

Burden of persuasion in argumentation: A meta-argumentation approach

Giuseppe Pisano¹, Roberta Calegari¹, Andrea Omicini² and Giovanni Sartor¹

¹*Alma AI – Alma Mater Research Institute for Human-Centered Artificial Intelligence, ALMA MATER STUDIORUM—Università di Bologna, Italy*

²*Dipartimento di Informatica – Scienza e Ingegneria (DISI), ALMA MATER STUDIORUM—Università di Bologna, Italy*

Abstract

This paper examines the view of the burden of persuasion as meta argument and elaborates the meta-argumentative aspects of a burden-of-persuasion semantics in argumentation. An argumentation framework composed of a *meta level* (dealing with the burden) and an *object level* (dealing with standard arguments) is proposed and discussed, and its equivalence with the burden-of-persuasion model in argumentation is proved. Finally, a computationally-feasible implementation of the meta-argumentation approach is presented.

Keywords

burdens of persuasion, argumentation, meta-argumentation, reasoning over burdens

1. Introduction

Given that argumentation also involves putting forward arguments about arguments, in this paper we start from the view that arguments and dialogues are inherently meta-logical, as already acknowledged by many research works [1]. For instance, a statement that serves as a justification of an argument is a statement about an argument: the argument for which the justification serves should itself be referred to in the justification. Accordingly, any proper formalisation of arguments and related abstractions should embrace this aspect.

A meta-level argumentation framework is instantiated by arguments that make statements about arguments, their interactions, and their evaluation at an object-level argumentation framework. Here we focus on the very concept of burden of persuasion and on the design of the meta model for dealing with burdens. Generally speaking, we can say that burdens of persuasion distribute dialectical responsibilities between the parties in a dialogue. In other words, when a party has a burden of persuasion of type b relative to a claim ϕ and does not provide the kind of arguments or evidence required by b , then the party will lose on ϕ . Losing on the burdened claim means that, for the purpose of

AI³'21: 5th Workshop on Advances In Argumentation In Artificial Intelligence

✉ g.pisano@unibo.it (G. Pisano); roberta.calegari@unibo.it (R. Calegari); andrea.omicini@unibo.it (A. Omicini); giovanni.sartor@unibo.it (G. Sartor)

🆔 0000-0003-0230-8212 (G. Pisano); 0000-0003-3794-2942 (R. Calegari); 0000-0002-6655-3869 (A. Omicini); 0000-0003-2210-0398 (G. Sartor)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

the dialectic interaction at stake, it will be assumed that ϕ has not been established, not even as a relevant possibility. Burdens of persuasion complement the analysis of dialectical frameworks that are provided by argumentation systems. In particular, they are important in adversarial contexts: they facilitate the process of reaching a single outcome in contexts of doubt and lack of information. This is obtained by ruling out (considering them as unacceptable) those arguments that fail to meet any applicable burden.

In this work we discuss the model of the burden of persuasion in structured argumentation [2, 3] under a meta-argumentative and meta-logic approach, which leads to (i) a clear separation of concerns in the model, (ii) a simpler and more efficient implementation, (iii) a natural model extension for dealing also with reasoning over the burden of persuasion concepts. Moving [2, 3] from a semantics standpoint to a meta-level approach enable flexibility in terms of opening the model to further extensions – which can also be captured at the meta level, i.e., fully interoperable – and enable the natural exploitation of all the argumentative mechanisms [4] in a manner similar to the way in which argumentation systems can be expanded to include argumentation about priorities.

Our approach relies on the works from [5, 6] 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 [5].

Accordingly, Section 2 introduces the meta-argumentation framework, defining the object-level language and concepts, the meta-level language and concepts, and bimodal graphs as the model for dealing with their connection. Section 3 formally defines the meta-argumentation framework for the burden of persuasion introducing related argument schemes and discusses its equivalence with the model presented in [3]. Finally, Section 4 introduces a technological reification in Arg2P. 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 [6, 5]. Accordingly, Subsection 2.1 presents the object-level argumentation language exploited by our model, leveraging on an ASPIC⁺-like argumentation framework [7]. 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 [8] 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 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 and bp predicates are brought into relation through defeasible rules.

Definition 2.2 (Defeasible rule). A **defeasible rule** r has the form: $\rho : \phi_1, \dots, \phi_n, \sim \phi'_1, \dots, \sim \phi'_m \Rightarrow \psi$ with $0 \leq n, m$, and where

- ρ is the unique identifier for r , denoted by $N(r)$;
- each $\phi_1, \dots, \phi_n, \phi'_1, \dots, \phi'_m, \psi$ is a literal or a bp predicate;
- $\phi_1, \dots, \phi_n, \sim \phi'_1, \dots, \sim \phi'_m$ 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 [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 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 2.4 (Defeasible theory). A **defeasible theory** is a tuple $\langle \text{Rules}, \succ \rangle$ 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 [9, 10, 11].

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

- A is the argument's unique identifier;
- A_1, \dots, A_n are arguments constructed from the defeasible theory $\langle \text{Rules}, \succ \rangle$;
- ϕ is the conclusion of the argument, denoted by $\text{Conc}(A)$;

- $r : \text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi$ is the top rule of A , denoted by $\text{TopRule}(A)$.

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

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 2.6 (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 $\text{TopRule}(A) \succ \text{TopRule}(B)$.

Arguments are put in relation according to the attack relation.

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

- A undercuts B (on B') iff $\text{Conc}(A) = \neg N(\rho)$ for some $B' \in \text{Sub}(B)$, where ρ is $\text{TopRule}(B')$
- A rebuts B (on B') iff
 - $\text{Conc}(A) = \bar{\phi}$ for some $B' \in \text{Sub}(B)$ of the form $B''_1, \dots, B''_M \Rightarrow \phi$ and $B' \neq A$, or
 - $\text{Conc}(A) = \phi$ for some $B' \in \text{Sub}(B)$ such that $\sim \phi \in \text{Antecedent}(\text{TopRule}(B'))$

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 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 $\{\text{IN}, \text{OUT}, \text{UND}\}$ -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 $\{\text{IN}, \text{OUT}, \text{UND}\}$ -**labelling** L of G is a total function $\mathcal{A}_G \rightarrow \{\text{IN}, \text{OUT}, \text{UND}\}$. The set of all $\{\text{IN}, \text{OUT}, \text{UND}\}$ -**labellings** of G will be denoted as $\mathcal{L}(\{\text{IN}, \text{OUT}, \text{UND}\}, G)$.

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 $\mathcal{L}(\{\text{IN}, \text{OUT}, \text{UND}\}, G)$, 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 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. $\forall R \in \mathcal{R}_O \exists B \in \mathcal{A}_M : (B, R) \in \mathcal{S}_R$

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 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*

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

Consequently, an object argument may be present 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 [6] – we leverage on argument schemes [8]. In a few words, argumentation schemes are commonly used patterns of reasoning. They can be formalised in a rules-like form [12] 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 which 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, 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: $s : P_1, \dots, P_n, \sim P'_1, \dots, \sim P'_m \Rightarrow Q$ with $0 \leq n, m$, and where*

- *each $P_1, \dots, P_n, P'_1, \dots, P'_m \in \mathcal{M} \cup \mathcal{O}$, 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: $A : A_1, \dots, A_n \Rightarrow_s P$ 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, \dots, A_n are arguments constructed from S and G ;*

- P is the conclusion of the argument, denoted by $\text{Conc}(A)$;
- we denote with $\text{CQ}(A)$ 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 2.17 (Meta-Attack). *An argument A attacks argument B (on B') iff*

- $\text{Conc}(A) = \bar{P}$ for some $B' \in \text{Sub}(B)$ of the form $B''_1, \dots, B''_M \Rightarrow P$ or
- $\text{Conc}(A) = P$ for some $B' \in \text{Sub}(B)$ such that $\sim P \in \text{Antecedent}(\text{TopRule}(B'))$.

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, 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 $\text{Conc}(B) = bp(\text{Conc}(A))$. This notion is modelled through the following argument scheme:

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

$$\textit{Are arguments } A \textit{ and } B \textit{ provable?} \quad (\text{CQ}_{\text{S0}})$$

where $bp(\phi)$ is a predicate stating ϕ is a literal with the allocation of the burden, $\text{conclusion}(A, \phi)$ is a structural meta-predicate stating that $\text{Conc}(A) = \phi$ holds, and $\text{burdened}(A)$ 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 CQ_{S0} .

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

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

$$\textit{Is argument } A \textit{ provable? Is argument } B \textit{ really unprovable?} \quad (\text{CQ}_{\text{S1}})$$

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**).

$$\text{burdened}(A) \Rightarrow bp_met(A) \quad (\text{S2})$$

$$\text{burdened}(A) \Rightarrow \neg bp_met(A) \quad (\text{S3})$$

$$\textit{Is argument } A \textit{ admissible?} \quad (\text{CQ}_{\text{S2}})$$

$$\textit{Is argument } A \textit{ refuted or undecidable?} \quad (\text{CQ}_{\text{S3}})$$

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 **S2**, we have that only if exists a burden of persuasion on argument A , and A is admissible (CQ_{S2}), then the burden is satisfied.

On the other side, the validity of 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:

$$attack(B, A), \sim (\neg bp_met(A)), \sim (\neg bp_met(B)) \Rightarrow effectiveAttack(B, A) \quad (S4)$$

Can we prove that argument A or B are not failing to meet their burden? (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 (Figure 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 [13], 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.

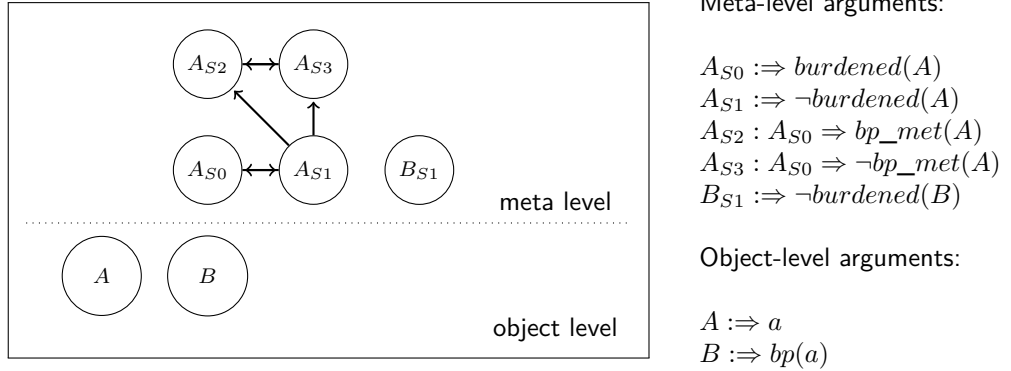


Figure 1: Object and meta level graphs from Example 1

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 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:

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

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:

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

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 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 ($CQ(A)$).*

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 [13]. 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 3.1. *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].*

Proof can be found in Appendix A.

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:

$$\begin{array}{lll}
e1 : ev1 & e2 : ev2 & e3 : ev3 \\
er1 : ev1 \Rightarrow \text{indiciaDiscrim} & er2 : ev2 \Rightarrow \neg \text{discrim} & er3 : ev3 \Rightarrow \text{discrim} \\
r1 : \text{indiciaDiscrim} \Rightarrow bp(\neg \text{discrim}) & &
\end{array}$$

We can then build the following object-level arguments:

$$\begin{array}{lll}
A_0 : \Rightarrow ev1 & B_0 : \Rightarrow ev2 & C_0 : \Rightarrow ev3 \\
A_1 : A_0 \Rightarrow \text{indiciaDiscrim} & B_1 : B_0 \Rightarrow \neg \text{discrim} & C_1 : C_0 \Rightarrow \text{discrim} \\
A_2 : A_1 \Rightarrow bp(\neg \text{discrim}) & &
\end{array}$$

and the following meta-level arguments:

$$\begin{array}{ll}
A_{0_{S1}} : \Rightarrow -\text{burdened}(A_0) & B_{0_{S1}} : \Rightarrow -\text{burdened}(B_0) \\
A_{1_{S1}} : \Rightarrow -\text{burdened}(A_1) & B_{1_{S0}} : \Rightarrow \text{burdened}(B_1) \\
A_{2_{S1}} : \Rightarrow -\text{burdened}(A_2) & B_{1_{S1}} : \Rightarrow -\text{burdened}(B_1) \\
C_{0_{S1}} : \Rightarrow -\text{burdened}(C_0) & B_{1_{S2}} : B_{1_{S0}} \Rightarrow bp_met(B_1) \\
C_{1_{S1}} : \Rightarrow -\text{burdened}(C_1) & B_{1_{S3}} : B_{1_{S0}} \Rightarrow \neg bp_met(B_1) \\
C_1 B_{1_{S4}} : \Rightarrow \text{effectiveAttack}(C_1, B_1) & B_1 C_{1_{S4}} : \Rightarrow \text{effectiveAttack}(B_1, C_1)
\end{array}$$

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 ($A_{0_{S1}}, A_{1_{S1}}, A_{2_{S1}}, B_{0_{S1}}, B_{1_{S0}}, C_{0_{S1}}, C_{1_{S1}}$) – where we can establish the absence of the burden for all of them –, and the rule $S4$ for every attack ($C_1 B_{1_{S4}}, B_1 C_{1_{S4}}$). 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. $\text{IN} = \{B_{1_{S1}}, C_1 B_{1_{S4}}, B_1 C_{1_{S4}}\}, \text{OUT} = \{B_{1_{S0}}, B_{1_{S2}}, B_{1_{S3}}\}, \text{UND} = \{\}$ —i.e., B_1 is not burdened;
2. $\text{IN} = \{B_{1_{S0}}, B_{1_{S2}}, C_1 B_{1_{S4}}, B_1 C_{1_{S4}}\}, \text{OUT} = \{B_{1_{S1}}, B_{1_{S3}}\}, \text{UND} = \{\}$ —i.e., B_1 is burdened and the burden is met;
3. $\text{IN} = \{B_{1_{S0}}, B_{1_{S3}}\}, \text{OUT} = \{B_{1_{S1}}, B_{1_{S2}}, C_1 B_{1_{S4}}, B_1 C_{1_{S4}}\}, \text{UND} = \{\}$ —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_{1S3} 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.

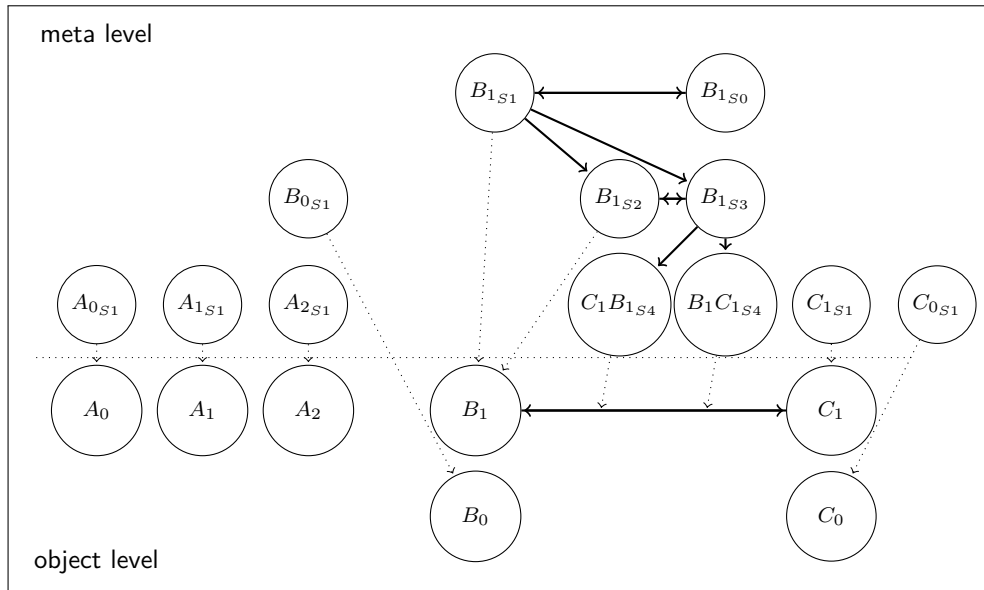


Figure 2: Argumentation graph (object- and meta- level) Example 2

4. Implementation in Arg2P

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 – it appears very inefficient from a computational perspective. Indeed the meta-level evaluation leads to a stable semantics computation, with a non-polynomial complexity [14]. For this reason, from a technological perspective, we reify the model presented in Section 3 into a more efficient resolution procedure.

Generally speaking, the stable semantics is exploited 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 consideration—the acceptable scenario is selected 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 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 the idea in [15]. Then, the graph is modified dynamically, leveraging the information on the burdened arguments. In a sense, we have a multi-stage evaluation that at every stage leads to the modification of the graph itself. First of all, burdened arguments are evaluated. Then, the original graph is modified to include the constraint (the burden). The entire process is so based on a grounded semantics—polynomial complexity [14]. The algorithm requires $m + 1$ evaluation stages to end, and, consequently, the final complexity is polynomial (where m is the number of burdened arguments).

Formally, 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 to meet the burden—expressed by S3 in the meta-model. A and B interaction is decided according to the A 's ability to satisfy the burden under grounded semantics:

- i)* iff A is IN, then the attack from A to B is introduced;
- ii)* vice versa, iff A is OUT or UND, the attack from B to A is introduced.

In other words, we are forcing the argument stating the burden (B) to defeat the burdened argument when the latter does not meet the burden (condition *ii*) and vice versa (condition *i*). Basically, through the first evaluation of the graph, the knowledge required to omit the contrast between arguments generated from schemes S3 and S2 is obtained—i.e. S2 becomes superfluous, leading to the possibility to avoid a stable semantics evaluation.

The algorithm has been tested and implemented in the Arg2P framework¹ [16, 17], Figure 4 shows the tool evaluation of the example discussed in Example 2.

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 [5, 6] introducing only the required abstraction at the meta level. In particular, [5] 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, [6] exploits bimodal graphs for dealing with arguments sources' trust. In [6] ASPIC+ is used instead of meta-ASPIC at the object level and on a set of meta-predicates related to

¹<http://arg2p.apice.unibo.it/>

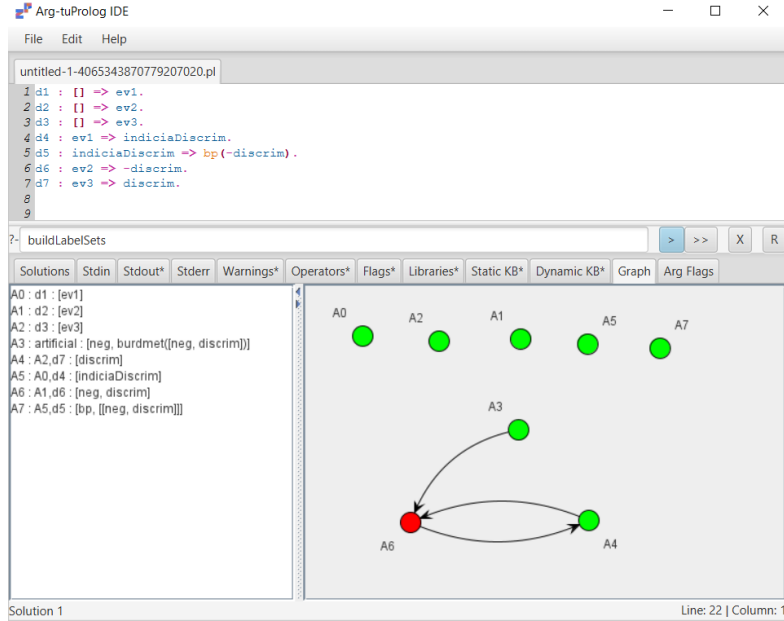


Figure 3: Arg2P evaluation of Example 2

the object level arguments and the schemes in the meta level, as in our approach. Both [5] and [6] use critical questions for managing attacks at the meta level.

Our framework and its model are based on these works with a clear definition of all the burdens abstractions at the meta-level. 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 nuance of burdens such as reasoning over the concept of the burden itself, thus leading to a full-fledged, interoperable framework open to further extensions. The feasibility of the model is tested by the implementation of Arg2P.

The model can be expanded in various ways; an open issue that we plan to address in future research concerns how to deal with defeat circles including burdened arguments. More generally, we plan to study 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.

Acknowledgments

The work has been supported by the “CompuLaw” project, funded by the European Re-

search Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 833647).

References

- [1] M. Wooldridge, P. McBurney, S. Parsons, On the meta-logic of arguments, in: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, Association for Computing Machinery, New York, NY, USA, 2005, pp. 560–567. doi:10.1145/1082473.1082558.
- [2] R. Calegari, G. Sartor, Burden of persuasion in argumentation, in: F. Ricca, A. Russo, S. Greco, N. Leone, A. Artikis, G. Friedrich, P. Fodor, A. Kimmig, F. Lisi, M. Maratea, A. Mileo, F. Riguzzi (Eds.), Proceedings 36th International Conference on Logic Programming (Technical Communications), ICLP 2020, volume 325 of *Electronic Proceedings in Theoretical Computer Science*, Open Publishing Association, Rende (CS), Italy, 2020, pp. 151–163. doi:10.4204/EPTCS.325.21.
- [3] R. Calegari, G. Sartor, A model for the burden of persuasion in argumentation, in: S. Villata, J. Harašta, P. Křemen (Eds.), Legal Knowledge and Information Systems. JURIX 2020: The Thirty-third Annual Conference, volume 334 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, Brno, Czech Republic, 2020, pp. 13–22. doi:10.3233/FAIA200845.
- [4] H. Prakken, C. Reed, D. N. Walton, Dialogues about the burden of proof, in: 10th International Conference on Artificial Intelligence and Law, ACM, Bologna, Italy, 2005, pp. 115–124. doi:10.1145/1165485.1165503.
- [5] J. Müller, A. Hunter, P. Taylor, Meta-level argumentation with argument schemes, in: International Conference on Scalable Uncertainty Management, volume 8078 of *Lecture Notes in Computer Science*, Springer, Springer, Washington, DC, USA, 2013, pp. 92–105. doi:10.1007/978-3-642-40381-1.
- [6] G. Ogunniye, A. Toniolo, N. 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*, Springer, Springer, Tokyo, Japan, 2018, pp. 585–593. doi:10.1007/978-3-030-03098-8.
- [7] H. Prakken, An abstract framework for argumentation with structured arguments, *Argument and Computation* 1 (2010) 93–124. doi:10.1080/19462160903564592.
- [8] D. Walton, C. Reed, F. Macagno, *Argumentation Schemes*, Cambridge University Press, United Kingdom, 2008. doi:10.1017/CB09780511802034.
- [9] S. Modgil, H. Prakken, The *ASPIC*⁺ framework for structured argumentation: a tutorial, *Argument & Computation* 5 (2014) 31–62. doi:10.1080/19462166.2013.869766.
- [10] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, *Artificial Intelligence* 171 (2007) 286–310. doi:10.1016/j.artint.2007.02.003.
- [11] G. Vreeswijk, Abstract argumentation systems, *Artificial Intelligence* 90 (1997) 225–279. doi:10.1016/S0004-3702(96)00041-0.

- [12] H. Prakken, Ai & law, logic and argument schemes, *Argumentation* 19 (2005) 303–320. doi:10.1007/s10503-005-4418-7.
- [13] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, *The knowledge engineering review* 26 (2011) 365–410. doi:10.1017/S0269888911000166.
- [14] M. Kröll, R. Pichler, S. 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*, ijcai.org, Melbourne, Australia, 2017, pp. 1145–1152. doi:10.24963/ijcai.2017/159.
- [15] G. Boella, D. M. Gabbay, L. van der Torre, S. Villata, Meta-argumentation modelling I: Methodology and techniques, *Studia Logica* 93 (2009) 297. doi:10.1007/s11225-009-9213-2.
- [16] G. Pisano, R. Calegari, A. Omicini, G. Sartor, Arg-tuProlog: A tuProlog-based argumentation framework, in: F. Calimeri, S. Perri, E. Zumpano (Eds.), *CILC 2020 – Italian Conference on Computational Logic. Proceedings of the 35th Italian Conference on Computational Logic*, volume 2719 of *CEUR Workshop Proceedings*, Sun SITE Central Europe, RWTH Aachen University, CEUR-WS, Aachen, Germany, 2020, pp. 51–66. URL: <http://ceur-ws.org/Vol-2710/paper4.pdf>.
- [17] G. Pisano, R. Calegari, A. Omicini, G. Sartor, A mechanism for reasoning over defeasible preferences in Arg2P, in: S. Monica, F. Bergenti (Eds.), *CILC 2021 – Italian Conference on Computational Logic. Proceedings of the 36th Italian Conference on Computational Logic*, volume 3002 of *CEUR Workshop Proceedings*, CEUR-WS, Parma, Italy, 2021, pp. 16–30. URL: <http://ceur-ws.org/Vol-3002/paper10.pdf>.

A. Proofs

Proof A.1. *The proof is straightforward. The burden of persuasion semantics acts like the grounded semantics, with the only difference that the burdened arguments that would have been UND for the latter are possibly 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 (S_1 or S_4). 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 will produce 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 admissible 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_{S_1} 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 object level.
- B is IN and A is OUT in L_G .
 - C2 If i), we have an inconsistency on CQ_{S1} —the burden is proved at object level and not at meta level.
 - C3 If ii), we have an inconsistency on CQ_{S2} since A is considered admissible 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 .²
 - B is IN and A is IN in L_G .
 - C4 If i), we have an inconsistency on CQ_{S1} —the burden is proved at object level and not at 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 an inconsistency on CQ_{S1} —the burden is proved at object level and not at meta level.
 - C7 If ii), we have an inconsistency since A is considered admissible 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 ²). 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 admissible w.r.t. critical questions. Moreover, the only admissible perspective coincides with the one generated from the bp-labelling in [3]—the burdened argument is labelled OUT in case of indecision (E4). Of course, the proof can be generalised to configurations taking into account any number of burdened arguments—where of course combinations grow exponentially with the number of burdened arguments.

²It 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