The Burden of Persuasion in Abstract Argumentation

Published Version:
The Burden of Persuasion in Abstract Argumentation / Kampik T.; Gabbay D.; Sartor G.. - ELETTRONICO. - 13040:(2021), pp. 224-243. (Intervento presentato al convegno 4th International Conference on Logic and Argumentation, CLAR 2021 tenutosi a Hangzhou, China nel 20-22 Ottobre 2021) [10.1007/978-3-030-89391-0_13].

This version is available at: https://hdl.handle.net/11585/873059 since: 2024-03-28

Published:

DOI: http://doi.org/10.1007/978-3-030-89391-0_13

The Burden of Persuasion in Abstract Argumentation

Timotheus Kampik[1], Dov Gabbay[2,3,4], and Giovanni Sartor[5,6]

[1] Umeå University, Sweden
[2] University of Luxembourg, Luxembourg
[3] King's College London, United Kingdom
[4] Bar Ilan University, Israel
[5] Università di Bologna, Italy
[6] European University Institute, Italy
tkampik@cs.umu.se, dov.gabbay@kcl.ac.uk, giovanni.sartor@eui.eu

Abstract. In this paper, we provide a formal framework for modeling the *burden of persuasion* in legal reasoning. The framework is based on abstract argumentation, a frequently studied method of non-monotonic reasoning, and can be applied to different argumentation semantics; it supports burdens of persuasion with arbitrary many levels, and allows for the placement of a burden of persuasion on any subset of an argumentation framework's arguments. Our framework can be considered an extension of related works that raise questions on how burdens of persuasion should be handled in some conflict scenarios that can be modeled with abstract argumentation. An open source software implementation of the introduced formal notions is available as an extension of an argumentation reasoning library.

Keywords: Formal Argumentation · Non-monotonic Reasoning · Legal Reasoning.

1 Introduction

Over the past decades, formal argumentation has emerged as a promising collection of methods for reasoning under uncertainty [4]. A particularly relevant application domain that can benefit from argumentation-based models of conflicts and contradictions is legal reasoning [9]. An important notion in legal argumentation – but also in other domains in which an outcome has to be reached under time and resource constraints, such as political debates – is the *burden of persuasion* [20]. By saying that an argument is burdened with persuasion we mean that the argument only is relevant when it is convincing, *i.e.* when it overcomes all relevant objections against it. If this is not the case, the argument has to be rejected for failing to meet its burden of persuasion. In an argumentation-based theory, the burden of persuasion may be placed on some of the arguments in the theory. Roughly speaking, if there are several conflicting conclusions (here and henceforth referred to as *extensions* to align with formal argumentation terminology), we can infer from the theory (considering constraints imposed by a basic inference function), the burden of persuasion dictates that we must be *less skeptical* towards unburdened arguments than towards burdened ones. If we are faced with conflicting extensions, one being only supported by burdened arguments and one being only supported by unburdened arguments, we select the latter. Moreover, any successful attacks against a burdened argument entail that the burdened argument is to be rejected[7]. In a recent paper, Calegari *et al.* present a model of the burden of persuasion that is based on a structured argumentation approach [11]; in their paper, the authors also highlight some limitations of their model, such as the inability to meaningfully model burdened arguments that are part of cyclic structures. This paper aims to address these limitations by introducing a model of the burden of persuasion that only relies on abstract argumentation and supports any abstract argumentation framework (where the burden of persuasion may be placed on any subset of the argumentation framework's arguments), as well as arbitrary many *levels of burdens*.

Let us introduce an example that gives an intuition of our approach.

Example 1. Usually patients have the burden of persuasion on the liability of medical doctors in order to be compensated for the harm they suffered as a consequence of an unsuccessful treatment. This follows from the general principle that the plaintiffs in a legal case should persuade the judge in order to get a favorable decision. Should the outcome remain uncertain, their claim has to be rejected. However, doctors do not have to pay compensation in case they were diligent in treating the patient and the failure of the treatment was not due to incompetence or carelessness. The possibility of doctors to avoid liability is limited by the fact that – at least in some legal systems – they have the burden of persuasion with regard to their diligence. Their arguments to this effect must be convincing. Otherwise they will be rejected: in case uncertainty remains on whether they were diligent or not, their liability will consequently be established. Note that this is a simplified representation of the matter at stake, since other aspects of the case may have to be considered, such as the difficult or extraordinary nature of the case of the patient.

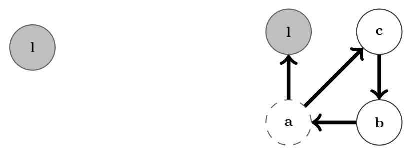
Let us assume however, that under the given normative framework a patient asks for compensation. The patient's argument l for the doctors' liability is based on the fact that the doctor subjected him to an unsuccessful and harmful therapy. Argument l is attacked by an expert witness in favor of the doctor, whose argument a claims that the doctor was diligent, since the adopted therapy is successful in the vast majority of cases; this was argued in a leading top scientific journal, the evidence of this journal being sufficient to guarantee the truth of the claim. The patient's expert witness attacks argument a through argument b, according to which a therapy with a higher success rate is available. The high success rate of the adopted treatment is insufficient to establish diligence, if an even more effective treatment is state-of-the art. The Court's expert witness attacks argument b through one further argument c, according to which the scientific evidence in favor of b is insufficient, being based on a restricted set of

[7] Here, we assume a model where an argument is either burdened or unburdened.

the scientific literature. Finally, argument c is attacked by argument a, which includes the claim that one single journal was sufficient to establish a scientific claim.

We end up with the following *argumentation framework* – a tuple consisting of a set of *arguments* AR and a set of *attacks* $AT \subseteq AR \times AR$ (Figure 1):

$$AF' = (AR', AT') = (\{l, a, b, c\}, \{(a, c), (a, l), (b, a), (c, b)\})$$



(a) AF. (b) AF'.

Fig. 1: We restrict AF' to $\{l\}$, generating AF, to reflect that the burden of persuasion rests on the rejection of l. Then, we infer $\{l\}$ from AF and check if we can infer an extension that entails $\{l\}$ from AF'. Since this is the case, we have to consider $\{l\}$ as valid. In the example, arguments with a gray background are unambiguously inferred; arguments with a white background and a solid border may be inferred (are part of at least one extension, considering the burden of persuasion approach); arguments with a dashed border are unambiguously rejected.

Intuitively, it is not clear which of the arguments are *valid* in this framework, so that their conclusion (extension) has to be endorsed, and in particular whether l is valid or not. As noted above, the patient should have the burden of persuasion on liability, but the doctor has the burden of persuasion on her diligence. We assume that it is uncontroversial that the patient has been harmed by the wrong therapy: there is no doubt that the patient has satisfied his burden of persuasion on this point. The issue is whether the doctor has satisfied her burden of persuasion relative to her diligence. She has no benefit of doubt in this regard: in case doubts remain on her diligence, her argument has to be rejected, and so her liability toward the patient will have to be established. The crucial point is then to establish whether there is doubt on her diligence based on the circle of arguments $\{a, b, c\}$.

Hence, we generate the following *argumentation framework sequence* from AF': $AFS = \langle AF, AF' \rangle$, where $AF = (\{l\}, \{\})$; we call AF the *restriction of* AF' *to* $\{l\}$. We first determine all possible extensions of AF, and trivially, there is only one, which is $\{l\}$. Then, we determine all extensions of AF'. Here, we have different options.

1. Assuming that the cycle of arguments "a attacks c attacks b attacks a" is a self-contradiction, we can say that the only extension is the empty set;

the traditional abstract argumentation semantics as introduced in Dung's seminal paper [14] behave accordingly. However, from a legal reasoning perspective, we need to employ a more credulous approach.

2. Again considering the cycle of arguments "a attacks c attacks b attacks c" as a self-contradiction, we can discard the arguments in this cycle, but then conclude that surely, l cannot be rejected; the recently introduced weak admissible set-based argumentation semantics family [8] formalizes this intuition, and allows us to again infer $\{l\}$ as the only extension. This result is aligned with common legal notions of the burden of persuasion *in our case*, because the practitioner's diligence is not beyond doubt[8].
3. We can assume that any of the arguments a, b, or c could be part of an extension, but that these three arguments are mutually exclusive, and hence infer that $\{a\}$, $\{b, l\}$ and $\{c, l\}$ are extensions. This intuition is formalized (for example) by CF2 [6] and SCF2 [12] semantics; not all extensions reject l; hence, the notion of the burden of persuasion constrains us to select one of the extensions that entail l, *i.e.* either $\{b, l\}$ or $\{c, l\}$. This means we have to accept l and we conclude that the doctor has not successfully persuaded the court that she has acted without negligence.

Let us highlight that our framework for modeling the burden of persuasion is not merely determining whether a set of arguments is *credulously accepted* – whether it is entailed by at least one extension – or *skeptically accepted*, *i.e.* whether it is entailed by all extensions. For this, we introduce an additional (abstract) example, which also illustrates how we can manage multiple levels of the burden of persuasion.

Example 2. Consider the following argumentation framework:

$$AF'' = (\{a, b, c, d, e\}, \{(a, b), (a, e), (b, a), (b, e), (c, d), (d, c), (e, a), (e, b)\})$$

and the following burdens of persuasion: *i)* a and b are unburdened; *ii)* c is burdened with a "light-weight" level 1 burden; *iii)* d and e are burdened with a "heavier" level 2 burden. Let us assume a credulous inference function allows for the following extensions[9] (given only AF'' and no burden of persuasion model):

$$\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{e, c\}, \{e, d\}$$

We take a look at the unburdened arguments and their attacks among each other, which gives us the argumentation framework $AF = (\{a, b\}, \{(a, b), (b, a)\})$. We

[8] For the sake of conciseness, we do not consider weak admissibility-based semantics in detail. However, let us claim that the simple example $AF = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$ illustrates that all weak admissible set-based semantics Baumann *et al.* may not be sufficiently credulous for many applications that require a model of the burden of persuasion.

[9] In this example, the inferences we draw from the abstract argumentation frameworks coincide, for example, with the *extensions* (sets of arguments) returned by CF2 [6] and SCF2 [12] semantics.

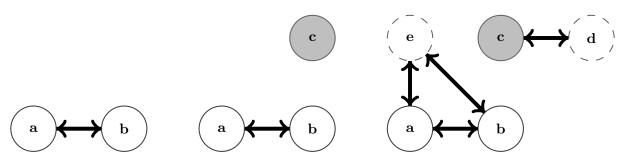
assume that from AF, we can infer either $\{a\}$ or $\{b\}$. This means that we need to consider all extensions that can be inferred from AF'', given they entail either $\{a\}$ or $\{b\}$. We "filter" the extensions accordingly and remain with the following sets[10]:

$$\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}$$

Now, we consider the arguments that carry the first-level burden of persuasion, *i.e.* $\{c\}$ and $AF' = (\{a, b, c\}, \{(a, b), (b, a)\})$. Because of the unburdened arguments, we have to be able to infer either $\{a\}$ or $\{b\}$. But surely, we can allow for this inference and still guarantee that we can infer $\{c\}$: we merely need to remove the extensions $\{a, d\}$ and $\{b, d\}$:

$$\{a, c\}, \{b, c\}$$

It follows that $\{a, c\}$ and $\{b, c\}$ are our final extensions; the arguments that carry the second-level burden of persuasion – d and e – are rejected. No unambiguous conclusion can be reached, as our final inference result is "either $\{a, c\}$ or $\{b, c\}$".



(a) AF. (b) AF'. (c) AF''.

Fig. 2: Multiple levels of burdens of persuasion.

A software implementation of the formal concepts we introduce in this paper is available at `https://git.io/JGueN`. The implementation relies on the abstract argumentation reasoner provided by the *Tweety* project [21].

The rest of this paper is organized as follows. Section 2 provides relevant theoretical preliminaries. Then, Section 3 introduces our formal framework for modeling the burden of persuasion in abstract argumentation. The suitability of applying different argumentation semantics, as well as the relevance of skeptical acceptance are discussed in Section 4. Finally, Section 5 discusses the framework in the context of related research, before Section 6 concludes the paper.

2 Preliminaries

This section introduces the preliminaries that our work is based upon. The central notion this paper uses is Dung's (abstract) argumentation framework [14].

[10] Let us note that there are some intricate details in the filtering approach that this example does not cover.

An argumentation framework AF is a tuple (AR, AT), such that AR is a set of *arguments* and AT is a set of *attacks*, $AT \subseteq AR \times AR$. We assume that the set of arguments in an argumentation framework is finite. For $(a, b) \in AT$, we say that "a attacks b". For $S \subseteq AR$, $b \in S$, and $a \in AR$, iff $(b, a) \in AT$, we say that "S attacks a" and iff $(a, b) \in AT$, we say that "a attacks S"; we denote $\{a | a \in AR, a \text{ attacks } S\}$ by S^- and $\{b | b \in AR, S \text{ attacks } b\}$ by S^+. For $S \subseteq AR$, $P \subseteq AR$ such that $\exists (a, b) \in AT, a \in S, b \in P$, we say that "$S$ attacks P". For $S \subseteq AR, a \in AR$, we say that "S defends a" iff $\forall b \in AR$, such that b attacks a it holds true that S attacks b. Given $S \subseteq AR$, we define $AF \downarrow_S = (S, AT \cap S \times S)$. We call $AF \downarrow_S$ the *restriction of AF to S*. Let us introduce some properties of sets of arguments in an argumentation framework.

Definition 1 (Conflict-free, Unattacked, and Admissible Sets [3]). *Let $AF = (AR, AT)$ be an argumentation framework. A set $S \subseteq AR$:* i) *is* conflict-free *iff $\nexists a, b \in S$ such that a attacks b;* ii) *is* unattacked *iff $\nexists a \in AR \setminus S$ such that a attacks S;* iii) *is* admissible *iff S is conflict-free and $\forall a \in S$, it holds true that S defends a.*

Argumentation framework expansions model the addition of new arguments and attacks to an argumentation framework.

Definition 2 (Argumentation Framework Expansions [7]). *Let $AF = (AR, AT)$ and $AF' = (AR', AT')$ be argumentation frameworks. AF' is an* expansion *of AF (denoted by $AF \preceq_E AF'$) iff $AR \subseteq AR'$ and $AT \subseteq AT'$. AF' is a* normal expansion *of AF (denoted by $AF \preceq_N AF'$) iff $AF \preceq_E AF'$ and $(AR \times AR) \cap (AT' \setminus AT) = \{\}$.*

While our formal framework does not rely on expansions or normal expansions, these notions can be used to establish the connection between our work and the research direction of *dynamics* in formal argumentation (see Section 5).

An argumentation semantics σ takes an argumentation framework as its input and determines sets of arguments (*extensions*) that can be considered valid conclusions. Dung's seminal paper introduces stable, preferred, complete, and grounded argumentation semantics.

Definition 3 (Dung's Argumentation Semantics [14]). *Let $AF = (AR, AT)$ be an argumentation framework. An* admissible *set $S \subseteq AR$ is a:*

- stable extension *of AF iff S attacks each argument that does not belong to S. $\sigma_{st}(AF)$ denotes all stable extensions of AF.*
- preferred extension *of AF iff S is a maximal (w.r.t. set inclusion) admissible subset of AR. $\sigma_{pr}(AF)$ denotes all preferred extensions of AF.*
- complete extension *of AF iff each argument that is defended by S belongs to S. $\sigma_{co}(AF)$ denotes all complete extensions of AF.*
- grounded extension *of AF iff S is the minimal (w.r.t. set inclusion) complete extension of AF. $\sigma_{gr}(AF)$ denotes all grounded extensions of AF.*

Given any argumentation semantics σ and any argumentation framework AF, we call a set $S \in \sigma(AF)$ a σ-extension of AF. If and only if for every argumentation framework AF it holds true that $|\sigma(AF)| \geq 1$ we say that σ is universally defined; if and only if for every argumentation framework AF it holds true that $|\sigma(AF)| = 1$ we say that σ is universally uniquely defined. Dung's semantics are all based on the notion of an admissible set. Later works introduce semantics based on naive ($\subseteq$-maximal conflict-free) sets.

Definition 4 (Naive and Stage Semantics [23]). *Let $AF = (AR, AT)$ be an argumentation framework and let $S \subseteq AR$.*

- *S is a naive extension of AF iff S is a maximal conflict-free subset of AR w.r.t. set inclusion. $\sigma_{naive}(AF)$ denotes all naive extensions of AF.*
- *S is a stage extension of AF iff S is conflict-free and $S \cup S^+$ is maximal w.r.t. set inclusion,* i.e. *$\nexists S' \subseteq AR$, such that S' is a conflict-free set and $S \cup S^+ \subset S' \cup S'^+$. $\sigma_{stage}(AF)$ denotes the stage extensions of AF.*

Given an argumentation framework AF and an argumentation semantics σ, the skeptically accepted set of arguments is the intersection of the σ-extensions of AF.

Definition 5 (Skeptical Acceptance). *Let $AF = (AR, AT)$ be an argumentation framework and let σ be an argumentation semantics. We call $\bigcap_{E \in \sigma(AF)} E$ the* skeptically accepted set of arguments of AF given σ *and denote it by $\sigma^{\cap}(AF)$.*

Let us introduce some preliminaries for so-called *SCC-recursive semantics*, starting with the notion of a path between arguments.

Definition 6 (Path between Arguments). *Let $AF = (AR, AT)$ be an argumentation framework. A path from an argument $a_0 \in AR$ to another argument $a_n \in AR$ is a sequence of arguments $P_{a_0,a_n} = \langle a_0, ..., a_n \rangle$, such that for $0 \leq i < n$, a_i attacks a_{i+1}.*

Based on this definition, we can define the notion of reachability.

Definition 7 (Reachability). *Let $AF = (AR, AT)$ be an argumentation framework. We say that given two arguments $a, b \in AR$, "b is reachable from a" iff there exists a path $P_{a,b}$ or $a = b$.*

Based on the notion of reachability, we can define *strongly connected components.*

Definition 8 (Strongly Connected Components (SCC)). *Let $AF = (AR, AT)$ be an argumentation framework. $S \subseteq AR$ is a strongly connected component of AF iff $\forall a, b \in S$, a is reachable from b and b is reachable from a and $\nexists c \in AR \setminus S$, such that a is reachable from c and c is reachable from a. Let us denote the strongly connected components of AF by $SCCS(AF)$.*

Another preliminary for SCC-recursive semantics is the UP function.

Definition 9 (UP [6]). *Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$, $S \subseteq AR$. We define $UP_{AF}(S, E) = \{a | a \in S, \nexists b \in E \setminus S \text{ such that } (b, a) \in AT\}$.*

Now, we can introduce the SCC-recursive and naive set-based CF2 semantics.

Definition 10 (CF2 Semantics [6]). *Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$. E is a CF2 extension iff:*

- *E is a naive extension of AF if $|SCCS(AF)| = 1$;*
- *$\forall S \in SCCS(AF)$, $(E \cap S)$ is a CF2 extension of $AF \downarrow_{UP_{AF}(S,E)}$, otherwise.*

$\sigma_{CF2}(AF)$ denotes all CF2 extensions of AF.

To give a rough intuition of how SCC-recursive semantics (and in particular: CF2 semantics) work, let us introduce an example.

Example 3. Consider $AF = (\{a, b, c\}, \{(a, b), (b, a), (a, c), (b, c)\})$. We have two SCCs: $\{a, b\}$ and $\{c\}$. Colloquially speaking, we traverse the SCC graph, starting with unattacked ("top-level") SCCs: first, we take the top-level SCC $\{a, b\}$ and determine $\sigma_{naive}(AF \downarrow_{\{a,b\}}) = \{\{a\}, \{b\}\}$. Then, $\forall E \in \{\{a\}, \{b\}\}$, we determine $UP_{AF}(S, E)$, where $S = \{c\}$, because $\{c\}$ is the "next" and only remaining SCC. Because
$UP_{AF}(\{c\}, \{a\}) = UP_{AF}(\{c\}, \{b\}) = \{\}$ and $\sigma_{naive}((\{\}, \{\})) = \{\{\}\}$, we remain with $\{a\}$ and $\{b\}$ as our CF2 extensions.

Stage2 is an SCC-recursive semantics that has been introduced to address some shortcomings of CF2 semantics, notably unintuitive behavior when resolving even-length cycles of length ≥ 6, roughly speaking (see Example 4, argumentation framework AF^{**}).

Definition 11 (Stage2 Semantics [15]). *Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$. E is a stage2 extension iff:*

- *E is a stage extension of AF if $|SCCS(AF)| = 1$;*
- *$\forall S \in SCCS(AF)$, $(E \cap S)$ is a stage2 extension of $AF \downarrow_{UP_{AF}(S,E)}$, otherwise.*

$\sigma_{stage2}(AF)$ denotes all stage2 extensions of AF.

Another "CF2 improvement attempt" is made by Cramer's and Van der Torre's SCF2 semantics [12]. The authors start by defining a notion that ignores self-attacking arguments.

Definition 12 (nsa(AF) [12]). *Let $AF = (AR, AT)$ be an argumentation framework. We define $nsa(AF) = AF \downarrow_{\{a | a \in AR \text{ and } (a,a) \notin AT\}}$.*

Based on this notion, Cramer and Van der Torre introduce nsa(CF2) semantics as an intermediate step on the way to SCF2 semantics.

Definition 13 (nsa(CF2) Semantics [12]). *Let $AF = (AR, AT)$ be an argumentation framework. A set $E \subseteq AR$ is an $nsa(CF2)$-extension of AF iff $E \in \sigma_{CF2}(nsa(AF))$. $\sigma_{nsa(CF2)}(AF)$ denotes all nsa(CF2) extensions of AF.*

This approach fixes some issues with CF2 semantics and self-attacking arguments. To tackle the problem with even-length cycles, we need to define some preliminaries.

Definition 14 (Attack Cycles). *Let $AF = (AR, AT)$ be an argumentation framework. An attack cycle C is a sequence of arguments $\langle a_0, ..., a_n \rangle$ where $(a_i, a_{i+1}) \in AT$ for $0 \leq i < n$ and $a_j \neq a_k$ for $0 \leq j < k \leq n$ if not $j = 0$ and $k = n$, and where $a_0 = a_n$. An attack cycle is* odd *iff n is odd and* even *iff n is even.*

Cramer and Van der Torre introduce a specific property to describe how a CF2-like semantics should ideally behave in the case of even cycles that are not "affected" by odd cycles, roughly speaking.

Definition 15 (Strong Completeness Outside Odd Cycles (Set) [12]). *Let $AF = (AR, AT)$ be an argumentation framework. A set $S \subseteq AR$ is strongly complete outside odd cycles iff $\forall a \in AR$, if no argument in $\{a\} \cup \{a\}^-$ is in an odd attack cycle and $S \cap \{a\}^- = \{\}$ then $a \in S$.*

To systematically analyze argumentation semantics, a range of formal argumentation principles have been defined [5,22]. Cramer and Van der Torre turn the *strong completeness outside odd cycles* property into a principle to "catch" unintuitive CF2 behavior.

Definition 16 (SCOOC Principle [12]). *An argumentation semantics σ is Strongly Complete Outside Odd Cycles (SCOOC) iff for every argumentation framework $AF, \forall E \in \sigma(AF), E$ is strongly complete outside odd cycles.*

Based on this principle and the notion of $nsa(CF2)$ semantics, SCF2 semantics is defined.

Definition 17 (SCF2 Semantics [12]). *Let $AF = (AR, AT)$ be an argumentation framework and let E be a set such that $E \subseteq AR$. E is an SCF2 extension iff:*

- *E is a naive extension of $nsa(AF)$ and E is strongly complete outside odd cycles if $|SCCS(nsa(AF))| = 1$;*
- *$\forall S \in SCCS(nsa(AF))$, $(E \cap S)$ is an SCF2 extension of $AF \downarrow_{UP_{nsa(AF)}(S,E)}$, otherwise.*

$\sigma_{SCF2}(AF)$ denotes all SCF2 extensions of AF.

Let us introduce some examples that illustrate the behaviors of – and highlights the difference between – stage, CF2, stage2, and SCF2 semantics. However, let us note that a detailed explanation of the semantics is beyond the scope of this paper and the reader may consult the original works instead.

Example 4. Let us consider the following argumentation frameworks: *i)* $AF' = (\{a, b, c\}, \{(a, b), (b, c), (c, c)\})$; *ii)* $AF'' = (\{a, b, c\}, \{(a, b), (a, c), (b, c), (c, a)\})$; *iii)* $AF^* = (\{a, b, c\}, \{(a, b), (b, c), (c, a), (c, c)\})$; *iv)* $AF^{**} = (\{a, b, c, d, e, f\}, \{(a, b), (b, c), (c, d), (d, e), (e, f), (f, a)\})$. Table 1 displays the extensions stage, CF2, stage2, and SCF2 semantics yield for these argumentation frameworks.

Argumentation principles that are relevant in the context of this paper are the admissibility and naivety principles.

	stage	**CF2**	**stage2**	**SCF2**
AF'	$\{a\}, \{b\}$	$\{a\}$	$\{a\}$	$\{a\}$
AF''	$\{a\}$	$\{a\}, \{b\}, \{c\}$	$\{a\}$	$\{a\}, \{b\}, \{c\}$
AF^*	$\{a\}, \{b\}$	$\{a\}, \{b\}$	$\{a\}, \{b\}$	{a}
AF^{**}	$\{a, c, e\}$, $\{b, d, f\}$	$\{a, c, e\}$, $\{b, d, f\}$ $\{a, d\}$, $\{b, e\}$, $\{c, f\}$	$\{a, c, e\}$, $\{b, d, f\}$	$\{a, c, e\}$, $\{b, d, f\}$

Table 1: Differences between stage, CF2, stage2, and SCF2 semantics (examples).

Definition 18 (Admissibility and Naivety Principles [5]). *Let σ be an argumentation semantics. σ satisfies the admissibility principle iff for every argumentation framework $AF = (AR, AT)$, $\forall E \in \sigma(AF), E$ is an admissible set. σ satisfies the naivety principle iff for every argumentation framework $AF = (AR, AT)$, $\forall E \in \sigma(AF), E$ is a maximal conflict-free subset (w.r.t. set inclusion) of AR.*

3 An Abstract Argumentation-based Burden of Persuasion

In this section, we introduce our formal framework for modeling burdens of persuasion in abstract argumentation.

Definition 19 (Burden of Persuasion-Framework (BPF)). *A Burden of Persuasion Framework (BPF) is a tuple $AF_{BP} = (ARS, AT)$, where:*

- *$ARS = \langle S_0, ..., S_n \rangle$ and each $S_i, 0 \leq i \leq n$ is a non-empty set of arguments, such that for each $S_j, 0 \leq j \leq n, i \neq j$, it holds true that $S_i \cap S_j = \{\}$;*
- *We denote $\bigcup_{0 \leq k \leq n} S_k$ by $ARGS(ARS)$;*
- *$AT \subseteq ARGS(ARS) \times ARGS(ARS)$.*

We assume that given a BPF $AF_{BPF}(ARS, AT)$, $ARGS(ARS)$ is finite. Let us introduce some short-hand notation that makes it easier to work with BPFs.

Definition 20 (BPF Short-hand Notation). *Let $AF_{BP} = (ARS, AT)$ be a BPF, such that $ARS = \langle S_0, ..., S_n \rangle$. Given $0 \leq i \leq n$, we denote $\bigcup_{0 \leq j \leq i} S_j$ by AR_i and $(AR_i, AT \cap (AR_i \times AR_i))$ by AF_i. Also, for any $AF_{BP} = (ARS, AT)$, such that $ARS = \langle S_0, ..., S_n \rangle$, we denote:*

$$AF_{BP-1} = \begin{cases} AF_{BP} & \text{if } n = 0; \\ (\langle S_0 \cup S_n \rangle, AT) & \text{if } n = 1; \\ (\langle S_0, ..., S_{n-2}, S_{n-1} \cup S_n \rangle, AT) & \text{otherwise.} \end{cases}$$

For a set of arguments $S \subseteq S_0$ we say that S is unburdened *and for any argument $a \in S_0$ we say that a is unburdened. For a set of arguments $S' \subseteq S_k, 0 < k \leq n$, we say that S' is burdened or that S' is level k-burdened, and for an argument $a' \in S_k$ we say that a' is burdened or that a' is level k-burdened.*

Let us introduce an example of a BPF.

Example 5. Consider Example 2. When modeling the argumentation frameworks that we have in the example as a BPF, we get:

- $AF_{BP} = (\langle\{a,b\},\{c\},\{d,e\}\rangle, \{(a,b),(a,e),(b,a),(b,e),(c,d),(d,c),$ $(e,a),(e,b)\})$;
- $AF_2 = (\{a,b,c,d,e\}, \{(a,b),(a,e),(b,a),(b,e),(c,d),(d,c),(e,a),(e,b)\})$;
- $AF_1 = (\{a,b,c\}, \{a,b),(b,a)\})$;
- $AF_0 = (\{a,b\}, \{(a,b),(b,a)\})$;
- $AF_{BP-1} = (\langle\{a,b\},\{c,d,e\}\rangle, \{(a,b),(a,e),(b,a),(b,e),(c,d),(d,c),(e,a),$ $(e,b)\})$.

The set of arguments $\{a,b\}$ is unburdened, $\{c\}$ is level 1-burdened and $\{d,e\}$ is level 2-burdened.

Before we can define a way to determine the extensions of BPFs, let us introduce the notion of $\subseteq$-maximal monotonic extensions.

Definition 21 ($\subseteq$-Maximal Monotonic Extensions). *Let AR and A be finite sets of arguments (extensions) and let $EXTS \subseteq 2^{AR}$ and $ES \subseteq 2^{A}$. We define the $\subseteq$-maximal monotonic extensions of $EXTS$ w.r.t. ES, denoted by $EXTS^{\subseteq-max}_{mon}(EXTS, ES)$, as follows:*

$$EXTS^{\subseteq-max}_{mon}(EXTS, ES) = \{E \mid E \in EXTS, \exists S \in ES \text{ such that } \forall E' \in EXTS, E' \cap S \subseteq E \cap S\}$$

Let us highlight that the notion of $\subseteq$-maximal monotonic extensions is purposefully different from the cardinality-based monotony measure and optimization approach [19] that we have recently introduced. Colloquially speaking, we can say that the $\subseteq$-maximal approach is more credulous. As an example, consider the argumentation frameworks $AF = (\{a,b,c\},\{\})$ and $AF' = (\{a,b,c,d,e\},$ $\{(d,a),(d,e),(e,b),(e,c),(e,d)\})$ and preferred semantics. $\sigma_{pr}(AF) = \{\{a,b,c\}\}$; the only cardinality-maximal monotonic extension of $\sigma_{pr}(AF')$ w.r.t. to $\{\{a,b,c\}\}$ is $\{b,d,c\}$, whereas we have two $\subseteq$-maximal monotonic extensions of $\sigma_{pr}(AF')$ w.r.t. to $\{\{a,b,c\}\}$, *i.e.* $\{b,d,c\}$ and $\{a,e\}$. Hence, $\subseteq$-maximal monotonic extensions are better aligned with the notion of the *burden of persuasion* in legal reasoning: intuitively, we cannot eliminate doubt in this scenario. However, we want to avoid the inclusion of extensions that are not Pareto optimal. Let us provide an example to illustrate this problem.

Example 6. Consider $EXTS = \{\{a,b\},\{\}\}$ and $ES = \{\{a\},\{c\}\}$. $EXTS^{\subseteq-max}_{mon}(EXTS, ES) = \{\{a,b\},\{\}\}$. However, intuitively, it makes sense to "drop" $\{\}$, because its absence does not affect the fact that c is not entailed by any set of arguments in $EXTS$, but its presence implies that we *may* select a set of arguments from $EXTS$ that does not entail a.

To address this issue, we define *Pareto optimal* $\subseteq$-maximal monotonic extensions.

Definition 22 (Pareto Optimal $\subseteq$-Maximal Monotonic Extensions). *Let AR and A be finite sets of arguments (extensions), let $EXTS \subseteq 2^{AR}$ and $ES \subseteq 2^{A}$. We define the Pareto optimal $\subseteq$-maximal monotonic extensions of $EXTS$ w.r.t. ES, denoted by $EXTS^{\subseteq-max}_{po-mon}(EXTS, ES)$, as follows:*

$$
\begin{aligned}
&EXTS^{\subseteq-max}_{po-mon}(EXTS, ES) = \{E|E \in EXTS \textit{ and} \\
&\quad \nexists E' \in EXTS, \textit{such that} \\
&\quad\quad \forall S \in ES, S \cap E \subseteq S \cap E' \textit{ and} \\
&\quad\quad \exists S' \in ES, \textit{ such that } S' \cap E \subset S' \cap E'\}
\end{aligned}
$$

Let us continue the previous example to illustrate the difference between the previous two definitions.

Example 7. Consider again $EXTS = \{\{a, b\}, \{\}\}$ and $ES = \{\{a\}, \{c\}\}$. $EXTS^{\subseteq-max}_{po-mon}(EXTS, ES) = \{\{a, b\}\}$.

Now, let us define a way to determine the extension of a BPF, given any universally defined argumentation semantics.

Definition 23 (BP Semantics and Extensions). *Let $AF_{BP} = (ARS, AT)$ be a BPF, such that $ARS = \langle S_0, ..., S_n\rangle$, and let σ be an argumentation semantics. We define the σ-extensions of AF_{BP} as returned by the BP semantics σ^{BP}, denoted by $\sigma^{BP}(AF_{BP})$, as follows:*

$$
\sigma^{BP}(AF_{BP}) =
\begin{cases}
\sigma(AF_0) & \textit{if } n = 0; \\
EXTS^{\subseteq-max}_{po-mon}(\sigma^{BP}(AF_{BP-1}), \sigma(AF_0) \cup ... \cup \sigma(AF_{n-1})) & \textit{otherwise.}
\end{cases}
$$

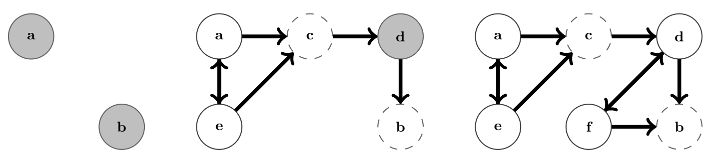
Let us provide an example of how BPF extensions are determined.

Example 8. Consider the BPF $AF_{BP} = (ARS, AT) = (\langle\{a, b\}, \{c, d, e\}, \{f\}\rangle, \{(a, c), (a, e), (c, d), (d, b), (d, f), (e, a), (e, c), (f, b), (f, d)\})$. Let us assume we apply SCF2 semantics[11] and first provide an intuition that strays from the recursive definition (Definition 23). Based on AF_{BP}, we generate the following argumentation frameworks: $AF_0 = (\{a, b\}, \{\})$; $AF_1 = (\{a, b, c, d, e\}, \{(a, c), (a, e), (c, d), (d, b), (e, a), (e, c)\})$; $AF_2 = (\{a, b, c, d, e, f\}, \{(a, c), (a, e), (c, d), (d, b), (d, f), (e, a), (e, c), (f, b), (f, d)\})$. Figure 3 depicts AF_0, AF_1, and AF_2. Then, we determine the CF2 extensions of AF_2 and AF_0: $\sigma_{SCF2}(AF_2) = \{\{a, d\}, \{a, f\}, \{e, d\}, \{e, f\}\}$ and $\sigma_{SCF2}(AF_0) = \{\{a, b\}\}$. $EXTS^{\subseteq-max}_{po-mon}(\sigma_{SCF2}(AF_2), \sigma_{SCF2}(AF_0)) = \{\{a, d\}, \{a, f\}\}$. Next, we determine the SCF2 extensions of AF_1: $\sigma_{SCF2}(AF_1) = \{\{a, d\}, \{e, d\}\}$.
$EXTS^{\subseteq-max}_{po-mon}(\sigma_{SCF2}(AF_2), \sigma_{SCF2}(AF_0) \cup \sigma_{SCF2}(AF_1)) = \{\{a, d\}\}$; hence our final result is $\sigma^{BP}_{SCF2}(AF_{BP}) = \{\{a, d\}\}$.

Following the recursive definition (Definition 23), we proceed as follows.

[11] Let us note that for this BPF, applying preferred semantics would not make a difference at any of the steps that follow. This may help the reader follow along.

1. $\sigma^{BP}_{SCF2}(AF_{BP}) =$
   $EXTS^{\subseteq -max}_{po-mon}(\sigma^{BP}_{SCF2}(AF_{BP-1}), \sigma_{SCF2}(AF_0) \cup \sigma_{SCF2}(AF_1))$;
2. $AF_{BP-1} = (\langle\{a, b\}, \{c, d, e, f\}\rangle, AT)$;
3. $\sigma^{BP}_{SCF2}(AF_{BP-1}) = EXTS^{\subseteq -max}_{po-mon}(\sigma^{BP}_{SCF2}(AF_{(BP-1)-1}), \sigma_{SCF2}(AF_0))$;
4. $AF_{(BP-1)-1} = (\langle\{a, b, c, d, e, f\}\rangle, AT)$;
5. $\sigma^{BP}_{SCF2}(AF_{(BP-1)-1}) = \sigma_{SCF2}(AF_2) = \{\{a, d\}, \{a, f\}, \{e, d\}, \{e, f\}\}$;
6. $\sigma_{SCF2}(AF_0) = \{\{a, b\}\}$;
7. $\sigma^{BP}_{SCF2}(AF_{BP-1}) = EXTS^{\subseteq -max}_{po-mon}(\sigma^{BP}_{SCF2}(AF_{(BP-1)-1}), \sigma_{SCF2}(AF_0)) =$
   $\{\{a, d\}, \{a, f\}\}$;
8. $\sigma^{BP}_{SCF2}(AF_{BP}) =$
   $EXTS^{\subseteq -max}_{po-mon}(\sigma^{BP}_{SCF2}(AF_{BP-1}), \sigma_{SCF2}(AF_0) \cup \sigma_{SCF2}(AF_1)) =$
   $EXTS^{\subseteq -max}_{po-mon}(\{\{a, d\}, \{a, f\}\}, \{\{a, b\}\} \cup \{\{a, d\}, \{e, d\}\}) = \{\{a, d\}\}$.



(a) AF_0. (b) AF_1. (c) AF_2.

Fig. 3: Example: given the $AF_{BP} = (\langle\{a, b\}, \{c, d, e\}, \{f\}\rangle, \{(a, c), (a, e), (c, d), (d, b), (d, f), (e, a), (e, c), (f, b), (f, d)\})$, the figure depicts AF_0, AF_1, and AF_2.

We can show that given an argumentation semantics σ that is universally defined, the corresponding BPF semantics σ^{BP} is universally defined as well.

Proposition 1. *Let σ be an argumentation semantics. If σ is universally defined then σ^{BP} is universally defined.*

Similarly, given an argumentation semantics σ that is universally uniquely defined, the corresponding BP semantics σ^{BP} is universally uniquely defined.

Proposition 2. *Let σ be an argumentation semantics. If σ is universally uniquely defined then σ^{BP} is universally uniquely defined.*

We provide the proofs in the Appendix. Let us claim that for every universally uniquely defined argumentation semantics σ, for every burden of persuasion-framework $AF_{BP} = (\langle S_0, ..., S_n\rangle, AT)$ it holds true that $\sigma^{BP}(AF_{BP}) = \sigma(AF_n)$. We call any argumentation semantics for which this condition holds true *burden agnostic* – every universally uniquely defined argumentation semantics is burden agnostic and for burden agnostic semantics, it does not make sense to construct burden of persuasion-frameworks.

4 Semantics Selection and Skeptical Acceptance

The formal framework we have introduced in the previous section can be applied together with any universally defined argumentation semantics (see Proposition 1)[12]. To analyze the feasibility of different argumentation semantics in the context of our framework, let us first give an overview of the three main abstract argumentation semantics families, using the argumentation framework $AF = (\{a, b, c, d\}, \{(a, b), (b, c), (c, a), (a, d)\})$ as an example that highlights key differences[13].

Admissible set-based semantics. The four argumentation semantics (stable, complete, preferred and grounded, see Definition 3) that Dung introduces in his seminal paper all satisfy the principle of admissibility (see Definition 18): any extension such a semantics yields must be an admissible set. Considering the example argumentation framework AF, the only set in 2^{AR} that is admissible is $\{\}$. Hence, we suggest that typically, admissible set-based semantics are too skeptical to be useful when applied to burden of persuasion frameworks. In the example, no matter where we place burdens of persuasion, we always have to infer the empty set. In case this skepticism is considered adequate in face of odd cycles, users may consider applying a universally defined admissible set-based semantics that is relatively credulous, such as preferred or complete semantics and should then consider ignoring self-attacking arguments (or abstaining from constructing argumentation frameworks that contain self-attacking arguments). However, let us note that even then, applying weak admissible set based semantics (see below) may be more suitable.

Weak admissible set-based semantics. Baumann *et al.* introduce the weak admissible set-based semantics family [8] to address a long-standing problem with admissible set-based semantics that Dung observes in his seminal paper. Consider the example argumentation framework AF, or the even simpler framework $AF' = (\{a, d\}, \{(a, a), (a, d)\})$ and assume that an argument that – roughly speaking – defeats itself should be rejected (which is, arguably, an intuition that motivates admissibility). According to this assumption, we want to reject a when considering AF', and a, b and c, when considering AF. Consequently, we should, for sure, be able to infer d from AF (and AF'). Weak admissible set-based semantics achieve this behavior by systematically relaxing admissibility. For the sake of conciseness, we do not introduce a formal perspective on weak admissible set-based semantics. Still, let us speculate that the application of weak admissible set-based semantics may be useful in the context of burden of persuasion frameworks, given we want to ensure skepticism in face of odd cycles.

---

[12] However, it does not make sense to apply the approach using universally uniquely defined semantics, see the previous section.

[13] Note that in this section, we merely provide intuitions that can guide a practical selection of argumentation semantics. These intuitions are informed by more thorough, overviews and principle-based analyses of abstract argumentation semantics, as for example surveyed by Baroni *et al.* [3] (argumentation semantics overview) and Van der Torre and Vesic [22] (overview of argumentation principles).

Naive set-based semantics. Naive set-based semantics, as initially introduced by Verheij [23] form the most credulous of the three semantics families; the naivety principle (see Definition 18) merely requires that every extension a semantics infers is a $\subseteq$-maximal conflict-free (*naive*) set. By definition, every extension that an admissible set-based or weak admissible set-based semantics yields is conflict-free and hence entailed by a naive set. Any of the naive set-based semantics whose definitions we provide in Section 2 infers the following three extensions from the example framework AF: $\{a\}$, $\{b, d\}$, and $\{c, d\}$. Naive set-based semantics start off with the naivety principle, and then typically formalize further constraints that are related to the notions of SCC-recursiveness (see Section 2) or *range*, *i.e.* $\subseteq$-maximality of an extension in union with the arguments the extension attacks. Among the four "reasonable" naive set-based semantics (not considering naive semantics, which does not impose any further constraint besides naivety), the two semantics that employ the notion of *range*, *i.e.* stage and stage2 semantics, can be considered more skeptical than the two semantics that are SCC-recursively defined, but do not use range (CF2 and SCF2 semantics). Consider AF'' as introduced by Example 4. Also, Example 4 highlights that stage, stage2, and CF2 semantics may behave counter-intuitively when self-attacking arguments are present; hence, self-attacking arguments should be avoided or ignored. Because of the well-known limitations (see Example 4 and also Dvorak and Gaggl [15], as well as Cramer and Van der Torre [12]), there is most likely no use-case that justifies the application of CF2 semantics; instead SCF2 semantics should be applied, or – if SCF2 semantics is deemed too complex – a stage semantics variant that ignores self-attacking arguments may be a reasonable and slightly more skeptical approximation.

In the context of our burden of persuasion framework, naive set-based semantics are arguably the most interesting abstract argumentation family, due to their relatively credulous behavior. This behavior can then be further constrained by the burden of persuasion model in a BPF. Still, in many scenarios, a naive set-based semantics yields several extensions for a given BPF, and hence is inconclusive. Then, we can use the notion of credulous and skeptical acceptance as an additional assessment layer; in particular, we may ask the following questions. *i)* Given a set of arguments that includes burdened arguments (or, in the case of multiple levels of burdens: arguments with a high level of burden), are these arguments entailed by the skeptical extension we can infer? *ii)* Given a set of arguments that are unburdened (or, in the case of multiple levels of burdens: unburdened arguments or arguments with a low level of burden), are these arguments entailed by at least one extension we can infer? Let us claim that in the case of naive set-based semantics, the notions of credulous and skeptical acceptance are more useful than the notion of *undecided* arguments in traditional labeling-based approaches (see, *e.g.*, Wu and Caminada [24]); all arguments that are not entailed by a naive-based extension are in conflict with this extension and hence, it is counter-intuitive to consider arguments that are not attacked by the extension – and consequently, are attackers of the extension – as undecided.

5 Discussion

From a formal theory perspective, our framework for modeling burdens of persuasion can be considered a contribution to the research area of *argumentation dynamics* (see Doutre and Mailly [13] for a survey). At first glance, this connection may not be obvious. However, let us observe that we can model a BPF $AF_{BP} = (\langle S_0, ..., S_1 \rangle, AT)$ as a sequence of normal expansions (see Definition 2) $\langle AF_0, ..., AF_n \rangle$, such that for $AF_i, 0 < i \leq n, AF_{i-1} \preceq_N AF_i$. For example, given the BPF $AF_{BP} = (\langle \{a\}, \{b\}, \{c\} \rangle, \{(a, b), (b, a), (b, c), (c, b)\})$, we have the sequence of normally expanding argumentation frameworks $\langle AF_0, AF_1, AF_2 \rangle = \langle (\{a\}, \{\}), (\{a, b\}, \{(a, b), (b, a)\}), (\{a, b, c\}, \{(a, b), (b, a), (b, c), (c, b)\}) \rangle$. Given this sequence (and an argumentation semantics σ), BP semantics applies an abstract argumentation semantics and returns $EXTS^{\subseteq-max}_{po-mon}(EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_2), \sigma(AF_0)), \sigma(AF_0) \cup \sigma(AF_1))$.

Let us note that the formal framework we provide is fundamentally different from traditional approaches to model preferences in formal argumentation, such as preference-based [1] and value-based [10] argumentation (where value-based argumentation is a generalization of preference-based argumentation). While the sequence of sets of arguments in a BPF can be considered as a total preference order on non-intersecting sets of arguments, the way this order is interpreted by BP semantics does not allow for the inference of sets of arguments that entail conflicts; the order merely gives us a way to treat uncertainty ("doubt") that is inherent in the corresponding abstract argumentation framework. In contrast, in preference-based argumentation, preferences may lead to a disregard of conflicts. Colloquially speaking, we can summarize that value-based and preference-based argumentation favor preferred arguments no matter what when drawing inferences in face of contradictions, whereas our burden of persuasion approach merely favors preferred sets of arguments *if in doubt*.

Still, let us note that our burden of persuasion frameworks and semantics reflect the idea of using preferences on the set of arguments in an argumentation framework to "narrow down" the extensions that an abstract argumentation semantics returns. Work in this direction has been conducted by Kaci *et al.* [18], as well as by Amgoud and Vesic [2]. For the sake of conciseness, let us informally claim here that each BPF can be mapped to a preference-based argumentation framework, but that the aforementioned approaches are fundamentally different to ours. For instance, let us claim that when considering the BPF $AF_{BP} = \langle \{a, c\}, \{b, d\} \rangle, \{(a, b), (a, c), (b, a), (b, d)\})$ and preferred semantics, neither Kaci *et al.*'s approach, nor the two approaches (*democratic* and *elitist*) introduced by Amgoud and Vesic allow for inferring only the extension $\{a, d\}$ but also infer the extension $\{b, c\}$. However, as b carries the burden of persuasion, it should not be able to defeat a, which then in turn can defeat the unburdened argument c. A formal, detailed comparison can be considered promising future work.

Similarly, our approach is different from *argumentation with many lives* in which arguments and attacks have numeric weights and an argument is defeated iff the sum of the weights of successful attacks on the argument exceeds the number of *lives* of the argument (roughly speaking) [17]. Similarly to value-based argumentation, argumentation with many lives allows for the inference of sets of arguments that are not conflict-free; also, it requires the assignment of weights (quantification) of arguments and attacks, which is not feasible in many legal use cases.

From a legal perspective, let us note that the burden of persuasion is related to, but different from, the *standard of persuasion* [16] which, from a formal argumentation perspective, relates more directly to the required strength of one or several attackers to defeat an argument. Modeling standards of persuasion in formal argumentation is certainly interesting future work, but not within the scope of this paper.

Considering previous research on formal models of burdens of persuasion, our work can be considered a continuation of recent research that introduces the burden of persuasion to structured argumentation [11]. This model of the burden of persuasion is based on grounded semantics and can be described – from an abstract argumentation perspective – as follows.

1. Given an abstract argumentation framework $AF = (AR, AT)$, we place the burden of persuasion on the arguments in a set $S \subseteq AR$.
2. We determine the grounded extension E_{gr} of AF and say that an argument $a \in AR$ is labeled as follows. IN if $a \in E_{gr}$; OUT if $a \in E_{gr}^{+}$; UND, otherwise. We denote all arguments labeled IN by $IN_{gr}(AF)$; all arguments labeled OUT by $OUT_{gr}(AF)$; all arguments labeled UND by $UND_{gr}(AF)$.
3. Based on the grounded labeling, we create the grounded burden of persuasion labeling (BP labeling). A BP-labeling is a 3-tuple $(IN^{BP}(AF), OUT^{BP}(AF), UND^{BP})$, such that $\forall a \in AR$, the following holds:
   If $a \in S$**.** $a \in IN^{BP}(AF)$ if $a \in E_{gr}$; $a \in OUT^{BP}(AF)$ if $a \in E_{gr}^{+}$ or $a \in (UND_{gr}(AF) \setminus S)^{+}$; $a \in UND^{BP}(AF)$, otherwise.
   If $a \notin S$**.** $a \in IN^{BP}(AF)$ if $a \notin E_{gr}^{+}$ and $\forall b \in IN^{BP}(AF), b$ does not attack a; $a \in OUT^{BP}(AF)$, otherwise.

This approach has shortcomings (even when only considering one burden of persuasion level as above). Below we give two examples that also illustrate how our framework addresses the issues.

Self-attacking arguments. Consider the argumentation framework $AF = (AR, AT) = (\{a, b, c, \}, \{(a, a), (a, b), (b, c)\})$ with the burden of persuasion placed on $\{b\}$. Considering the approach by Calegari *et al.*, we have: *i)* a is UND; *ii)* b is initially undecided, and because it carries the burden of persuasion, it is finally out; *iii)* hence, c is in. This is problematic, because a as a self-defeating argument should arguably not defeat b, even if the burden of persuasion lies on b. In contrast, when using our approach we have the following BPF: $AF_{BP} = (\langle\{a, c\}, \{b\}\rangle, AT)$. $\sigma_{SCF2}^{BP}(AF_{BP}) = \{\{b\}\}$; *i.e.*, we infer $\{b\}$ because the burden of persuasion is not strong enough to allow for the defeat of b by a self-attacking argument.

Consistent defeat from inconsistent arguments. Consider the abstract argumentation framework $AF' = (AR', AT') = (\{a, b, c, d, e\}, \{(a, b), (b, a), (a, c), (b, c), (c, d), (d, e)\})$. What we have in this framework is a phenomenon that we can colloquially describe as *consistent defeat from inconsistent arguments*. We place the burden of persuasion on argument $\{d\}$. Let us apply the approach by Calegari *et al.* a and b attack each other and are hence undecided, but both arguments *consistently* attack c. Again considering three-valued labeling and grounded semantics, we have d is out and e is in. However, we claim that we should conclude that c is out, because it is attacked by both a and b, and that consequently, d is in and e is out. Let us highlight the difference to the previous example. In the previous example, we maintain it should be impossible to infer a because a is inconsistent with itself. However, in this example, we maintain it should be impossible to infer "not d", because we have to infer "either a or b", which implies the defeat of c. Our approach supports this intuition: $AF'_{BP} = (\langle\{a, b, c, d\}, \{e\}\rangle, AT')$ and $\sigma^{BP}_{SCF2}(AF'_{BP}) = \{\{a, d\}, \{b, d\}\}$.

6 Conclusion

In this paper, we have introduced a formal framework for modeling the burden of persuasion in abstract argumentation, which is accompanied by an open source software implementation. The framework supports arbitrary many levels of burdens, can be combined with any universally defined argumentation semantics, and addresses some open issues that previous works have identified in models of burdens of persuasion for structured argumentation. By abstracting from structured argumentation specifics, the framework can be applied to a range of formal argumentation variants.

Acknowledgments. We thank the anonymous reviewers for their thoughtful and useful feedback. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

1. Amgoud, L., Cayrol, C.: Inferring from inconsistency in preference-based argumentation frameworks. Journal of Automated Reasoning **29**(2), 125–169 (Jun 2002). https://doi.org/10.1023/A:1021603608656
2. Amgoud, L., Vesic, S.: Rich preference-based argumentation frameworks. International Journal of Approximate Reasoning **55**(2), 585–606 (2014). https://doi.org/https://doi.org/10.1016/j.ijar.2013.10.010, `https://www.sciencedirect.com/science/article/pii/S0888613X13002326`
3. Baroni, P., Caminada, M., Giacomin, M.: Abstract argumentation frameworks and their semantics. In: Baroni, P., Gabbay, D., Massimiliano, G., van der Torre, L. (eds.) Handbook of Formal Argumentation. College Publications, chap. 4, pp. 159–236. College Publications (2018)
4. Baroni, P., Gabbay, D.M., Giacomin, M., van der Torre, L.: Handbook of formal argumentation. College Publications (2018)
5. Baroni, P., Giacomin, M.: On principle-based evaluation of extension-based argumentation semantics. Artificial Intelligence **171**(10), 675 – 700

(2007). https://doi.org/https://doi.org/10.1016/j.artint.2007.04.004, http://www.sciencedirect.com/science/article/pii/S0004370207000744, argumentation in Artificial Intelligence
6. Baroni, P., Giacomin, M., Guida, G.: Scc-recursiveness: a general schema for argumentation semantics. Artificial Intelligence **168**(1), 162 – 210 (2005). https://doi.org/https://doi.org/10.1016/j.artint.2005.05.006
7. Baumann, R., Brewka, G.: Expanding argumentation frameworks: Enforcing and monotonicity results. COMMA **10**, 75–86 (2010)
8. Baumann, R., Brewka, G., Ulbricht, M.: Revisiting the foundations of abstract argumentation-semantics based on weak admissibility and weak defense. In: AAAI. pp. 2742–2749 (2020)
9. Bench-Capon, T., Prakken, H., Sartor, G.: Argumentation in legal reasoning. In: Argumentation in artificial intelligence, pp. 363–382. Springer (2009)
10. Bench-Capon, T.J.: Persuasion in practical argument using value-based argumentation frameworks. Journal of Logic and Computation **13**(3), 429–448 (2003)
11. Calegari, R., Riveret, R., Sartor, G.: The burden of persuasion in structured argumentation. In: Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law. ICAIL '21, Association for Computing Machinery, New York, NY, USA (2021)
12. Cramer, M., van der Torre, L.: Scf2-an argumentation semantics for rational human judgments on argument acceptability. In: Proceedings of the 8th Workshop on Dynamics of Knowledge and Belief (DKB-2019) and the 7th Workshop KI\& Kognition (KIK-2019) co-located with 44nd German Conference on Artificial Intelligence (KI 2019), Kassel, Germany, September 23, 2019. pp. 24–35 (2019)
13. Doutre, S., Mailly, J.G.: Constraints and changes: A survey of abstract argumentation dynamics. Argument & Computation **9**, 223–248 (2018). https://doi.org/10.3233/AAC-180425
14. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artificial intelligence **77**(2), 321–357 (1995)
15. Dvořák, W., Gaggl, S.A.: Stage semantics and the SCC-recursive schema for argumentation semantics. Journal of Logic and Computation **26**(4), 1149–1202 (02 2014). https://doi.org/10.1093/logcom/exu006, https://doi.org/10.1093/logcom/exu006
16. Friedman, R.D.: Standards of persuasion and the distinction between fact and law. Northwestern University Law Review **86**, 916 (1991)
17. Gabbay, D.M., Rozenberg, G.: Introducing abstract argumentation with many lives. FLAP **7**(3), 295–336 (2020)
18. Kaci, S., van der Torre, L., Villata, S.: Preference in abstract argumentation. In: Computational Models of Argument, pp. 405–412. IOS Press (2018)
19. Kampik, T., Gabbay, D.: The "Degrees of Monotony"-Dilemma in Abstract Argumentation. In: Vejnarová, J., Wilson, N. (eds.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty 2021, to appear. Springer International Publishing, Cham (2021)
20. Prakken, H., Sartor, G.: A logical analysis of burdens of proof. Legal evidence and proof: Statistics, stories, logic pp. 223–253 (2009)
21. Thimm, M.: Tweety: A comprehensive collection of java libraries for logical aspects of artificial intelligence and knowledge representation. In: Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning. p. 528–537. KR'14, AAAI Press (2014)

22. van der Torre, L., Vesic, S.: The principle-based approach to abstract argumentation semantics. IfCoLog Journal of Logics and Their Applications **4**(8), 2735–2778 (October 2017)
23. Verheij, B.: Two approaches to dialectical argumentation: admissible sets and argumentation stages. Proc. NAIC **96**, 357–368 (1996)
24. Wu, Y., Caminada, M.: A labelling-based justification status of arguments. Studies in Logic **3**(4), 12–29 (2010)

Appendix - Proofs

Proposition 1. *Let σ be an argumentation semantics. If σ is universally defined then σ^{BP} is universally defined.*

Proof. Let $AF_{BP} = (ARS, AT)$ be a BPF and $ARS = \langle S_0, ..., S_n \rangle$. If $n = 0$, by definition of σ^{BP} (Definition 23) it holds true that $\sigma^{BP}(AF_{BP}) = \sigma(AF_0)$. Hence, the proposition holds true for $n = 0$. For $n > 0$, we provide a proof by induction on n.
Base case: $n = 1$**.** By definition of σ^{BP}, it holds true that $\sigma^{BP}(AF_{BP}) = EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_1), \sigma(AF_0))$. Because σ is universally defined, by definition of $EXTS^{\subseteq-max}_{po-mon}$ (Definition 22), it holds true that $|EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_1), \sigma(AF_0))| \geq 1$. Hence, the proposition holds true for the base case.
Inductive case: $n = k + 1$**.** By definition of σ^{BP}, it holds true that $\sigma^{BP}(AF_{BP}) = EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_{BP-1}), \sigma(AF_0) \cup ... \cup \sigma(AF_{k+1}))$. Because σ is universally defined it holds true that $|\sigma(AF_0) \cup ... \cup \sigma(AF_{k+1})| \geq 1$ and from the base case and from the definition of $EXTS^{\subseteq-max}_{po-mon}$ it follows that $|\sigma(AF_{BP-1})| \geq 1$. Hence, $\sigma^{BP}(AF_{BP})$ is universally defined for $n = k + 1$ and the proof follows from the inductive case. $\square$

Proposition 2. *Let σ be an argumentation semantics. If σ is universally uniquely defined then σ^{BP} is universally uniquely defined.*

Proof. Let $AF_{BP} = (ARS, AT)$ be a BPF and $ARS = \langle S_0, ..., S_n \rangle$. If $n = 0$, by definition of σ^{BP} (Definition 23) it holds true that $\sigma^{BP}(AF_{BP}) = \sigma(AF_0)$. Hence, the proposition holds true for $n = 0$. For $n > 0$, we provide a proof by induction on n.
Base case: $n = 1$**.** By definition of σ^{BP}, it holds true that $\sigma^{BP}(AF_{BP}) = EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_1), \sigma(AF_0))$. Because σ is universally uniquely defined, by definition of $EXTS^{\subseteq-max}_{po-mon}$ (Definition 22), it holds true that $|EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_1), \sigma(AF_0))| = 1$. Hence, the proposition holds true for the base case.
Inductive case: $n = k + 1$**.** By definition of σ^{BP}, it holds true that $\sigma^{BP}(AF_{BP}) = EXTS^{\subseteq-max}_{po-mon}(\sigma(AF_{BP-1}), \sigma(AF_0) \cup ... \cup \sigma(AF_{k+1}))$. Because σ is universally uniquely defined it holds true that $|\sigma(AF_0) \cup ... \cup \sigma(AF_{k+1})| \geq 1$ and from the base case and from the definition of $EXTS^{\subseteq-max}_{po-mon}$ it follows that $|\sigma(AF_{BP-1})| = 1$. Hence, $\sigma^{BP}(AF_{BP})$ is universally uniquely defined for $n = k + 1$ and the proof follows from the inductive case. $\square$