Defeasible theory)**.** *A* ***defeasible theory*** *is a tuple* $\langle Rules, \succ \rangle$ *where Rules is a set of rules, and* $\succ$ *is a superiority relation over Rules.*

Given a defeasible theory, we can construct arguments by chaining rules from the theory [9, 7, 17].

Definition 2.5 (Argument)**.** *An* ***argument*** A *constructed from a defeasible theory* $\langle Rules, \succ \rangle$ *is a finite construct of the form:* $A : A_1, \ldots A_n \Rightarrow_r \phi$ *with* $0 \leq n$*, where*

- A *is the argument's unique identifier;*
- $A_1, \ldots, A_n$ *are arguments constructed from the defeasible theory* $\langle Rules, \succ \rangle$*;*
- ϕ *is the* conclusion *of the argument, denoted by* $\mathsf{Conc}(A)$*;*
- $r : \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \phi$ *is the top rule of* A*, denoted by* $\mathsf{TopRule}(A)$*.*

Notation 2. *Given an argument* $A : A_1, \ldots A_n \Rightarrow_r \phi$ *as in Definition 2.5,* $\mathsf{Sub}(A)$ *denotes the set of subarguments of* A*, i.e.,* $\mathsf{Sub}(A) = \mathsf{Sub}(A_1) \cup \ldots \cup \mathsf{Sub}(A_n) \cup \{A\}$*.* $\mathsf{DirectSub}(A)$ *denotes the direct subarguments of* A*, i.e.,* $\mathsf{DirectSub}(A) = \{A_1, \ldots, A_n\}$*.*

Preferences over arguments are defined via a last-link ordering: argument A is preferred over argument B if the top rule of A is stronger than the top rule of B.

Definition 2.6 (Preference relation)**.** *A* ***preference relation*** $\succ$ *is a binary relation over a set of arguments* $\mathcal{A}$*: argument* A *is preferred to argument* B *– denoted by* $A \succ B$ *– iff* $\mathsf{TopRule}(A) \succ \mathsf{TopRule}(B)$*.*

Arguments are put in relation with each other according to the attack relation.

Definition 2.7 (Attack)**.** *Argument* A ***attacks*** *argument* B *iff* A *undercuts or rebuts* B*, where*

- A *undercuts* B *(on B') iff* $\mathsf{Conc}(A) = \neg \mathsf{N}(\rho)$ *for some* $B' \in \mathsf{Sub}(B)$*, where* ρ *is* $\mathsf{TopRule}(B')$
- A *rebuts* B *(on B') iff either* (i) $\mathsf{Conc}(A) = \bar{\phi}$ *for some* $B' \in \mathsf{Sub}(B)$ *of the form* $B'_1, ..., B'_M \Rightarrow \phi$ *and* $B' \not\succ A$*, or* (ii) $\mathsf{Conc}(A) = \phi$ *for some* $B' \in \mathsf{Sub}(B)$ *such that* $\sim\phi \in Antecedent(\mathsf{TopRule}(B'))$

In short, arguments can be attacked either on a conclusion of a defeasible inference (*rebutting* attack) or on a defeasible inference step itself (*undercutting* attack).

Definition 2.8 (Argumentation graph)**.** *An* ***argumentation graph*** *is a tuple* $\langle \mathcal{A}, \rightsquigarrow \rangle$*, where* $\mathcal{A}$ *is the set of all arguments, and* $\rightsquigarrow$ *is* attack *relation over* $\mathcal{A}$*.*

Notation 3. *Given an argumentation graph* $G = \langle \mathcal{A}, \rightsquigarrow \rangle$*, we write* $\mathcal{A}_G$ *and* $\rightsquigarrow_G$ *to denote the graph's arguments and attacks, respectively.*

Now, let us introduce the notion of the $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-labelling of an argumentation graph, where each argument in the graph is labelled IN, OUT, or UND, depending on whether it is accepted, rejected, or undecided, respectively.

Definition 2.9 (Labelling)**.** *Let* G *be an argumentation graph. An* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-***labelling*** L *of* G *is a total function* $\mathcal{A}_G \to \{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$*.* $\mathcal{L}(\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}, G)$ *denotes the set of all* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-***labellings*** *of* G*.*

A labelling-based semantics prescribes a set of labellings for any argumentation graph according to some criterion embedded in its definition.

Definition 2.10 (Labelling-based semantic)**.** *Let* G *be an argumentation graph. A labelling-based semantics* S *associates with* G *a subset of* $\mathcal{L}(\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}, G)$*, denoted as* $L_S(G)$*.*

2.2 Object and meta level connection: bimodal graphs

In this section we recall the main definitions of bimodal graphs as the model of interaction between object and meta level. Bimodal graphs make it possible to capture scenarios where arguments are categorised in multiple levels—two in our case, the object and the meta level. Accordingly, a bimodal graph is composed of two components: an argumentation graph for the meta level and an argumentation graph for the object level, along with a relation of support that originates from the meta level and targets attacks and arguments on the object level. Every object-level argument and every object-level attack is supported by at least one meta-level argument. Meta-level arguments can only attack meta-level arguments, and object-level arguments can only attack object-level arguments.

Definition 2.11 (Bimodal argumentation graph)**.** *A* ***bimodal argumentation graph*** *is a tuple* $\langle \mathcal{A}_O, \mathcal{A}_M, \mathcal{R}_O, \mathcal{R}_M, \mathcal{S}_A, \mathcal{S}_R \rangle$ *where*

1. $\mathcal{A}_O$ *is the set of object-level arguments*
2. $\mathcal{A}_M$ *is the set of meta-level arguments*
3. $\mathcal{R}_O \subseteq \mathcal{A}_O \times \mathcal{A}_O$ *represents the set of object-level attacks*

4. $\mathcal{R}_M \subseteq \mathcal{A}_M \times \mathcal{A}_M$ *represents the set of meta-level attacks*
5. $\mathcal{S}_A \subseteq \mathcal{A}_M \times \mathcal{A}_O$ *represents the set of supports from meta-level arguments into object-level arguments*
6. $\mathcal{S}_R \subseteq \mathcal{A}_M \times \mathcal{R}_O$ *represents the set of supports from meta-level arguments into object-level attacks*
7. $\mathcal{A}_O \cap \mathcal{A}_M = \emptyset$
8. $\forall A \in \mathcal{A}_O \; \exists \; B \in \mathcal{A}_M \; : \; (B, A) \in \mathcal{S}_A$
9. $\forall R \in \mathcal{R}_O \; \exists \; B \in A_M \; : \; (B, R) \in \mathcal{S}_R$

The object-level argument graph is represented by the couple $(\mathcal{A}_O, \mathcal{R}_O)$, while the meta-level argument graph is represented by the couple $(\mathcal{A}_M, \mathcal{R}_M)$. The two distinct components are connected by the support relations represented by $\mathcal{S}_A$ and $\mathcal{S}_R$. These supports are the only structural interaction between the meta and the object levels. Condition (8) above ensures that every object-level argument is supported by at least one meta-level argument, whereas condition (9) ensures that every object-level attack is supported by at least one meta-level argument.

Perspectives of the object-level graph can be defined as:

Definition 2.12 (Perspective)**.** *Let $G = \langle \mathcal{A}_O, \mathcal{A}_M, \mathcal{R}_O, \mathcal{R}_M, \mathcal{S}_A, \mathcal{S}_R \rangle$ be a bimodal argumentation graph and let L_S be a labelling semantics. A tuple $\langle \mathcal{A}'_O, \mathcal{R}'_O \rangle$ is an L_S-perspective of G if $\exists \; l \in L_S(\langle \mathcal{A}_M, \mathcal{R}_M \rangle)$ such that*

- $\mathcal{A}'_O = \{ \; A | \exists B \in \mathcal{A}_M \;\; s.t. \; l(B) = \mathsf{IN}, (B, A) \in \mathcal{S}_A \}$
- $\mathcal{R}'_O = \{ \; R | \exists B \in \mathcal{A}_M \;\; s.t. \; l(B) = \mathsf{IN}, (B, R) \in \mathcal{S}_R \}$

Consequently, an object argument may occur in one perspective and not in another according to the results yielded by the meta-level argumentation graph. Under this setting, the role of conditions (8) and (9) becomes clear: every element in a lower level must be relevant w.r.t. the meta-level argumentation process—i.e. we can not have arguments that in no case can be part of a perspective.

2.3 Argument schemes for meta-level argumentation

A fundamental aspect to consider when dealing with a multi-level argumentation graph is how the higher-level graphs can be built starting from the object-level ones. For the purpose, in this work – following the example in [11] – we leverage on argument schemes [18]. In short, argumentation schemes are commonly-used

patterns of reasoning. They can be formalised in a rule-like form [14] where every argument scheme consists of a set of conditions and a conclusion. If the conditions are met, then the conclusion holds. Each scheme comes with a set of *critical questions* (CQ), identifying possible exceptions to the admissibility of arguments derived from the schemes.

Definition 2.13 (Meta-predicate)**.** *A meta-predicate P_M is a symbol that represents a property or a relation between object-level arguments. Let be $\mathcal{M}$ the set of all P_M.*

Definition 2.14 (Object-relation meta-predicate)**.** *An object-relation meta-predicate O_M is a predicate stating the existence of a relation at the object level—e.g., attacks, preferences, and conclusions. Let be $\mathcal{O}$ the set of all O_M.*

Moving from the above definitions we can define an argument scheme as:

Definition 2.15 (Argument Scheme)**.** *An **argument scheme** s has the form:*

$$s : P_1, ..., P_n, \sim P'_1, ..., \sim P'_m \Rightarrow Q$$

with $0 \leq n, m$, and where

- *each $P_1, \ldots, P_n, P'_1, \ldots, P'_m \in \mathcal{M} \cup \mathcal{O}$, while $Q \in \mathcal{M}$*
- *$\sim P$ denotes weak negation (negation by failure) of P—i.e., P is an exception that would block the application of the rule whose antecedent includes $\sim P$*
- *we denote with CQ_s the set of critical questions associated to scheme s.*

Using argument schemes we can build meta-arguments.

Definition 2.16 (Meta-Argument)**.** *A **meta-argument** A constructed from a set of argument schemes S and an object-level argumentation graph G is a finite construct of the form: $A : A_1, \ldots A_n \Rightarrow_s P$ with $0 \leq n$, where*

- *A is the argument's unique identifier;*
- *$s \in S$ is the scheme used to build the argument;*
- *$A_1, \ldots, A_n$ are arguments constructed from S and G;*
- *P is the* conclusion *of the argument, denoted by $\mathsf{Conc}(A)$.*

$CQ(A)$ denotes the critical questions associated to scheme s. The same notation introduced for standard arguments in Notation 2 also applies to meta-arguments.

We can now define attacks over meta-arguments, or, *meta-attacks*.

Definition 2.17 (Meta-Attack)**.** *An argument A* ***attacks*** *argument B (on B') iff either* (i) $\mathsf{Conc}(A) = \bar{P}$ *for some* $B' \in \mathsf{Sub}(B)$ *of the form* $B''_1, ..., B''_M \Rightarrow P$*, or* (ii) $\mathsf{Conc}(A) = P$ *for some* $B' \in \mathsf{Sub}(B)$ *such that* $\sim P \in Antecedent(\mathsf{TopRule}(B'))$*.*

The same definition of *argumentation graph* and *labellings* introduced for standard argumentation in Definitions 2.8, 2.9, 2.10 also holds for meta-arguments and for the meta level.

3 Burden of persuasion as meta-argumentation

Informally, we can say that when we talk about the notion of the burden of persuasion concerning an argument, we intuitively argue over that argument according to a meta-argumentative approach.

Let us consider, for instance, an argument A: if we allocate the burden over it, we implicitly impose the duty to prove its admissibility on A. Thus, moving the analysis up to the meta level of the argumentation process is like having two arguments, let them be F_{BP} and S_{BP}, reflecting the burden of persuasion status. According to this perspective, F_{BP} states that "the burden is not satisfied if A *fails* to prove its admissibility" – i.e. A should be rejected or undefined – and, of course, F_{BP} is not compatible with A being accepted. Alongside, S_{BP} states that "A is acceptable since it *satisfies* its burden". F_{BP} and S_{BP} have a contrasting conclusion and thus they attack each other.

Analysing the burden from this perspective makes immediately clear that the notions that the meta model should deal with are:

N1 the notion of the burden itself expressing the possibility for an argument to be allocated with a burden of persuasion (i.e., *burdened argument*)

N2 the possibility that this burden is satisfied (that is, a *burden met*) or not satisfied

N3 the possibility of making *attacks* involving burdened arguments ineffective.

The outline of that multi-part evaluation scheme for burdens of persuasion in argumentation is now visible and can be formally designed. In the following, we formally define these concepts by exploiting bimodal argument graphs as techniques for expressing the two main levels of the model – meta and object level – and the relationships between the two.

In particular, we are going to define each set of the bimodal argument graph tuple $\langle \mathcal{A}_O, \mathcal{A}_M, \mathcal{R}_O, \mathcal{R}_M, \mathcal{S}_A, \mathcal{S}_R \rangle$. With respect to $\mathcal{A}_O$ and $\mathcal{R}_O$, representing respectively

the set of object-level arguments and attacks, they are built accordingly to the argumentation framework discussed in Subsection 2.1. Hence, our analysis focuses on the meta-level graph $\langle \mathcal{A}_M, \mathcal{R}_M \rangle$ and on the support sets connecting the two levels ($\mathcal{S}_A$ and $\mathcal{S}_R$).

3.1 Meta-level graph

We now proceed to detail all the argumentation schemes used to build arguments in the meta-level graph. Every scheme comes along with its critical questions. As we will see in the next sections, all the critical questions have to be interpreted as kind of "presumptions": they are believed to be true during the construction and evaluation of the argumentation framework – i.e., they are not used as possible attack dimensions –, but their post hoc verification invalidates the entire solution.

Let us first introduce the basic argumentation scheme enabling the definition and representation of an argument with an allocation of the burden of persuasion (i.e., reifying **N1**). We say that an object-level argument A has the burden of persuasion on it if exists an object-level argument B such that $\mathsf{Conc}(B) = bp(\mathsf{Conc}(A))$. This notion is modeled through the following argument scheme:

$$conclusion(A, \phi), conclusion(B, bp(\phi)) \Rightarrow burdened(A) \tag{S0}$$

$$\textit{Is argument } B \textit{ provable?} \tag{$\mathsf{CQ_{S0}}$}$$

where $bp(\phi)$ is a predicate stating ϕ is a literal with the allocation of the burden, $conclusion(A, \phi)$ is a structural meta-predicate stating that $\mathsf{Conc}(A) = \phi$ holds, and $burdened(A)$ is a meta-predicate representing the allocation of the burden on A. Clearly, an argument produced using this scheme only holds if both the arguments A and B on which the inference is based hold—critical question $\mathsf{CQ_{S0}}$.

Analogously, we introduce the scheme S1 representing the absence of such an allocation:

$$conclusion(A, \phi) \Rightarrow \neg burdened(A) \tag{S1}$$

$$\textit{Is argument } A \textit{ provable? Are arguments concluding } bp(\phi) \textit{ not provable?} \tag{$\mathsf{CQ_{S1}}$}$$

Then, as informally introduced at the beginning of this section, we have two schemes reflecting the possibility for a burdened argument to meet or not the burden (**N2**).

$$burdened(A) \Rightarrow bp\_met(A) \tag{S2}$$

$$burdened(A) \Rightarrow \neg bp\_met(A) \tag{S3}$$

Is argument A provable? ($\mathsf{CQ_{S2}}$)

Is argument A always refuted or undecidable? ($\mathsf{CQ_{S3}}$)

where *bp_met* is the meta-predicate stating the burden has been met. It is important to notice that the two schemes above reach opposite conclusions from the same grounds—i.e., the presence of the burden on argument A. The discriminating elements are the critical questions they are accompanied by. In the case of S2, we have that only if a burden of persuasion on argument A exists, and A is acceptable ($\mathsf{CQ_{S3}}$), then the burden is satisfied. On the other side, the validity of S3 is bound to the missing admissibility of argument A. We will see in Section 3.3 how the meta-arguments and the associated questions concur to determine the model results.

Let us now consider attacks between arguments and their relation with the burden of persuasion allocation. When a burdened argument fails to meet the burden, the only thing affecting the argument's acceptability is the burden itself—i.e., attacks from other arguments do not influence the status of the burdened argument, which only depends on its inability to satisfy the burden. The same applies to attacks issued by an argument that fails to meet the burden: the failure implies argument rejection and, as a direct consequence, the inability to effectively attack other arguments. In order to capture the nuance of discerning between effective and ineffective object-level attacks w.r.t. the concept of burden of persuasion (**N3**), we define the following scheme:

$$attack(B, A), \sim(\neg bp\_met(A)), \sim(\neg bp\_met(B)) \Rightarrow effectiveAttack(B, A) \quad (\mathsf{S4})$$

Can we prove arguments A or B do not fail to meet their burden? ($\mathsf{CQ_{S4}}$)

where *attack* is a structural meta-predicate stating an attack relation at the object level, whereas *effectiveAttack* is a meta-predicate expressing that an attack should be taken into consideration according to the burden of persuasion allocation. In other words, if an object-level attack involves burdened arguments, and one of these fails to satisfy the burden, then the attack is considered not effective w.r.t. the allocation of the burden.

The aforementioned schemes can be used to create a meta-level graph containing all the information about constraints related to the burden of persuasion concept thus leading to a clear separation of concerns, as shown in the following example.

Example 1 (Base)**.** *Let us consider two object-level arguments A and B, concluding the literals a and bp(a) respectively. Using the schemes in Subsection 3.1 we can build the following meta-level arguments:*

- A_{S0} *representing the allocation of the burden on argument* A.
- A_{S1} *and* B_{S1} *standing for the absence of a burden on arguments* A *and* B *respectively. The scheme used to build those arguments exploits weak negation in order to cover those scenarios where an argument concluding a* bp *literal exists at the object-level, but it is found not acceptable.*
- A_{S2} *and* A_{S3} *sustaining that* (i) A *was capable of meeting the burden on it,* (ii) A *was not capable of meeting its burden.*

The meta-level graph (Figure 1) points out the relations actually implicit in the notion of the burden of persuasion over an argument, where, intuitively, we argue over the consequences of A's possibly succeeding/failing to meet the burden. At the meta level, all the possible scenarios can be explored by applying different semantics over the meta-level graph.

Considering for instance Dung's preferred semantics [1], we can obtain two distinct outcomes: (1) the burden is not satisfied, i.e., argument A_{S3} *is accepted, and consequently,* A_{S2} *is rejected, or (2) we succeed in proving* A_{S2}*, i.e., the burden is met and* A_{S3} *is rejected (*A_{S0}*,* A_{S1} *are accepted and rejected accordingly). Although the example is really simple – only basic schemes for reasoning on the burden are considered at the meta-level – it clearly demonstrates the possibility of reasoning over the burdens, since, i.e., it establishes whether or not there is a burden on a literal* ϕ *– argument* B *in the example – and enables the evaluation of the consequences of a burdened argument to meet or not its burden.*

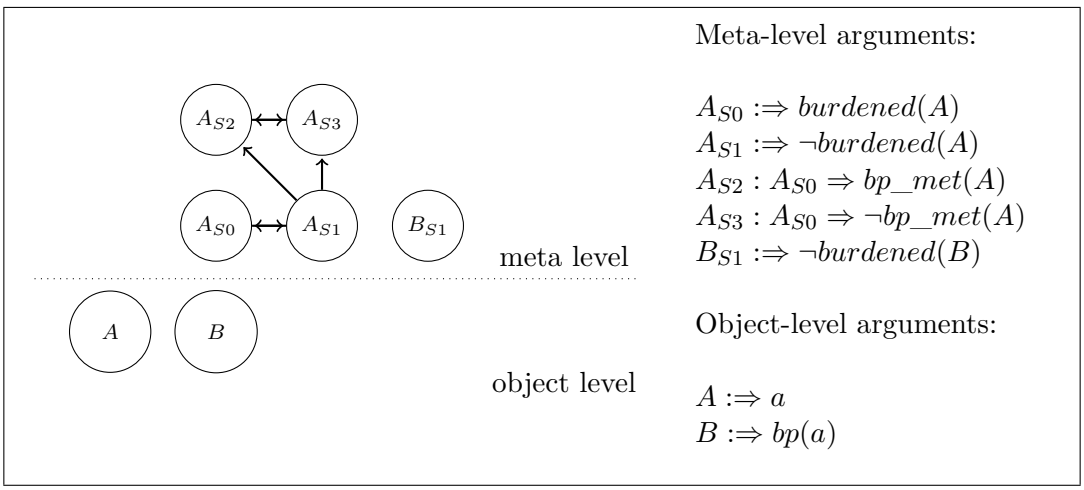

Figure 1: Object and meta level graphs from Example 1

3.2 Object- and meta-level connection: supporting sets

Let us now define how the meta level and the object level interact. Indeed, it is not enough to reason on the consequences of the burden of persuasion allocation only concerning the burdened argument, but the results of the argument satisfying or not such a burden constraint should affect the entire object-level graph. According to the standard bimodal graph theory, defining how the object level and the meta level interact is the role of the argument support relation $\mathcal{S}_A$ and of the attack support relation $\mathcal{S}_R$, respectively. According to Definition 2.11 (Subsection 2.2), every element at level n is connected to an argument at level $n+1$ by a support edge in $\mathcal{S}_A$ or $\mathcal{S}_R$, depending on whether it is either an argument or an attack.

Let us define the support set $\mathcal{S}_A$ of meta arguments supporting object-level arguments as:

$$\mathcal{S}_A = \{(Arg_1, Arg_2) \mid Arg_1 \in \mathcal{A}_M, Arg_2 \in \mathcal{A}_O, \\ (\mathsf{Conc}(Arg_1) = bp\_met(Arg_2) \vee \mathsf{Conc}(Arg_1) = \neg burdened(Arg_2))\}$$

Intuitively, an argument A at the object level is supported by arguments at the meta level claiming that either the burden on A is satisfied (S2) or there is no burden allocated on it (S1).

The set $\mathcal{S}_R$ of meta arguments supporting object-level attacks is defined as:

$$\mathcal{S}_R = \{(Arg_1, (B, A)) \mid Arg_1 \in \mathcal{A}_M, (B, A) \in \mathcal{R}_O, \\ \mathsf{Conc}(Arg_1) = effectiveAttack(B, A)\}$$

In other words, an object-level attack is supported by arguments at the meta level claiming its effectiveness w.r.t. the burden of persuasion allocation (S4).

3.3 Equivalence with burden of persuasion semantics

The defined meta-framework can be used to achieve the same results as the original burden of persuasion labelling semantics [5].

Let us first introduce the notion of *CQ-consistency* for a bimodal argumentation graph G.

Definition 3.1 (CQ-consistency)**.** *Let $G = \langle \mathcal{A}_O, \mathcal{A}_M, \mathcal{R}_O, \mathcal{R}_M, \mathcal{S}_A, \mathcal{S}_R \rangle$ be a bimodal argumentation graph, and let $L_S(G)$ be a labelling-based semantics. P is the set of corresponding L_S-perspectives. A perspective $p \in P$ is CQ-consistent if every* IN *argument A in the corresponding meta-level labelling satisfies its critical questions ($CQ(A)$).*

Before proceeding, let us ground the Critical Questions introduced in Subsection 3.1 within the context of L_S-perspectives and labelling based semantics.

CQ_{S0} Given a L_S-perspective p and one of its labelling l, is $l(B) = \mathsf{IN}$?

CQ_{S1} Given a L_S-perspective p and one of its labelling l, is $l(A) = \mathsf{IN}$? If an argument B such that $\mathsf{Conc}(B) = bp(\phi)$ does exist, is $l(B) \in \{\mathsf{UND}, \mathsf{OUT}\}$?

CQ_{S2} Given a L_S-perspective p and one of its labelling l, is $l(A) = \mathsf{IN}$?

CQ_{S2} Given all L_S-perspectives p and the set of their labellings L, does $\forall l \in L, l(A) \in \{\mathsf{UND}, \mathsf{OUT}\}$ hold?

CQ_{S3} Given a L_S-perspective p and one of its labelling l, are $l(A) = \mathsf{IN}$ and $l(B) = \mathsf{IN}$?

Using this new definition we can introduce the concept of *BP-perspective*.

Definition 3.2 (BP-perspective)**.** *Let $G = \langle \mathcal{A}_O, \mathcal{A}_M, \mathcal{R}_O, \mathcal{R}_M, \mathcal{S}_A, \mathcal{S}_R \rangle$ be a bimodal argumentation graph, and P the set of its L_{stable}-perspectives [1]. We say that $p \in P$ is a BP-perspective of G iff p is CQ-consistent.*

Example 2 (Antidiscrimination law)**.** *Let us consider a case in which a woman claims to have been discriminated against in her career on the basis of her sex, as she was passed over by male colleagues when promotions came available ($ev1$), and brings evidence showing that in her company all managerial positions are held by men ($ev3$), even though the company's personnel includes many equally qualified women, having worked for a long time in the company, and with equal or better performance ($ev2$). Assume that this practice is deemed to indicate the existence of gender-based discrimination ($indiciaDiscrim$) and that the employer fails to provide prevailing evidence that the woman was not discriminated against ($\neg discrim$). It seems that it may be concluded that the woman was indeed discriminated against on the basis of her sex.*

Consider, for instance, the following formalisation of the European nondiscrimination law, that, in case of presumed discrimination, requires prevailing evidence that no offence was committed—i.e., $bp(\neg discrim)$:

$e1 : ev1$	$e2 : ev2$	$e3 : ev3$
$er1 : ev1 \Rightarrow indiciaDiscrim$	$er2 : ev2 \Rightarrow \neg discrim$	$er3 : ev3 \Rightarrow discrim$
$r1 : indiciaDiscrim \Rightarrow bp(\neg discrim)$		

We can then build the following object-level arguments:

$A_0 :\Rightarrow ev1$ $B_0 :\Rightarrow ev2$ $C_0 :\Rightarrow ev3$
$A_1 : A_0 \Rightarrow indiciaDiscrim$ $B_1 : B_0 \Rightarrow \neg discrim$ $C_1 : C_0 \Rightarrow discrim$
$A_2 : A_1 \Rightarrow bp(\neg discrim)$

and the following meta-level arguments:

$A_{0_{S1}} :\Rightarrow -burdened(A_0)$ $B_{0S1} :\Rightarrow -burdened(B_0)$
$A_{1_{S1}} :\Rightarrow -burdened(A_1)$ $B_{1_{S0}} :\Rightarrow burdened(B_1)$
$A_{2_{S1}} :\Rightarrow -burdened(A_2)$ $B_{1_{S1}} :\Rightarrow -burdened(B_1)$
$C_{0_{S1}} :\Rightarrow -burdened(C_0)$ $B_{1_{S2}} : B_{1_{S0}} \Rightarrow bp\_met(B_1)$
$C_{1_{S1}} :\Rightarrow -burdened(C_1)$ $B_{1_{S3}} : B_{1_{S0}} \Rightarrow \neg bp\_met(B_1)$
$C_1B_{1_{S4}} :\Rightarrow effectiveAttack(C_1, B_1)$ $B_1C_{1_{S4}} :\Rightarrow effectiveAttack(B_1, C_1)$

The resulting graph is depicted in Figure 2. In this case, at the object level, since there are indicia of discrimination (A_1), we can infer the allocation of the burden on non-discrimination (A_2). Moreover, we can build both arguments for discrimination (C_1) and non-discrimination (B_1), leading to a situation of undecidability.

At the meta level we can apply the rule S1 for every argument at the object level ($A_{0_{S1}}, A_{1_{S1}}, A_{2_{S1}}, B_{0S1}, B_{1_{S0}}, C_{0_{S1}}, C_{1_{S1}}$) – where we can establish the absence of the burden for all of them –, and the rule S4 for every attack ($C_1B_{1_{S4}}, B_1C_{1_{S4}}$). By exploiting B_1 and A_2, we can also apply schema S0, and consequently rules S2 and S3. In a few words, we are concluding the meta argumentative structure given by the allocation of the burden of persuasion on argument B_1.

We can now apply the stable labelling to the meta-level graph, thus obtaining three distinct results. For clarity reasons, in the following, we ignore the arguments that are acceptable under every solution.

1. $\mathsf{IN} = \{B_{1_{S1}}, C_1B_{1_{S4}}, B_1C_{1_{S4}}\}, \mathsf{OUT} = \{B_{1_{S0}}, B_{1_{S2}}, B_{1_{S3}}\}, \mathsf{UND} = \{\}$—*i.e., B_1 is not burdened;*

2. $\mathsf{IN} = \{B_{1_{S0}}, B_{1_{S2}}, C_1B_{1_{S4}}, B_1C_{1_{S4}}\}, \mathsf{OUT} = \{B_{1_{S1}}, B_{1_{S3}}\}, \mathsf{UND} = \{\}$—*i.e., B_1 is burdened and the burden is met;*

3. $\mathsf{IN} = \{B_{1_{S0}}, B_{1_{S3}}\}, \mathsf{OUT} = \{B_{1_{S1}}, B_{1_{S2}}, C_1B_{1_{S4}}, B_1C_{1_{S4}}\}, \mathsf{UND} = \{\}$—*i.e., B_1 is burdened and the burden is not met.*

Then, the meta-level results can be reified to the object-level perspectives taking into account the CQ *we have to impose on the solutions and the results given by the perspective evaluation under the grounded semantics. Let us first consider solutions 1 and 2. They lead to the same perspective on the object-level graph—the graph remains unchanged w.r.t. the original graph. If we consider the critical questions attached to the* IN *arguments, both these solutions are not valid. Indeed, according*

to solution 1 the burden is not allocated on argument B_1, but this is in contrast with argument A_2's conclusion (A_2 is IN *under grounded labelling)—i.e., CQ_{S1} is not satisfied. Analogously, solution 2 concludes that B_1 is allocated with the burden and its success to meet the burden, but at the same time, argument B_1 is found undecidable at the object level (B_1 is* UND *under the grounded semantics)—i.e., CQ_{S2} is not satisfied.*

The only acceptable result is the one given by solution 3. In this case, argument B_1 is not capable to meet the burden – $B_{1_{S3}}$ is IN *– and, consequently, it is rejected and deleted from the perspective. Indeed, CQ_{S3} is satisfied. As a consequence, argument C_1 is labelled* IN*. In other words, the argument for non-discrimination fails and the argument for discrimination is accepted.*

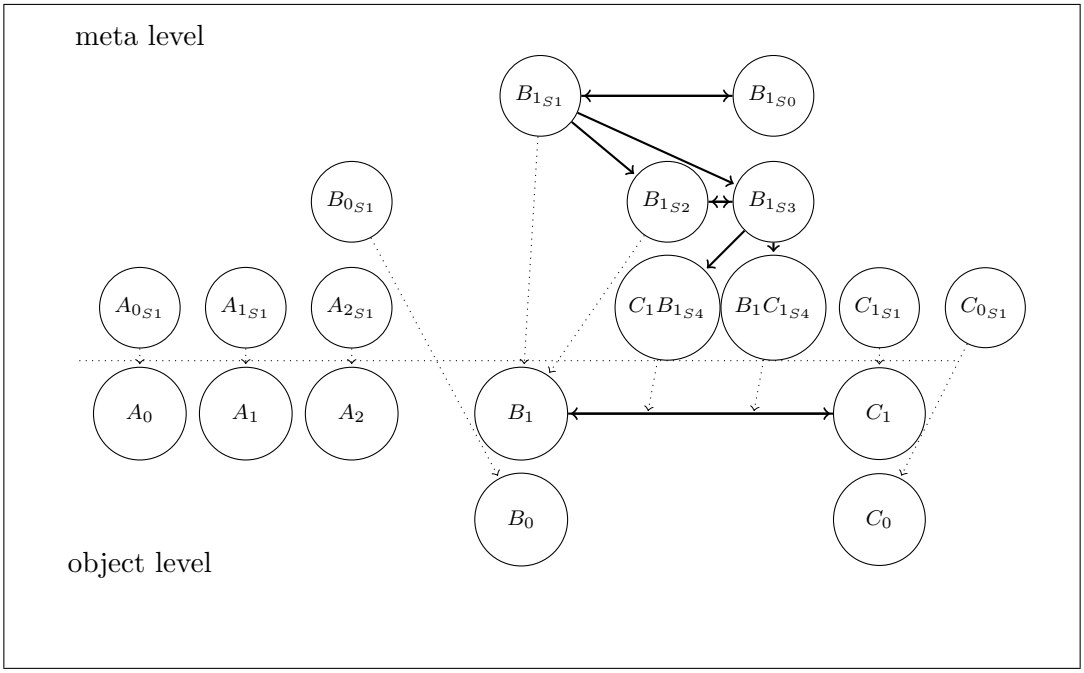

Figure 2: Argumentation graph (object- and meta- level) from Example 2

Before proceeding, let us recall the main definitions from Calegari and colleagues [5], who, in their work, present a semantics dealing with the burden of persuasion allocation on members of the argumentation language.

Definition 3.3 (BP-defeat)**.** *Given a set of burdens of persuasion* BurdPers*, A* ***bp-defeats*** *B iff there exists a subargument B' of B such that:*

1. $\mathsf{Conc}(A) = \overline{\mathsf{Conc}(B')}$ *and*
   (a) $\mathsf{Conc}(A) \notin \mathsf{BurdPers}$, *and* $B' \not\succ A$, *or*
   (b) $\mathsf{Conc}(A) \in \mathsf{BurdPers}$ *and* $A \succ B'$.
2. $\mathsf{Conc}(A) = \neg\mathsf{N}(\rho)$, *where* ρ *is* $\mathsf{TopRule}(B')$.

Definition 3.4 (Grounded BP-labelling)**.** *A grounded* ***BP-labelling*** *of an argumentation graph* G*, relative to a set of burdens* $\mathsf{BurdPers}$*, is a* $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$*-labelling* l *s.t. the set of* UND *arguments is minimal and* $\forall \mathsf{A} \in \mathcal{A}_G$ *with* $\mathsf{Conc}(A) = \phi$

1. $l(A) = \mathsf{IN}$ *iff* $\forall \mathsf{B} \in \mathcal{A}_G$ *such that* B *bp-defeats* $\mathsf{A} : l(B) = \mathsf{OUT}$
2. $l(A) = \mathsf{OUT}$ *iff*
   (a) $\phi \in \mathsf{BurdPers}$ *and* $\exists\ B \in \mathcal{A}_G$ *s.t.* B *bp-defeats* A *and* $l(B) \neq \mathsf{OUT}$
   (b) $\phi \notin \mathsf{BurdPers}$ *and* $\exists\ B \in \mathcal{A}_G$ *such that* B *bp-defeats* A *and* $l(B) = \mathsf{IN}$
3. $l(A) = \mathsf{UND}$ *otherwise.*

Proposition 3.1. *If* $\nexists A, B \in \mathcal{A}_O$ *such that both* A *and* B *have a burden of persuasion on them and* A *is reachable from* B *through* $\mathcal{R}_O$*, the results yielded by the grounded evaluation of* G*'s* BP*-perspectives are congruent with the evaluation of the object-level graph* $\langle \mathcal{A}_O, \mathcal{R}_O \rangle$ *under the Grounded BP-labelling as in Definition 3.4 [5].*

Proof. The burden of persuasion semantics acts like the grounded semantics, with the only difference being that the burdened arguments that would have been UND for the latter could be OUT/IN for the former. So, it is a matter of fact that burdened arguments and arguments connected to them through attack relation can change their state.

Let us consider an argumentation graph $AF\langle \mathcal{A}, \rightsquigarrow \rangle$, and let L_G be the grounded labelling resulting from the evaluation of AF under a grounded semantics. With respect to our framework, and in particular, to the bimodal argumentation graph $G = \langle \mathcal{A}_O, \mathcal{A}_M, \mathcal{R}_O, \mathcal{R}_M, \mathcal{S}_A, \mathcal{S}_R \rangle$, we have, by construction, that every node at the object level, if not burdened, has an undisputed supporting argument at the meta level (S1 or S4). As a consequence, the meta level has no influences on no burdened arguments, and – in the absence of burdened arguments – the evaluation of the object level graph under the grounded semantics would be equal to L_G. It is a matter of fact that the meta level influences only the burdened arguments' state. Accordingly, the extent of this influence and the consequences on the object-level graph will be considered in the following.

Let us consider a single argument $A \in \mathcal{A}$ allocated with the burden of persuasion, thus having the additional argument $B \in \mathcal{A}$ stating the burden on A (as depicted in Figure 1). Computing the stable semantics on the meta-level graph produces the following scenarios:

Stable.a burden on A cannot be proved;

Stable.b burden on A can be proved and the burden is met;

Stable.c burden on A can be proved and the burden is not met.

Accordingly, the stable evaluation of the meta-graph produces three different perspectives of the object level:

(i) argument A is supported—it is not burdened;

(ii) argument A is supported—it satisfies the burden;

(iii) argument A is not supported, and then it is excluded from the object-level graph—it does not meet the burden then it is refuted.

In particular, we have that Stable.a induces *(i)*, Stable.b leads to *(ii)*, while Stable.c induces *(iii)*. Let L_{BP} be this new object-level labelling (obtained by the meta-level stable semantics reification at the object level). Also, let us compare L_{BP} with the initial object-level grounded labelling L_G. Then, the following cases can occur (E is exploited for valid solutions with labelling equivalence, while C is exploited for solutions to be discarded).

- B is OUT or UND in L_G.

  E1 If *(i)* the burden is not allocated and cannot be proven, the meta level does not influence the object level supporting all unburdened arguments. CQ_{S1} is satisfied and L_{BP} is equivalent to L_G.

  C1 If *(ii)* or *(iii)*, in both cases CQ_{S0} is not satisfied—the burden is proved at the meta level and not at the object level.

- B is IN and A is OUT in L_G.

  C2 If *(i)* we have inconsistency on CQ_{S1}—the burden is proved at the object level and not at the meta level.

  C3 If *(ii)* we have inconsistency on CQ_{S2} since A is considered IN at the meta level (supported by the meta-argument) but A is OUT at the object level.

E2 If *(iii)* A is not supported, i.e., removed from the object-level graph. CQ_{S0} and CQ_{S3} are both satisfied. Then, under the grounded semantics, the removal of an OUT argument from a graph is not influent w.r.t. its evaluation, i.e., L_{BP} is equivalent to L_G.[1]

- B is IN and A is IN in L_G.

  C4 If *(i)*, we have inconsistency on CQ_{S1}—the burden is proved at the object level and not at the meta level.

  E3 If *(ii)*, then CQ_{S0} and CQ_{S2} are both satisfied and L_{BP} is equal to L_G.

  C5 If *(iii)* we have an inconsistency because CQ_{S3} is not satisfied.

- B is IN and A is UND in L_G.

  C6 If *(i)*, we have inconsistency on CQ_{S1}—the burden is proved at the object level and not at the meta level.

  C7 If *(ii)*, we have an inconsistency since A is considered IN at the meta level (supported by the meta-argument) but A is UND at the object level—CQ_{S2} is not satisfied.

  E4 If *(iii)* A is not supported, i.e., is removed from the object level, i.e., it can be labelled as OUT in L_{BP} (see [1]). CQ_{S0} and CQ_{S2} are satisfied.

As made evident by the proof, the reification of the meta level upon the object level generates multiple solutions: yet, only one solution for each case can be considered valid w.r.t. critical questions. Moreover, the only valid perspective coincides with the one generated from the bp-labelling in [5]—the burdened argument is labelled OUT in case of indecision (E4). Obviously, the proof can be generalised to configurations taking into account any number of burdened independent arguments—where combinations grow exponentially with the number of burdened arguments. □

4 Burden Inversion

Let us consider a situation in which one argument A is presented for a claim ϕ being burdened, and A (or one of its subarguments) is attacked by a counterargument B, of which the conclusion ψ is also burdened. Intuitively, if both arguments fail to

[1] It can trivially be proved considering that – in the grounded semantics – an OUT argument does not affect other arguments' state, i.e., it is irrelevant and can be removed; of course, also the dual proposition holds, i.e., if L_{BP} build in the meta-frameworks does not consider an argument it can be labelled as OUT in the grounded bp-labelling

satisfy the burden of persuasion, both of them are to be rejected. This is not the case if the inversion of the burden is taken into account [5]—i.e., if no convincing argument for ψ is found, then the attack fails, and the uncertainty on ψ does not affect the status of A. Accordingly, B is rejected for failing to meet its burden, thus leaving A free to be accepted also if it was not able to satisfy the burden of persuasion in the beginning.

The model we propose in this work is able to correctly deal with the inversion of the proof, as we discuss in the next example adapted from [5].

Example 3 (Inversion of the burden)**.** *Let us consider a case in which a doctor caused harm to a patient by misdiagnosing his case. Assume that there is no doubt that the doctor harmed the patient (harm), but it is uncertain whether the doctor followed the guidelines governing this case. Assume that, under the applicable law, doctors are liable for any harm suffered by their patients (liable), but they can avoid liability if they show that they exercised due care in treating the patient (dueDiligence). Let us also assume that a doctor is considered to be diligent if he/she follows the medical guidelines that govern the case (guidelines). The doctor has to provide a convincing argument that he/she was diligent* ($bp(dueDiligence)$)*, and the patient has to provide a convincing argument for the doctor's liability* ($bp(liable)$)*.*

We can formalise the case as follows:

$$
\begin{array}{ll}
f1 : guidelines & f2 : \neg guidelines \\
f3 : harm & r1 : \neg guidelines \Rightarrow \neg dueDiligence \\
r2 : guidelines \Rightarrow dueDiligence & r3 : harm, \sim dueDiligence \Rightarrow liable \\
bp1 : bp(dueDiligence) & bp2 : bp(liable)
\end{array}
$$

We can then build the following object-level arguments:

$$
\begin{array}{ll}
A_0 :\Rightarrow bp(dueDiligence) & A_1 :\Rightarrow bp(liable) \\
A_2 :\Rightarrow guidelines & A_3 :\Rightarrow harm \\
A_4 :\Rightarrow \neg guidelines & A_5 : A_2 \Rightarrow dueDiligence \\
A_1 : A_0 \Rightarrow indiciaDiscrim & B_1 : B_0 \Rightarrow \neg discrim \\
C_1 : C_0 \Rightarrow discrim & A_6 : A_3 \Rightarrow liable \\
A_7 : A_4 \Rightarrow \neg dueDiligence &
\end{array}
$$

According to the original burden semantics, the argument for the doctor's due diligence (A_5) fails to meet its burden of persuasion. Consequently, following the inversion principle, it fails to defeat the argument for the doctor's liability (A_6), which is then able to meet its burden of persuasion.

Let's now analyse the case from the meta-model perspective. Using argument schemes defined in Section 3 we can build the following meta-arguments:

$A_{0_{S1}} :\Rightarrow -burdened(A_0)$ $\quad$ $A_{1_{S1}} :\Rightarrow -burdened(A_1)$

$A_{2_{S1}} :\Rightarrow -burdened(A_2)$ $\quad$ $A_{3_{S1}} :\Rightarrow -burdened(A_3)$

$A_{4_{S1}} :\Rightarrow -burdened(A_4)$ $\quad$ $A_{7_{S1}} :\Rightarrow -burdened(A_7)$

$A_2A_{7_{S4}} :\Rightarrow effectiveAttack(A_2, A_7)$ $\quad$ $A_2A_{4_{S4}} :\Rightarrow effectiveAttack(A_2, A_4)$

$A_4A_{2_{S4}} :\Rightarrow effectiveAttack(A_4, A_2)$

$A_7A_{5_{S4}} :\Rightarrow effectiveAttack(A_7, A_5)$ $\quad$ $A_5A_{7_{S4}} :\Rightarrow effectiveAttack(A_5, A_7)$

$A_4A_{5_{S4}} :\Rightarrow effectiveAttack(A_4, A_5)$ $\quad$ $A_5A_{6_{S4}} :\Rightarrow effectiveAttack(A_5, A_6)$

$A_{5_{S0}} :\Rightarrow burdened(A_5)$ $\quad$ $A_{5_{S1}} :\Rightarrow -burdened(A_5)$

$A_{5_{S2}} : A_{5_{S0}} \Rightarrow bp\_met(A_5)$ $\quad$ $A_{5_{S3}} : A_{5_{S0}} \Rightarrow \neg bp\_met(A_5)$

$A_{6_{S0}} :\Rightarrow burdened(A_6)$ $\quad$ $A_{6_{S1}} :\Rightarrow -burdened(A_6)$

$A_{6_{S2}} : A_{6_{S0}} \Rightarrow bp\_met(A_6)$ $\quad$ $A_{6_{S3}} : A_{6_{S0}} \Rightarrow \neg bp\_met(A_6)$

Connecting the object- and meta-level arguments we obtain the graph in Figure 3. Let us now consider the extensions obtained applying stable semantics to the meta-level graph:

1. $\{A_{6_{S0}}, A_{6_{S2}}, A_{5_{S0}}, A_{5_{S3}}\}$
2. $\{A_{6_{S0}}, A_{6_{S3}}, A_{5_{S0}}, A_{5_{S3}}\}$
3. $\{A_{6_{S0}}, A_{6_{S2}}, A_{5_{S0}}, A_{5_{S2}}, A_5A_{6_{S4}}, A_5A_{7_{S4}}, A_7A_{5_{S4}}, A_4A_{5_{S4}}\}$
4. $\{A_{6_{S0}}, A_{6_{S3}}, A_{5_{S0}}, A_{5_{S2}}, A_5A_{7_{S4}}, A_7A_{5_{S4}}, A_4A_{5_{S4}}\}$
5. $\{A_{6_{S0}}, A_{6_{S2}}, A_{5_{S1}}, A_5A_{6_{S4}}, A_5A_{7_{S4}}, A_7A_{5_{S4}}, A_4A_{5_{S4}}\}$
6. $\{A_{6_{S0}}, A_{6_{S3}}, A_{5_{S1}}, A_5A_{7_{S4}}, A_7A_{5_{S4}}, A_4A_{5_{S4}}\}$
7. $\{A_{6_{S1}}, A_{5_{S0}}, A_{5_{S2}}, A_5A_{6_{S4}}, A_5A_{7_{S4}}, A_7A_{5_{S4}}, A_4A_{5_{S4}}\}$
8. $\{A_{6_{S1}}, A_{5_{S1}}, A_5A_{6_{S4}}, A_5A_{7_{S4}}, A_7A_{5_{S4}}, A_4A_{5_{S4}}\}$
9. $\{A_{6_{S1}}, A_{5_{S0}}, A_{5_{S3}}\}$

The only extensions that produce a CQ-consistent perspective are the first and the second, given that all the others violate at least one of the constraints imposed by the critical questions—e.g. CQ_{S1} for $5, 6, 7, 8, 9$ and CQ_{S2} for $3, 4$. The first perspective acts exactly like the original semantics from [5]—i.e., the argument for the doctor's due diligence (A_5) fails to meet the burden ($A_{5_{S3}}$), and consequently, the argument for doctor's liability (A_6) is able to satisfy its own burden ($A_{6_{S2}}$). However, the model delivers a second result according to which both A_5 and A_6 fail to meet their burden of persuasion ($A_{6_{S3}}$ and $A_{5_{S3}}$). It is the result that we would have expected in absence of the inversion principle.

The example highlights the meta-argumentation model is able to provide both a solution that follows the inversion principle and one not considering it. When the inversion principle is taken into account the number of burdened arguments is maximised in the final extension. Accordingly, we can provide a generalisation of Property 3.1:

Proposition 4.1. *Given the results yielded by the grounded evaluation of G's BP-perspectives, the results that maximise the number of burdened arguments in the* IN *set are congruent with the evaluation of the object-level graph* $\langle \mathcal{A}_O, \mathcal{R}_O \rangle$ *under the grounded-bp semantics as in Definition 3.4 [5].*

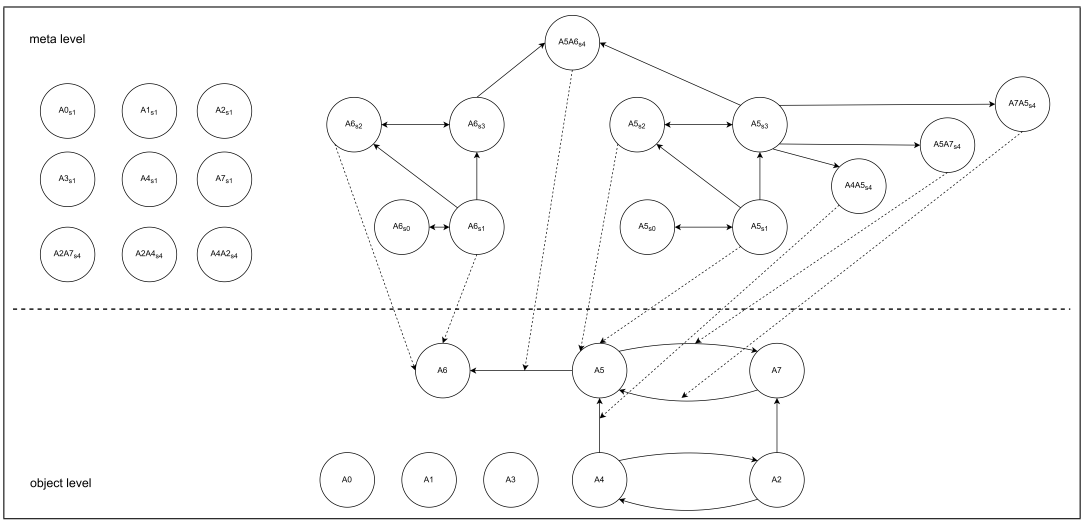

Figure 3: Argumentation graph (object- and meta- level) from Example 3

5 Technological Reification

Despite the benefits of the meta-approach discussed in Section 3 – such as clear separation of concerns, encapsulations of argumentation abstractions and naturalness in terms of human thinking – the method is quite inefficient from a computational perspective. Indeed, the meta-level evaluation leads to a stable semantics computation, with a non-polynomial complexity [8]. This is why, from a technological perspective, the model presented in Section 3 has been reified into a more efficient resolution method.

In a nutshell, the proposed approach exploits the stable semantics to explore the search space at the meta level. Then, in order to identify the final solution, the

grounded assessment of the object level is taken into account—selecting the acceptable scenario according to the critical questions. The idea behind the technological refinement is exactly to leverage the information of those arguments to guide the search—i.e., to exploit the grounded assessment of the object level as an *a priori* constraint. Following this idea, the computation algorithm becomes really simple. The two argumentation levels (object and meta) are collapsed in a single graph, following [3]. Then, the graph is modified dynamically, leveraging the information on the burdened arguments. In a sense, we have a multi-stage evaluation that leads to the modification of the graph itself at every stage.

Let us consider the framework of Example 2. There, two arguments exist, namely B_1 and C_1, attacking each other. Then, another argument, A_2, concludes the presence of the burden on B_1. The grounded evaluation of this framework would lead to a single extension containing argument A_2—i.e., the burden on B_1 has been proved, and we should proceed to verify B_1's compliance with the constraint. According to the model presented in Section 3, the graph should be used to build the meta-level framework expressing all the possible outcomes the burden could lead to. Then, the one leading to an object-level perspective that satisfies all the attached *Critical Questions* would be the correct one. This kind of assessment has one major drawback: we already know from the initial grounded evaluation that B_1 does not satisfy its burden; however, through stable semantics, we explore also the scenarios in which B_1's burden is satisfied, just to discard them later using the *Critical Questions*. The main idea of the technological reification presented in this Section is exactly to use the information generated by the initial grounded assessment to produce a new graph including all the new meta-knowledge.

Let us test this approach with the theory in Example 2. We know that B_1 has a burden on it, but it has not been able to satisfy it. As in the original model, we can use this info to build the argument $B_{1_{S3}}$ using the scheme S3. Intuitively, this new argument claims that "B_1 should be rejected for not being able to defend itself" and, consequently, it throws a new attack against it. If we add these new elements to the original framework, we obtain a new framework containing both object- and meta-arguments on the same level. Its evaluation under grounded semantics leads to the expected result: B_1 is rejected, while C_1 and $B_{1_{S3}}$ are both accepted.

More generally, what we are doing is verifying the *Critical Questions* associated with a meta scheme using the grounded evaluation of the original framework. In this way, we do not need stable semantics to explore all the possible scenarios, but, instead, we can directly select the correct one. For instance, in the case of Example 2, $B_{1_{S3}}$ satisfies its critical questions, while $B_{1_{S2}}$ does not. In the case B_1 were able to satisfy its burden, then just $B_{1_{S2}}$ would have been instantiated, and consequently, no new attacks would have been introduced in the framework.

Summing up, given a constraint $bp(x)$, then for every argument A having x as its conclusion a new argument B can be introduced in the graph. This argument represents the possibility of A failing/succeeding to meet the burden—expressed by S3 and S2 in the meta-model. A and B's interaction is decided according to the A's ability to satisfy the burden under the grounded semantics:

i) iff A is OUT or UND, then B is an instance of scheme S3, and consequently an attack from B to A is introduced;

ii) iff A is IN, then B is an instance of scheme S2, then no attack is introduced.

Basically, through the first evaluation of the graph, the knowledge required to choose between schemes S3 and S2 is obtained—i.e. the stable semantics evaluation becomes superfluous.

Let us now apply the new approach to Example 3 to see whether the inversion principle is supported or not. If we consider the grounded evaluation of the object-level framework, we obtain two burdened arguments, A_5 and A_6, both failing to satisfy the persuasion constraint. According to our algorithm, we can introduce two meta-arguments based on scheme S3 in the framework, one attacking A_5, and the other A_6. The evaluation of this framework under grounded semantics would of course lead to an undesirable result—i.e. both arguments A_5 and A_6 are rejected.

The enforcement of the inversion mechanism requires a procedural evaluation of the burdened arguments—i.e., we should first evaluate those arguments whose acceptability does not depend on burdened arguments not yet evaluated, and then we apply the algorithm again until all the burdens have been evaluated. For instance, in Example 3 we should first introduce argument $A_{5_{S3}}$ in the graph, and then use the results of this new framework to evaluate the consequences on A_6. Accordingly, the dependencies among burdened arguments are respected—i.e., we enforce the inversion principle.

More formally, given an argumentation framework $AF = \langle \mathcal{A}, \rightsquigarrow \rangle$ along with its grounded extension E_G, we can define the set of burdens to evaluate B_e as

$$\{A_0 \in E_G | \mathsf{Conc}(A_0) = bp(a) \textit{ and } \nexists b \in E_G \textit{ s.t. } \mathsf{Conc}(b) = bp\_met(A_1) \textit{ or } \neg bp\_met(A_1) \textit{ with } Conc(A_1) = a\}$$

Then we can define the reduction R_{B_e} of B_e as:

$$\{bp(a) \in B_e \mid \nexists bp(b) \in B_e \textit{ s.t. } a \textit{ is reachable from } b \textit{ through} \rightsquigarrow\}$$

In simpler terms, the reduction set contains all the burdens on the arguments whose status does not depend on other burdened arguments. Then, given an AF and its

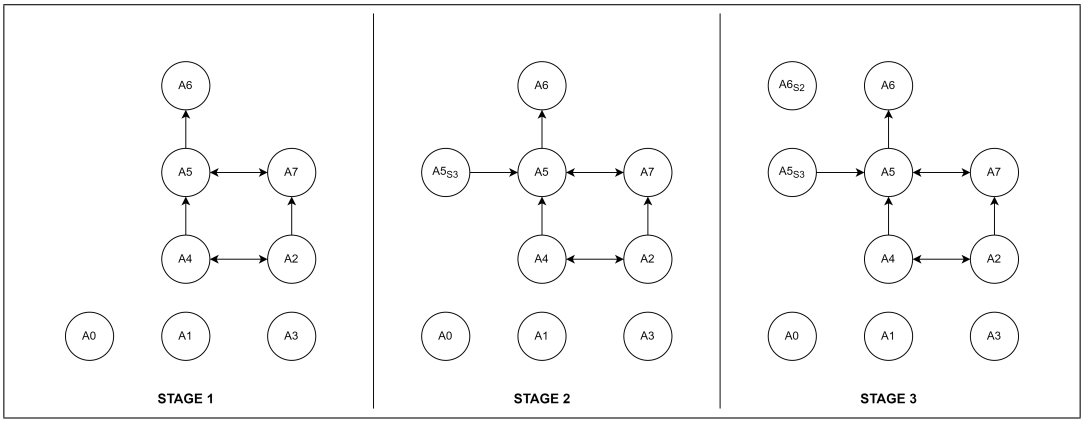

Figure 4: Staged evaluation of Example 3

grounded extension we can use the reduction set to produce a new framework AF_1 containing the meta-arguments for the burdens in the set. We can then recursively apply the same procedure on AF_1 until no elements remain to be evaluated in the reduction set. Understandably, the procedure requires the absence of cycles in the burdened arguments in order to derive a partial ordering over the burdens to evaluate. When all the elements in B_e are independent, the reduction set R_{B_e} is the same as B_e—i.e., the procedure is a generalisation of the naive algorithm introduced at the beginning of this section and used in the evaluation of Example 2.

Figure 4 shows Example 3's evaluation steps. The graph on the left is obtained from the initial theory. We can compute the set of burdens ($\{A_0, A_1\}$) and its reduction ($\{A_0\}$). The new knowledge is used to build the framework in the middle by adding an instance of scheme S3 relative to argument A_5 and its attack. Again, we compute the set of burdens ($\{A_0\}$) and its reduction ($\{A_0\}$), and use it to instantiate scheme S2 in the graph on the right. Now the set of burdens to evaluate is empty and we have our final result: argument A_5 fails to satisfy its burden and it is rejected, thus making it possible for A_6 to satisfy its burden.

5.1 Implementation in Arg2P

The algorithm has been tested and implemented in the Arg2P framework[2] [4, 12]. Please note that the equivalence of the optimised procedure with the formal model presented in the paper has for now only been conjectured, thus remaining still unproven. Figure 5 shows the tool evaluation of the example discussed in Example 2.

[2] `http://arg2p.apice.unibo.it/`

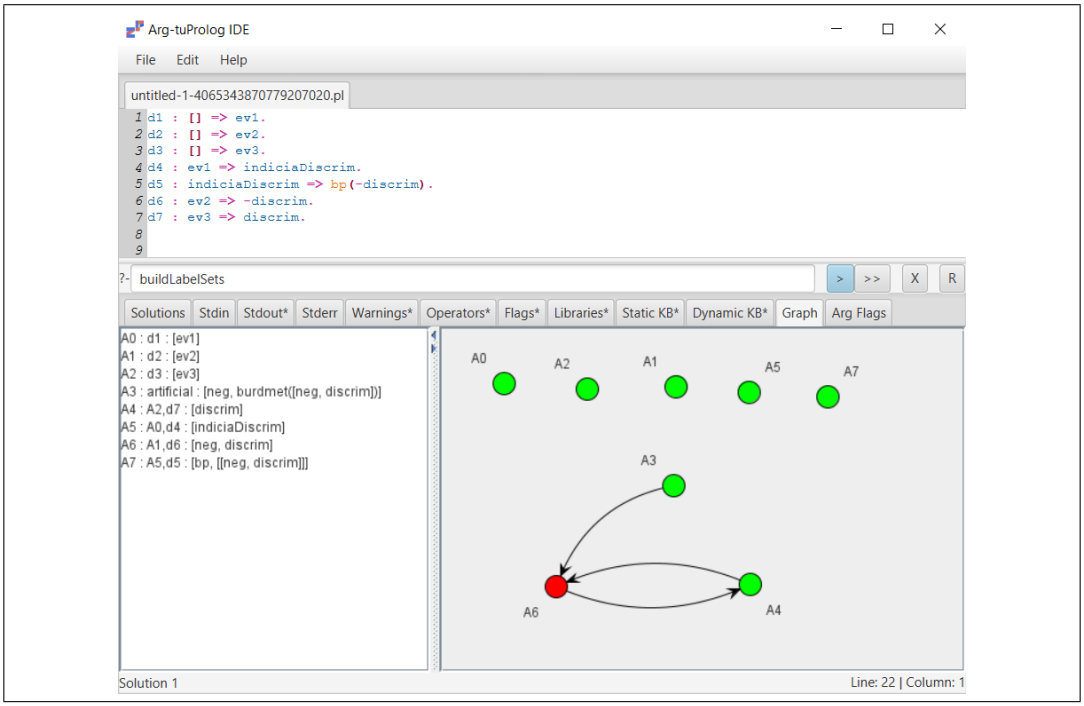

Figure 5: Arg2P evaluation of Example 2

So, the entire process is based on grounded semantics and reachability checking—both polynomial complexity [8]. The algorithm requires $m + 1$ evaluation stages to end – where m is the number of connected burdened arguments –, then the final complexity is polynomial.

6 Related

Our approach relies on the work from [10, 11] introducing the required abstraction at the meta level. In particular, the first formalisation of meta-argumentation synthesising bimodal graphs, structured argumentation, and argument schemes in a unique framework is presented in [10]. There, a formal definition of the meta-ASPIC framework is provided as a model for representing object arguments. Along the same line, bimodal graphs are exploited in [11] for dealing with arguments sources' trust. In [11] ASPIC+ is used instead of meta-ASPIC at the object level and on a set of meta-predicates related to the object level arguments and the schemes in the meta level, as in our approach. Both [10] and [11] use critical questions for managing

attacks at the meta level.

Our framework and its model mix the two approaches by exploiting bimodal graphs in ASPIC+ and defining all the burdens abstractions at the meta-level. The reification of the meta level at the object level allows the concept of the burden of persuasion to be properly dealt with—i.e., arguments burdened with persuasion have to be rejected when there is uncertainty about them. As a consequence, those arguments become irrelevant to the argumentation framework including them: not only do they fail to be included in the set of accepted arguments, but they also are unable to affect the status of the arguments they attack.

An interesting connection with our work could be drowned with the multi-sorted argumentation networks proposed in [16], and their reification in the modal fibring approach from [2]. The main idea of their work is to allow different parts of a framework – called cells – to be evaluated under different semantics. In a nutshell, a set of arguments is a multi-sorted extension only if it is the union of the extensions computed on the qualified arguments – i.e., arguments not defeated and defended from attacks coming from other cells – of the single cells composing the framework. The modal fibring approach from [2] allows every cell to be represented as a separate argumentation framework, with the possibility of modality used to express inter-cell attacks within these frameworks. Their work could appear similar to the bimodal approach in the way different graphs are used to derive the final results, but there is an important difference to consider: the nature of the relation used to connect the different graphs. Bimodal graphs exploit a support relation to model the dependency of an N-level argument on an N+1-level argument, while multi-sorted networks are based on inter-cell attacks. A naive transposition of our work in a multi-sorted setting would require three steps:

1. the use of the supports to build the attack set connecting meta and object level in order to compose a single graph made of two cells (object and meta);

2. enumeration of the multi-sorted extensions using grounded semantics for the object-cell and stable semantics for the meta-cell;

3. evaluation of the extensions using the *Critical Questions* connected to the meta-argument in them.

However, the transposition would bring no real benefits, while at the same time losing the encapsulation and clarity given by the multi-level structuring of the problem.

7 Conclusions

In this paper we present a meta-argumentation approach for the burden of persuasion in argumentation, discussing interconnections with the state of the art. We show how this model easily deals with all the nuances of burdens such as reasoning over the concept of the burden itself, thus leading to a full-fledged, interoperable framework open to further extensions. Moreover, the model correctly deals with the inversion of the burden.

Future research will be devoted to studying the properties of our meta framework and the connection of our framework with meta-ASPIC for argumentation. We also plan to inquire about the way in which our model fits into legal procedures and enables their rational reconstruction.

References

[1] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.

[2] Howard Barringer, Dov M. Gabbay, and John Woods. Modal and temporal argumentation networks. *Argument & Computation*, 3(2-3):203–227, 2012.

[3] Guido Boella, Dov M. Gabbay, Leendert van der Torre, and Serena Villata. Meta-argumentation modelling I: Methodology and techniques. *Studia Logica*, 93(2–3):297, 2009.

[4] Roberta Calegari, Giuseppe Pisano, Andrea Omicini, and Giovanni Sartor. Arg2P: An argumentation framework for explainable intelligent systems. *Journal of Logic and Computation*, 32(2):369–401, March 2022. Special Issue from the 35th Italian Conference on Computational Logic (CILC 2020).

[5] Roberta Calegari, Regis Riveret, and Giovanni Sartor. The burden of persuasion in structured argumentation. In Juliano Maranhão and Adam Zachary Wyner, editors, *ICAIL'21: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, ICAIL'21, pages 180–184. ACM, June 2021.

[6] Roberta Calegari and Giovanni Sartor. A model for the burden of persuasion in argumentation. In Serena Villata, Jakub Harašta, and Petr Křemen, editors, *Legal Knowledge and Information Systems. JURIX 2020: The Thirty-third Annual Conference*, volume 334 of *Frontiers in Artificial Intelligence and Applications*, pages 13–22, Brno, Czech Republic, 9-11 2020. IOS Press.

[7] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5—6):286–310, 2007.

[8] Markus Kröll, Reinhard Pichler, and Stefan Woltran. On the complexity of enumerating the extensions of abstract argumentation frameworks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017*, pages 1145–1152, Melbourne, Australia, 2017.

[9] Sanjay Modgil and Henry Prakken. The *ASPIC*$^+$ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014.

[10] Jann Müller, Anthony Hunter, and Philip Taylor. Meta-level argumentation with argument schemes. In *International Conference on Scalable Uncertainty Management*, volume 8078 of *Lecture Notes in Computer Science*, pages 92–105, Washington, DC, USA, 2013. Springer.

[11] Gideon Ogunniye, Alice Toniolo, and Nir Oren. Meta-argumentation frameworks for multi-party dialogues. In *International Conference on Principles and Practice of Multi-Agent Systems*, volume 11224 of *Lecture Notes in Computer Science*, pages 585–593, Tokyo, Japan, 2018. Springer.

[12] Giuseppe Pisano, Roberta Calegari, Andrea Omicini, and Giovanni Sartor. A mechanism for reasoning over defeasible preferences in Arg2P. In Stefania Monica and Federico Bergenti, editors, *CILC 2021 – Italian Conference on Computational Logic. Proceedings of the 36th Italian Conference on Computational Logic*, volume 3002 of *CEUR Workshop Proceedings*, pages 16–30, Parma, Italy, 7-9 September 2021.

[13] Giuseppe Pisano, Roberta Calegari, Andrea Omicini, and Giovanni Sartor. Burden of persuasion in argumentation: A meta-argumentation approach. In Marcello D'Agostino, Fabio Aurelio D'Asaro, and Costanza Larese, editors, *Advances in Argumentation in Artificial Intelligence 2021*, volume 3086 of *CEUR Workshop Proceedings*, pages 5:1–5:19, February 2022. Proceedings of the Workshop on Advances in Argumentation in Artificial Intelligence (AI^3 2021), co-located with the 19th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2021), Milan, November 29, 2021.

[14] Henry Prakken. AI & Law, logic and argument schemes. *Argumentation*, 19(3):303–320, 2005.

[15] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[16] Tjitze Rienstra, Alan Perotti, Serena Villata, Dov M. Gabbay, and Leendert W. N. van der Torre. Multi-sorted argumentation. In Sanjay Modgil, Nir Oren, and Francesca Toni, editors, *Theorie and Applications of Formal Argumentation – First International Workshop, TAFA 2011*, volume 7132 of *Lecture Notes in Computer Science*, pages 215–231. Springer, 2011.

[17] Gerard Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90(1–2):225–279, 1997.

[18] Douglas Walton, Christopher Reed, and Fabrizio Macagno. *Argumentation Schemes*. Cambridge University Press, United Kingdom, 2008.

 Received 31 May 2022