Burden of Persuasion in Meta-Argumentation

Giuseppe Pisano[1][0000−0003−0230−8212], Roberta Calegari[1][0000−0003−3794−2942], Andrea Omicini[2][0000−0002−6655−3869], and Giovanni Sartor[1][0000−0003−2210−0398]

[1] Alma AI – Alma Mater Research Institute for Human-Centered Artificial Intelligence, Alma Mater Studiorum—Università di Bologna, Italy
{g.pisano, roberta.calegari, giovanni.sartor}@unibo.it
[2] Dipartimento di Informatica – Scienza e Ingegneria (DISI), Alma Mater Studiorum—Università di Bologna, Italy
andrea.omicini@unibo.it

Abstract. This work defines a burden of persuasion meta-argumentation model interpreting the 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). Finally, an example in the law domain addressing the problem of burden inversion is discussed in detail.

Keywords: burdens of persuasion · argumentation · meta-argumentation

1 Introduction

In this work we discuss the model of the burden of persuasion in structured argumentation [2, 3] 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 dealing also with reasoning over the burden of persuasion concepts.

The work finds its foundation in the approaches of meta-argumentation that emphasize the inner nature of arguments and dialogues as inherently meta-logical [6, 7]. Our approach relies on the works from [6, 7] introducing 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 interaction between the object level and the meta level—following the account in [6].

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 [3]. Finally, Section 4 discuss a real case study in the law domain dealing with the problem of burden inversion. Conclusion are drawn in Section 5.

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 both to define meta-level concepts and to understand the interactions of object-level and meta-level arguments [7, 6]. Accordingly, Subsection 2.1 presents the object-level argumentation language exploited by our model, leveraging on an ASPIC$^+$-like argumentation framework [9]. Then, Subsection 2.2 introduces bimodal argumentation graphs main definitions. Finally, in Subsection 2.3, the meta-level argumentation language based on the use of argument schemes [11] is introduced.

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 $\bar{\phi} = \neg p$, while if ϕ is $\neg p$, then $\bar{\phi}$ is p.*

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

Definition 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 and *bp* predicates are brought into relation through defeasible rules.

Definition 2 (Defeasible rule). *A **defeasible rule** r has the following form: $\rho : \phi_1, ..., \phi_n, \sim \phi'_1, ..., \sim \phi'_m \Rightarrow \psi$ 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 $Antecedent(r)$ and ψ 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 correspondingly to specify that the rule is inapplicable [5].

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

Definition 3 (Superiority relation). *A **superiority relation** $\succ$ over a set of rules Rules is an antireflexive and antisymmetric binary relation over Rules.*

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

Definition 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, by chaining rules from the theory, we can construct arguments [5, 4, 10].

Definition 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 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: an argument A is preferred over another argument B if the top rule of A is stronger than the top rule of B.

Definition 6 (Preference relation). *A **preference relation** $\succ$ is a binary relation over a set of arguments $\mathcal{A}$: an 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 according to the attack relation.

Definition 7 (Attack). *An 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*
  - *$\mathsf{Conc}(A) = \bar{\phi}$ for some $B' \in \mathsf{Sub}(B)$ of the form $B''_1, ..., B''_M \Rightarrow \phi$ and $B' \not\succ A$, or*
  - *$\mathsf{Conc}(A) = \phi$ for some $B' \in \mathsf{Sub}(B)$ such that $\sim \phi \in Antecedent(\mathsf{TopRule}(B'))$*

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

Definition 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 9 (Labelling). *Let G be an argumentation graph. An $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-**labelling** L of G is a total function $\mathcal{A}_G \to \{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$. The set of all $\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}$-**labellings** of G will be denoted as $\mathcal{L}(\{\mathsf{IN}, \mathsf{OUT}, \mathsf{UND}\}, G)$.*

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

Definition 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 allow capturing scenarios in which arguments are categorised in multiple levels—only 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 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 $\mathcal{S}_R$. This supports are the only structural interaction between the meta and the object levels. Condition (8) in the above definition ensures that every object-level argument is supported by at least one meta-level argument, while 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 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 bi-modal 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*

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

Consequently, an object argument may occur in one perspective and not in another according to the results yielded by the meta-level argumentation graph.

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. At this purpose, in this work – following the example in [7] – we leverage on argument schemes [11]. In a few words, argumentation schemes are commonly used patterns of reasoning. They can be formalised in a rules-like form [8] 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 13 (Meta-predicate). *A meta-predicate P_M is a symbol which represents a property or a relation between object-level arguments. Let be $\mathcal{M}$ the set of all P_M.*

Definition 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, conclusions. Let be $\mathcal{O}$ the set of all O_M.*

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

Definition 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 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* $\mathsf{Conc}(A)$*;*
– *we denote with* $CQ(A)$ *the critical questions associated to scheme* s*.*

The same notation introduced for standard arguments in Notation 2 applies also to meta-arguments. We can now define attacks over meta-arguments.

Definition 17 (Meta-Attack). *An argument* A ***attacks*** *argument* B *(on* B'*) iff*

– $\mathsf{Conc}(A) = \bar{P}$ *for some* $B' \in \mathsf{Sub}(B)$ *of the form* $B''_1, ..., B''_M \Rightarrow P$ *or*
– $\mathsf{Conc}(A) = P$ *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 8, 9, 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, it 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 admissible 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:

N.1 the notion of the burden itself expressing the possibility for an argument to be allocated with a burden of persuasion (i.e., *burdened argument*)
N.2 the possibility that this burden is satisfied (that is, a *burden met*) or not satisfied
N.3 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 level 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.

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 **N.1**). 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 modelled through the following argument scheme:

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

$$Are\ arguments\ A\ and\ B\ provable? \quad (\mathsf{CQ_{S0}})$$

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. Of course an argument produced using this scheme holds only if both the arguments A and B on which the inference is based hold—critical question $\mathsf{CQ_{S0}}$.

Analogously, we introduce the scheme $\mathsf{S1}$ representing the absence of such an allocation:

$$conclusion(A, \phi), \sim conclusion(B, bp(\phi)) \Rightarrow \neg burdened(A) \quad (\mathsf{S1})$$

$$Is\ argument\ A\ provable?\ Is\ argument\ B\ really\ unprovable? \quad (\mathsf{CQ_{S1}})$$

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 (**N.2**).

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

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

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

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

where bp_met is the meta-predicate stating the burden has been met. It is important to notice that these two schemes 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 $\mathsf{S2}$, we have that only if exists a burden of persuasion on argument A, and A is admissible ($\mathsf{CQ_{S3}}$), then the burden is satisfied. On the other side, the validity of $\mathsf{S3}$ is linked 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 acceptability is the burden itself—i.e., attacks from other arguments do not influence the burdened argument status that 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 the argument rejection and, as a direct consequence, the inability to effectively attack other arguments. In order to capture the nuance to differentiate among effective or ineffective object level attacks w.r.t the concept of burden of persuasion (**N.3**), 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 not fail to meet their burden? $(\mathsf{CQ_{S4}})$

where *attack* is a structural meta-predicate stating an attack relation at the object level, while *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 fail to satisfy the burden, then the attack is considered not effective w.r.t. the allocation of the burden.

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

Example 1 (Base Example).

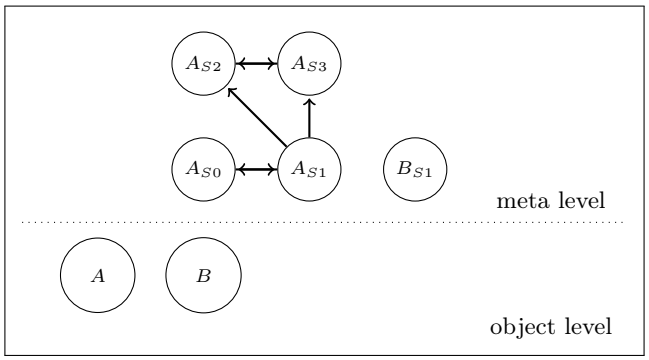
Let us consider two object-level arguments A and B, concluding the literals a and $bp(a)$ respectively. Using the schemes in Subsection 3.1 we can build the following meta-level arguments:

- 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 in which an argument concluding a *bp* literal exists at the object-level, but it is found not admissible.
- 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 (Fig. 1) points out the relations actually implicit in the notion of 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 the Dung's preferred semantics [1], we can obtain two distinct outcomes: the burden is not satisfied, i.e., argument A_{S3} is accepted, and consequently, A_{S2} is rejected, or 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 discussed 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, i.e., 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.



Meta-level arguments:

$A_{S0} :\Rightarrow burdened(A)$
$A_{S1} :\Rightarrow \neg burdened(A)$
$A_{S2} : A_{S0} \Rightarrow bp\_met(A)$
$A_{S3} : A_{S0} \Rightarrow \neg bp\_met(A)$
$B_{S1} :\Rightarrow \neg burdened(B)$

Object-level arguments:

$A :\Rightarrow a$
$B :\Rightarrow bp(a)$

Fig. 1. Argumentation graph (object- and meta-level) for 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 11 (Subsection 2.2), every node 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 the node is 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 the burden on A is satisfied (S2) or that there is no burden allocated on it (S1).

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

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

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 of the original burden of persuasion labelling semantics [3].

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

Definition 18 (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 ($CQ(A)$).*

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

Definition 19 (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 w.r.t. the results given by the grounded evaluation of p, p is CQ-consistent.*

Proposition 1. *If $\nexists (A, B) \in \mathcal{R}_O$ such that both A and B have a burden of persuasion on them, 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 semantics as presented in [3].*

Example 2 (Antidiscrimination law example).

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 and that the employer fails to provide prevailing evidence that the woman was not discriminated against. 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:

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

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:

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

The resulting graph is depicted in Fig. 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 ($A_{0_{S1}}, A_{1_{S1}}, A_{2_{S1}}, B_{0}S1, B_{1_{S0}}, C_{0_{S1}}, C_{1_{S1}}$) – 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 admissible under every solution.

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

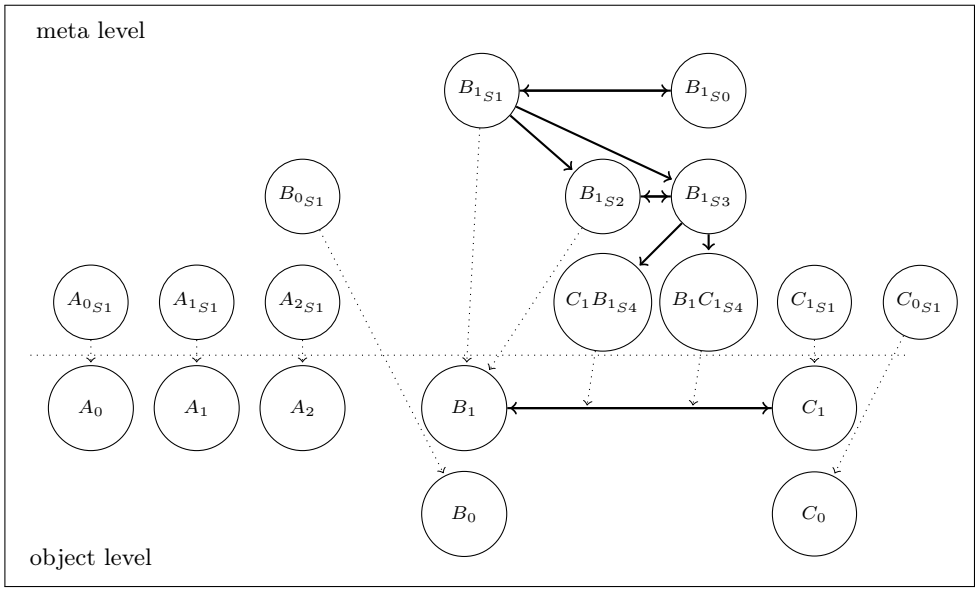


Fig. 2. Argumentation graph (object- and meta-level) for Example 2

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 satisfy the burden of persuasion, we would have to reject both of them. This is not the case if we take into account the inversion of the burden [2]—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 [2].

Example 3 (Inversion of the burden of proof).

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, 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, but they can avoid liability if they show that they exercised due care in treating the patient. Let also assume that a doctor is considered to be diligent if he/she follows the medical guidelines that govern the case. The doctor has to provide a convincing argument that he/she was diligent, and the patient has to provide a convincing argument for the doctor's liability.

We can formalise the case as follows:

$f1 : guidelines$ $f2 : \neg guidelines$ $f3 : harm$
$r1 : \neg guidelines \Rightarrow \neg dueDiligence$ $r2 : guidelines \Rightarrow \neg dueDiligence$
$r3 : harm, \sim dueDiligence \Rightarrow liable$
$bp1 : bp(dueDiligence)$ $bp2 : bp(liable)$

We can then build the following object-level arguments:

$A_0 :\Rightarrow bp(dueDiligence)$ $A_1 :\Rightarrow bp(liable)$ $A_2 :\Rightarrow guidelines$
$A_3 :\Rightarrow harm$ $A_4 :\Rightarrow \neg guidelines$ $A_5 : A_2 \Rightarrow dueDiligence$
$A_1 : A_0 \Rightarrow indiciaDiscrim$ $B_1 : B_0 \Rightarrow \neg discrim$ $C_1 : C_0 \Rightarrow discrim$
$A_6 : A_3 \Rightarrow liable$ $A_7 : A_4 \Rightarrow \neg dueDiligence$

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 under the meta-model perspective. Using argument schemes defined in Section 3 we can build the following meta-arguments:

$A_{0_{S1}} :\Rightarrow -burdened(A_0)$ $A_{1_{S1}} :\Rightarrow -burdened(A_1)$
$A_{2_{S1}} :\Rightarrow -burdened(A_2)$ $A_{3_{S1}} :\Rightarrow -burdened(A_3)$
$A_{4_{S1}} :\Rightarrow -burdened(A_4)$ $A_{7_{S1}} :\Rightarrow -burdened(A_7)$
$A_2A_{7_{S4}} :\Rightarrow effectiveAttack(A_2, A_7)$ $A_2A_{4_{S4}} :\Rightarrow effectiveAttack(A_2, A_4)$
$A_4A_{2_{S4}} :\Rightarrow effectiveAttack(A_4, A_2)$
$A_7A_{5_{S4}} :\Rightarrow effectiveAttack(A_7, A_5)$ $A_5A_{7_{S4}} :\Rightarrow effectiveAttack(A_5, A_7)$
$A_4A_{5_{S4}} :\Rightarrow effectiveAttack(A_4, A_5)$ $A_5A_{6_{S4}} :\Rightarrow effectiveAttack(A_5, A_6)$
$A_{5_{S0}} :\Rightarrow burdened(A_5)$ $A_{5_{S1}} :\Rightarrow -burdened(A_5)$
$A_{5_{S2}} : A_{5_{S0}} \Rightarrow bp\_met(A_5)$ $A_{5_{S3}} : A_{5_{S0}} \Rightarrow \neg bp\_met(A_5)$
$A_{6_{S0}} :\Rightarrow burdened(A_6)$ $A_{6_{S1}} :\Rightarrow -burdened(A_6)$
$A_{6_{S2}} : A_{6_{S0}} \Rightarrow bp\_met(A_6)$ $A_{6_{S3}} : A_{6_{S0}} \Rightarrow \neg bp\_met(A_6)$

Connecting the object- and meta- level arguments we obtain the graph in Fig. 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, 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 [2]—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 the one not considering it. When the inversion principle is taken into account the number of burdened arguments are maximised in the final extension. Accordingly, we can provide a generalisation of Property 1:

Proposition 2. *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 presented in [3].*

5 Conclusions

In this paper we present a meta-argumentation approach for the burden of persuasion in argumentation. Our approach relies on the work from [6, 7] introducing the required abstraction at the meta level. In particular, [6] presents the first formalisation of meta-argumentation synthesising bimodal graphs, structured argumentation, and argument schemes in a unique framework. There, a formal definition of the meta-ASPIC framework is provided as a model for representing object arguments. Along the same line, [7] exploits bimodal graphs for dealing with arguments sources' trust. In [7] 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 [6] and [7] use critical questions for managing attacks at the meta level.

Our framework and its model mix the two approaches exploiting bimodal graphs in ASPIC+ and defining all the burdens abstractions at the meta-level.

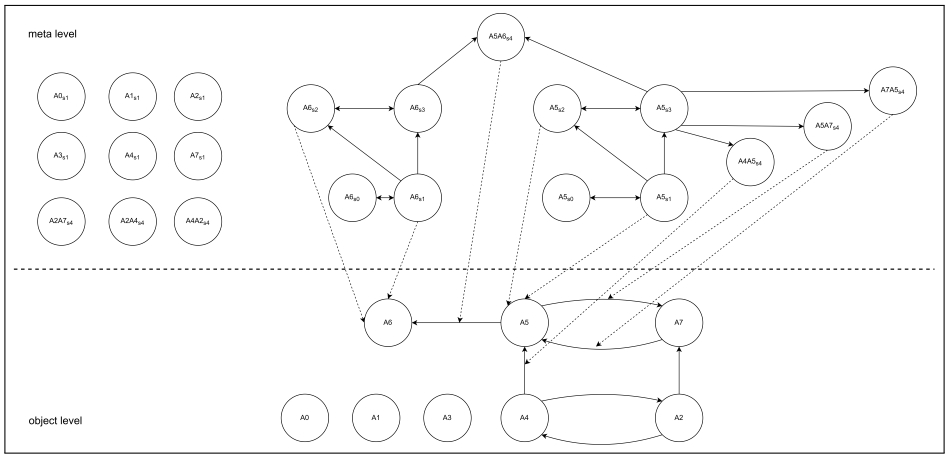


Fig. 3. Argumentation graph (object- and meta-level) for Example 3

The reification of the meta level at the object level allows the concept of 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 they fail to be included in the set of the accepted arguments, but they also are unable to affect the status of the arguments they attack.

We show how this model easily deals with all the nuances of burdens – w.g., 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 of proof.

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

Acknowledgements 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).

References

1. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. The Knowledge Engineering Review **26**(4), 365–410 (2011). https://doi.org/10.1017/S0269888911000166
2. Calegari, R., Riveret, R., Sartor, G.: The burden of persuasion in structured argumentation. In: Maranhão, J., Wyner, A.Z. (eds.) 17th International Conference on Artificial Intelligence and Law (ICAIL'21). pp. 180–184. ACM (Jun 2021). https://doi.org/10.1145/3462757.3466078

3. Calegari, R., Sartor, G.: A model for the burden of persuasion in argumentation. In: Villata, S., Harašta, J., Křemen, P. (eds.) Legal Knowledge and Information Systems, Frontiers in Artificial Intelligence and Applications, vol. 334, pp. 13–22. IOS Press, Brno, Czech Republic (2020). https://doi.org/10.3233/FAIA200845
4. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artificial Intelligence **171**(5–6), 286–310 (2007). https://doi.org/10.1016/j.artint.2007.02.003
5. Modgil, S., Prakken, H.: The $ASPIC^{+}$ framework for structured argumentation: a tutorial. Argument & Computation **5**(1), 31–62 (2014). https://doi.org/10.1080/19462166.2013.869766
6. Müller, J., Hunter, A., Taylor, P.: Meta-level argumentation with argument schemes. In: International Conference on Scalable Uncertainty Management. Lecture Notes in Computer Science, vol. 8078, pp. 92–105. Springer (2013). https://doi.org/10.1007/978-3-642-40381-1
7. Ogunniye, G., Toniolo, A., Oren, N.: Meta-argumentation frameworks for multi-party dialogues. In: PRIMA 2018: Principles and Practice of Multi-Agent Systems. Lecture Notes in Computer Science, vol. 11224, pp. 585–593. Springer (2018). https://doi.org/10.1007/978-3-030-03098-8_45
8. Prakken, H.: Ai & Law, logic and argument schemes. Argumentation **19**(3), 303–320 (2005). https://doi.org/10.1007/s10503-005-4418-7
9. Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation **1**(2), 93–124 (2010). https://doi.org/10.1080/19462160903564592
10. Vreeswijk, G.: Abstract argumentation systems. Artificial Intelligence **90**(1–2), 225–279 (1997). https://doi.org/10.1016/S0004-3702(96)00041-0
11. Walton, D., Reed, C., Macagno, F.: Argumentation Schemes. Cambridge University Press, United Kingdom (2008). https://doi.org/10.1017/CBO9780511802034