A comprehensive account of the burden of persuasion in abstract argumentation

TIMOTHEUS KAMPIK, *Department of Computing Science, Umeå University, Umeå 90187, Sweden.*
E-mail: tkampik@cs.umu.se

DOV GABBAY, *Department of Informatics, King's College London, London WC2R 2, UK, and Department of Computer Science, University of Luxembourg, Esch-sur-Alzette 4364, Luxembourg.*
E-mail: dov.gabbay@kcl.ac.uk

GIOVANNI SARTOR, *Department of Law, European University Institute, Florence 50139, Italy, and Department of Legal Studies, Università di Bologna, Bologna 40126, Italy.*
E-mail: 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. A theoretical analysis shows that our approach can be generalized to a novel method for the preference-based selection of extensions from argumentation frameworks.

1 Introduction

Over the past decades, formal argumentation has emerged as a promising collection of methods for reasoning under uncertainty [5]. A particularly relevant application domain that can benefit from argumentation-based models of conflicts and contradictions is legal reasoning [12]. 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* [26]. By saying that an argument carries the burden of 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 build from the theory (considering constraints imposed by a basic inference function), the burden of persuasion

Vol. 33, No. 2,
Advance Access Publication on 27 January 2023 https://doi.org/10.1093/logcom/exac094

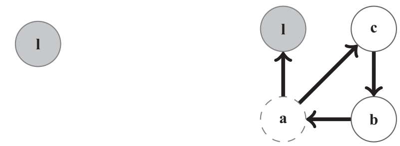
dictates that we must be *less skeptical* towards unburdened arguments than towards burdened ones. If we are faced with conflicting extensions, one being only supported by burdened arguments and one being only supported by unburdened arguments, we select the latter. Moreover, any successful attacks against a burdened argument entail that the burdened argument is to be rejected.[1] In a recent paper, Calegari *et al.* present a model of the burden of persuasion that is based on a structured argumentation approach [14]; in their paper, the authors 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*. From a purely technical perspective, the results provide a novel approach to extension-selection in preference-based argumentation.

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 obtain 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 still 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, so that arguments are developed for and against the doctor's liability. 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 unburdened, as the the burden is placed on the doctor's absence of liability, i.e. the rejection of l. 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 affirmed 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. According to this witness, 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 the scientific literature. Finally, argument c is attacked by argument a, which includes the claim that one single journal article was sufficient to establish a scientific claim; we assume that c does not attack a because the Court's expert witness has not claimed that relying on a late-breaking research result is insufficient, whereas argument a has advanced a general claim that leads to the emergence of the attack from a to c.

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



(a) AF. **(b)** AF'.

FIGURE 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.

We end up with an *argumentation framework*—a tuple consisting of a set of *arguments* AR' and a set of *attacks* $AT' \subseteq AR' \times AR'$ (Figure 1)—as follows:

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

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 about her diligence based on the cycle 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 [19] 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 admissibility-based argumentation semantics family [10] 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.[2]

[2] 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 admissibility-based semantics as introduced by Baumann *et al.* [10] may not be sufficiently credulous for many applications that require a model of the burden of persuasion.

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 [7] and SCF2 [17] 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 example, which also illustrates how we can manage multiple levels of the burden of persuasion.

EXAMPLE 2
Let us consider a criminal law scenario. Assume the prosecution has raised two claims against a suspect in a homicide case: (i) that the suspect was at the scene of a crime around the time the crime was committed, in which case he may be considered to be at least an accomplice in the homicide, and (ii) that he had the knowledge and skills to handle the weapon that was used, in which case he could also be considered the material author of the crime. With respect to (i), three witnesses—w_a, w_b, and w_e—make mutually inconsistent claims: w_a and w_b claim they have seen the defendant at different (mutually inconsistent) locations at the given time; the claims are denoted by arguments a and b, respectively. w_e claims that the defendant *was* at the scene of the crime, denoted by argument e; a, b and e attack each other. With respect to (ii), the defendant claims that he does not (and did not) have the knowledge and skills to handle the weapon that was used (argument c), whereas another witness claims the contrary (d). The arguments e and d are attacking the position of the defendant. Consequently, they carry the highest level of burden. Argument c is produced by the defendant and hence carries a higher level of burden than the arguments a and b that support the position of the defendant, but are put forward by independent witnesses.

We can model the scenario as the following argumentation framework:

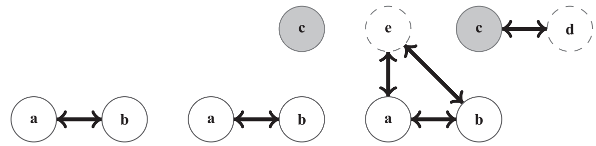
$$AF'' = (\{a, b, c, d, e\}, \{(a, b), (a, e), (b, a), (b, e), (c, d), (d, c), (e, a), (e, b)\})$$

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[3] (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 $AF = (\{a, b\}, \{(a, b), (b, a)\})$. From AF, we can infer either $\{a\}$ or $\{b\}$.

[3] In this example, the inferences we draw from the abstract argumentation frameworks coincide, e.g. with the *extensions* (sets of arguments) returned by CF2 [7] and SCF2 [17] semantics.



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

FIGURE 2. Multiple levels of burdens of persuasion.

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[4] :

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

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

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 $\{b, c\}$'. However, in our legal reason scenario, the inference results are sufficient to establish that the claims against the defendant cannot be upheld.

A software implementation of the formal burden of persuasion framework and semantics that we introduce in this paper is available at https://git.io/JGueN.

Let us highlight that we do not provide examples that use structured argumentation for the sake of simplicity, even though this may raise some questions about modeling decisions on abstract level: generally, the examples are constructed to motivate the need for a generic approach to drawing inferences from argumentation frameworks with burdened arguments and hence focus on particularly intricate phenomena, such as the handling of odd loops or multiple levels of burden. Although we expect that applications to legal reasoning make use of structured argumentation, we separate the abstract-level challenges from structure-level concerns, which are addressed by other works. Indeed, the approach we provide in this paper is motivated by some limitations of a similar method that has its roots in structured argumentation [14, 16] and it is interoperable with structured argumentation (see Section 7 for a detailed discussion).

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. Sections 5 and 6 then formally analyze the introduced formal framework from the perspective of argumentation dynamics and preference-based argumentation, respectively. Finally, Section 7 discusses the framework in the context of related research on burdens of persuasion, before Section 8 concludes the paper. Note

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

that the paper revises and extends a shorter paper that has previously been disseminated in conference proceedings [25]; in particular, this version contains additional legal reasoning examples and provides a formal analysis of our burden of persuasion frameworks and semantics from the perspective of argumentation dynamics and preference-based argumentation.

2 Preliminaries

This section introduces the preliminaries that our work is based upon, in particular abstract argumentation and its semantics, as well as preference-based argumentation. Let us highlight that in order to follow the paper, the reader does not need to understand the behavior of all of the introduced semantics.

2.1 Abstract argumentation

The central notion this paper uses is Dung's (abstract) argumentation framework [19]. 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.

Let us provide the definition of the restriction of an argumentation framework $AF = (AR, AT)$ to a set of arguments $S \subseteq AR$.

DEFINITION 1 (Restriction [7]).
Let $AF = (AR, AT)$ be an argumentation framework. Given a set $S \subseteq AR$, let $AF \downarrow_S$ be defined as $(S, AT \cap S \times S)$. 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 2 (Conflict-free, Unattacked and Admissible Sets [4]).
Let $AF = (AR, AT)$ be an argumentation framework. A set $S \subseteq AR$:

- is *conflict-free* iff $\nexists a, b \in S$ such that a attacks b;
- is *unattacked* iff $\nexists a \in AR \setminus S$ such that a attacks S;
- is *admissible* iff S is conflict-free and $\forall a \in S$, it holds true that S defends a.

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 (Stable, Preferred, Complete and Grounded Semantics [19]).
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 [29]).
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 all 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 $SCCS(AF)$.

Let us define a final preliminary for SCC-recursive semantics: the *UP* function.

DEFINITION 9 (UP [7]).
Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$, $S \subseteq AR$. We define $UP_{AF}(S, E) = \{a | a \in S, \nexists b \in E \setminus S$ such that $(b, a) \in AT\}$.

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

DEFINITION 10 (CF2 Semantics [7]).
Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$. E is a CF2 extension of AF 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 [20]).
Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$. E is a stage2 extension of AF 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 [17]. The authors start by defining a notion that ignores self-attacking arguments.

DEFINITION 12 (nsa(AF) [17]).
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 [17]).
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 $0 \leq j < k < n$, 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) [17]).
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 [6, 28]. Cramer and Van der Torre turn the *strong completeness outside odd cycles* property into a principle to ‘catch’ unintuitive CF2 behavior.

DEFINITION 16 (Strong Completeness Outside Odd Cycles (SCOOC Principle) [17]).
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 [17]).
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:

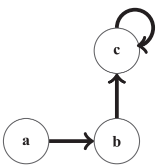
- E is a naive extension of $nsa(AF)$ and E is strongly complete outside odd cycles if $|SCCS(nsa(AF))| = 1$;
- $\forall S \in SCCS(nsa(AF))$, $(E \cap S)$ is an SCF2 extension of $AF \downarrow_{UP_{nsa(AF)}(S,E)}$, otherwise.

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

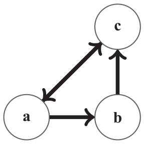
Let us introduce some examples that illustrate the behaviors of—and highlights the difference between—stage, CF2, stage2 and SCF2 semantics. However, let us note that a detailed explanation of the semantics is beyond the scope of this paper and the reader may consult the original works instead.

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

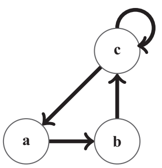
σ / AF	**stage**	**CF2**	**stage2**	**SCF2**
AF'	$\{a\},\{b\}$	$\{a\}$	$\{a\}$	$\{a\}$
AF''	$\{a\}$	$\{a\},\{b\},\{c\}$	$\{a\}$	$\{a\},\{b\},\{c\}$
AF^*	$\{a\},\{b\}$	$\{a\},\{b\}$	$\{a\},\{b\}$	{a}
AF^{**}	$\{a,c,e\}$, $\{b,d,f\}$	$\{a,c,e\}$, $\{b,d,f\}$ $\{a,d\}$, $\{b,e\}$, $\{c,f\}$	$\{a,c,e\}$,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



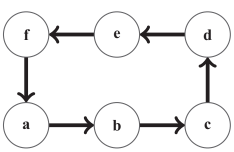
(a) AF'.



(b) AF''.



(c) AF^*.



(d) AF^{**}.

FIGURE 3. Examples that highlight the differences between stage, CF2, stage2 and SCF2 semantics, see Table 1.

EXAMPLE 4
Let us consider the following argumentation frameworks (Figure 3):

- $AF' = (\{a,b,c\},\{(a,b),(b,c),(c,c)\})$;
- $AF'' = (\{a,b,c\},\{(a,b),(a,c),(b,c),(c,a)\})$;
- $AF^* = (\{a,b,c\},\{(a,b),(b,c),(c,a),(c,c)\})$;
- $AF^{**} = (\{a,b,c,d,e,f\},\{(a,b),(b,c),(c,d),(d,e),(e,f),(f,a)\})$.

Table 1 displays the extensions stage, CF2, stage2 and SCF2 semantics yield for these argumentation frameworks.

Recently, weak admissibility-based semantics have been defined, which are based on a relaxation of admissibility.[5]

To introduce this semantics family, we need to provide some preliminaries. Given an argumentation framework and a subset E of its arguments, an E-reduct restricts the argumentation framework to the arguments that are neither in E nor attacked by E.

[5]The underlying formal notion was first defined by Kakas and Mancarella [23].

Definition 18 (*E*-Reduct [10]).
Let $AF = (AR, AT)$ be an argumentation framework and let $E \subseteq AR$. The E-reduct of AF is the argumentation framework $AF^E = AF \downarrow_{E^*}$, where $E^* = AR \setminus (E \cup \{a | a \in AR, E \text{ attacks } a\})$.

A weakly admissible set E is a set of arguments that is conflict-free and defends itself against all attackers that appear in a weakly admissible set of the corresponding argumentation framework's E-reduct.

Definition 19 (Weak Admissibility [10]).
Let $AF = (AR, AT)$ be an argumentation framework. $E \subseteq AR$ is *weakly admissible* in AF, denoted by $E \in ad^w(AF)$, iff E is conflict-free and $\forall a \in AR$, such that a attacks E it holds true that $a \notin \bigcup_S S \in ad^w(AF^E)$.

Based on weak admissibility, we can define the notion of *weak defense*.

Definition 20 (Weak Defense [10]).
Let $AF = (AR, AT)$ be an argumentation framework and let $E, S \subseteq AR$. E *weakly defends* S iff $\forall a \in AR$, such that a attacks S the following statement holds true:

$$E \text{ attacks } a \text{ or}$$

$$\Big(a \notin \bigcup_S S \in ad^w(AF^E), a \notin E \text{ and } S \subseteq S' \in ad^w(AF)\Big).$$

Roughly speaking, weakly preferred, weakly complete and weakly grounded semantics are the weak admissibility-based counterparts to preferred, complete and grounded semantics.

Definition 21 (Weakly Preferred, Weakly Complete and Weakly Grounded Semantics and Extensions [10]).
Let $AF = (AR, AT)$ be an argumentation framework. $E \subseteq AR$ is a:

- *weakly preferred extension* of AF iff E is $\subseteq$-maximal in $ad^w(AF)$. Weakly preferred semantics $\sigma_{wp}(AF)$ denotes all weakly preferred extensions of AF.
- *weakly complete extension* of AF iff $E \in ad^w(F)$ and for any set S, such that $E \subseteq S$ and S is weakly defended by E, it holds true that $S \subseteq E$. Weakly complete semantics $\sigma_{wc}(AF)$ denotes all weakly complete extensions of AF.
- *weakly grounded extension* of AF iff E is $\subseteq$-minimal in $\sigma_{wc}(AF)$. Weakly grounded semantics $\sigma_{wg}(AF)$ denotes all weakly grounded extensions of AF.

Let us introduce an example that illustrates the differences between weakly preferred, weakly complete and weakly grounded extensions.

Example 5
Consider the argumentation framework $AF = (\{a, b, c, d, e\}, \{(a, b),$
$(b, c), (c, a), (c, d), (d, e), (e, d)\})$ (Figure 4). The framework yields the following extensions:

- $\sigma_{wp}(AF) = \{\{e\}, \{d\}\}$;

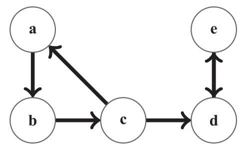


FIGURE 4. Differences between weak admissibility-based semantics.

- $\sigma_{wc}(AF) = \{\{\}, \{e\}, \{d\}\}$;
- $\sigma_{wg}(AF) = \{\{\}\}$.

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

DEFINITION 22 ((Weak) Admissibility and Naivety Principles [10, 28]).
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 (in AF).
- σ satisfies the weak admissibility principle iff for every argumentation framework $AF = (AR, AT)$, $\forall E \in \sigma(AF), E$ is a weakly admissible set (in AF).
- 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 (in AF).

Each of these three principles 'covers' a semantics family (see [28] and [10]):

- Stable, preferred, complete and grounded semantics satisfy admissibility, i.e. these semantics are part of the admissibility-based semantics family.
- Weakly stable, weakly preferred, weakly complete and weakly grounded semantics satisfy weak admissibility, i.e. these semantics are part of the weak admissibility-based semantics family.
- Naive, stage, CF2, stage2, nsa(CF2) and SCF2 semantics satisfy naivety, i.e. these semantics are part of the naive set-based semantics family.

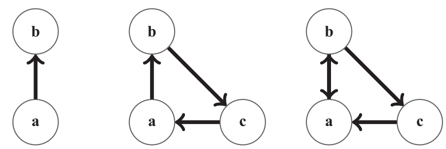
Combined, the three semantics families include all semantics that are defined in Section 2, as well as all semantics that are typically surveyed in the literature [4, 28].

To model change in argumentation frameworks (*argumentation dynamics*), Baumann and Brewka introduce the notion of argumentation framework *expansions*.

DEFINITION 23 ((Normal) Expansions and Expansion Chains [9]).
Let $AF = (AR, AT)$ and $AF' = (AR', AT')$ be argumentation frameworks.

- AF' is an *expansion* of AF, denoted by $AF \preceq AF'$, iff $AR \subseteq AR'$ and $AT \subseteq AT'$.
- AF' is a *normal expansion* of AF, denoted by $AF \preceq_N AF'$, iff $AF \preceq AF'$ and $(AT' \setminus AT) \cap (AR \times AR) = \emptyset$.

A sequence of argumentation frameworks $\langle AF_0, ..., AF_n\rangle$ is an *expansion chain* iff for $0 \leq i < n$ it holds true that $AF_i \preceq_N AF_{i+1}$.



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

FIGURE 5. Expansions and normal expansions: $AF \preceq AF'$ and $AF \preceq_N AF'$; $AF \preceq AF''$ but $AF \npreceq_N AF'$; $AF' \preceq AF''$ but $AF' \npreceq_N AF''$.

Intuitively, an expansion adds new arguments or attacks to an argumentation framework, but does not remove any arguments or attacks; a normal expansion is an expansion that does not add attacks between existing arguments, i.e. each new attack must originate from or attack a new argument. 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).

Let us introduce an example that illustrates the differences between expansions and normal expansions.

EXAMPLE 6
Consider the following argumentation frameworks (Figure 5):

- $AF = (AR, AT) = (\{a, b\}, \{(a, b)\})$;
- $AF' = (AR', AT') = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$;
- $AF'' = (AR'', AT'') = (\{a, b, c\}, \{(a, b), (b, a), (b, c), (c, a)\})$.

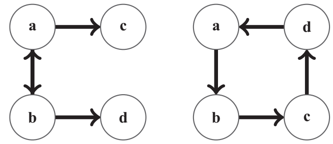
$AF'$is an expansion of AF ($AF \preceq AF'$) because $AR \subseteq AR'$ and $AT \subseteq AT'$, and AF' is also a normal expansion of AF ($AF \preceq_N AF'$) because in addition, $(AT' \setminus AT) \cap (AR \times AR) = \emptyset$. AF'' is an expansion of AF ($AF \preceq AF''$) because $AR \subseteq AR''$ and $AT \subseteq AT''$, but not a normal expansion of AF ($AF \npreceq_N AF''$), because $(AT'' \setminus AT) \cap (AR \times AR) = \{(b, a)\}$. Analogously, it holds that $AF' \preceq AF''$, but $AF' \npreceq_N AF''$.

2.2 *Preference-based Argumentation*

Amgoud and Cayrol extend abstract argumentation frameworks to support the modeling of a preference relation on an argumentation framework's set of arguments.

DEFINITION 24 (Preference-based Argumentation Framework [1]).
A Preference-based Argumentation Framework (PAF) is a triple $AF_P = (AR, AT, Pref)$, where AR is a set of arguments $AT \subseteq AR \times AR$ ('attacks'), and $Pref$ is a preorder (a reflexive and transitive) binary relation on AR.

Given a PAF $AF_P = (AR, AT, Pref)$ and two arguments $a, b \in AR$, we denote $(a, b) \in Pref$ by $a \succeq b$ (in AF_P) and $((a, b) \in Pref$ and $(b, a) \notin Pref)$ by $a \succ b$ (in AF_P); given $a \succeq b$ we say that 'a is preferred over b', and given $a \succ b$ we say that 'a is strictly preferred over b'. Recall that because $Pref$ is reflexive and transitive for x, y, z it holds true that (i) $x \succeq x$; (ii) if $x \succeq y$ and $y \succeq z$ then $x \succeq z$. Note that henceforth, we assume reflexivity without explicitly modeling it.



(a) AF_P. **(b)** AF'_P.

FIGURE 6. Examples: extension-selecting approaches to preference-based argumentation.

In initial approaches to preference-based argumentation, inferences are drawn by considering an attack from an argument a to an argument b as unsuccessful if b is strictly preferred over a (roughly speaking). This has an obvious impact on the interpretation of conflicts and can notably lead to the inference of extensions that are not conflict-free. In the context of our work, approaches are more relevant that use preferences to instead select some of the extensions an abstract argumentation semantics returns for a given argumentation framework.

DEFINITION 25 (Extension-Selection in Preference-based Argumentation [2, 22]).
Let $AF_P = (AR, AT, Pref)$ be a PAF, let $A \subseteq 2^{AR}$, let $E, E' \in A$ and let σ be an argumentation semantics. Given AF_P and A we say that:

- 'E is at least as good as E'' iff $\forall a, b \in AR$, it does not hold true that $a \succ b$ and $a \in E' \setminus E$ and $b \in E \setminus E'$; we say that 'E is better than E'' iff E is at least as good as E' and E' is not at least as good as E; we say that 'E is best among A' iff $\nexists E'' \in A$ such that E'' is better than E. $\sigma^{P,best}(AF_P)$ denotes the best extensions among $\sigma((AR, AT))$ given *Pref*.
- 'E is at least as democratic as E'' iff $\forall b \in E' \setminus E, \exists a \in E \setminus E'$, such that $a \succ b$; we say that 'E is more democratic than E'' iff E is at least as democratic as E' and E' is not at least as democratic as E; we say that 'E is most democratic among A' iff $\nexists E'' \in A$ such that E'' is more democratic than E. $\sigma^{P,dem}(AF_P)$ denotes the most democratic extensions among $\sigma((AR, AT))$ given *Pref*.
- 'E is at least as elitist as E'' iff $\forall a \in E \setminus E', \exists b \in E' \setminus E$, such that $a \succ b$; we say that 'E is more elitist than E'' iff E is at least as elitist as E' and E' is not at least as elitist as E; we say that 'E is most elitist among A' iff $\nexists E'' \in A$ such that E'' is more elitist than E. $\sigma^{P,el}(AF_P)$ denotes the most elitist extensions among $\sigma((AR, AT))$ given *Pref*.

Note that we have aligned the notations of the approaches by Kaci *et al.* and Amgoud and Vesic to facilitate comparability. The example below illustrates how the extension-selection approaches to preference-based argumentation work.

EXAMPLE 7
Consider the PAF $AF_P = (AR, AT, Pref) = (\{a, b, c, d\}, \{(a, b), (b, a),$ $(a, c), (b, d)\}, \{a \succeq b, a \succeq c, a \succeq d, b \succeq d, b \succeq c\})$ (see Figure 6a for the argumentation graph) and complete semantics σ_{co}. $\sigma_{co}((AR, AT)) = \{\{\}, \{a, d\}, \{b, c\}\}$.

Using the extension-selection approaches, we can infer the following extensions from AF_P:

- $\sigma_{co}^{P,best}(AF_P) = \{\{\}, \{a, d\}, \{b, c\}\}$. Note that $\{\}$ is an extension because there exists no argument b' in $\{\}$, such that there exists an argument a' in any $E \in \sigma_{co}((AR, AT))$ and $a' \succ b'$;

- $\sigma_{co}^{P,dem}(AF_P) = \{\{a,d\}\}$. Intuitively, the empty set is the least democratic among the complete extensions and $\{a,d\}$ is more democratic than $\{b,c\}$ because $a \succ b$ and $a \succ c$;
- $\sigma_{co}^{P,el}(AF_P) = \{\{\}\}$. Somewhat counter-intuitively, only $\{\}$ is most elitist: the elitist condition applies the universal quantifier on the empty set, thus trivially satisfying the condition; conversely, for $\{a,d\}$ and $\{b,c\}$ the existential quantification of the condition is not satisfied because it must be applied to the empty set (roughly speaking).

For an example that does not depend on the empty set 'edge case' and that highlights the difference between $\sigma^{P,dem}$ and $\sigma^{P,el}$ on the one hand and $\sigma^{P,best}$ on the other, consider the PAF $AF'_P = (AR', AT', Pref') = (\{a,b,c,d\}, \{(a,b),(b,c),$ $(c,d),(d,a)\}, \{a \succeq b\})$ (see Figure 6b for the argumentation graph) and preferred semantics. Note that $\sigma_{pr}(AR', AT') = \{\{a,c\},\{b,d\}\}$. a is strictly preferred over b and this is the only established preference. But neither every argument in $\{a,c\}$ is strictly preferred over at least one argument in $\{b,d\}$ (or vice versa) nor for every argument a' in $\{b,d\}$, there is an argument in $\{a,c\}$ that is strictly preferred over a'. Hence, we have $\sigma_{co}^{P,dem}(AF'_P) = \sigma_{co}^{P,el}(AF'_P) = \{\{a,c\},\{b,d\}\}$ but $\sigma_{co}^{P,best}(AF_P) = \{\{a,c\}\}$.

3 Burden of Persuasion Frameworks

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

DEFINITION 26 (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 27 (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.

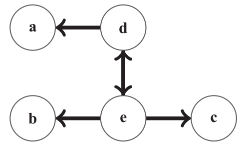


FIGURE 7. $AF' = (\{a,b,c,d,e\}, \{(d,a),(d,e),(e,b),(e,c),(e,d)\})$. $\subseteq$-maximal monotonic extensions of $\sigma_{pr}(AF')$ w.r.t. $\{\{a,b,c\}\}$: $\{a,e\}$ and $\{b,c,d\}$.

Let us introduce an example of a BPF.

EXAMPLE 8
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 28 ($\subseteq$-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 $EXTS_{mon}^{\subseteq-max}(EXTS, ES)$, as follows:

$$EXTS_{mon}^{\subseteq-max}(EXTS, ES) = \{E | E \in EXTS, \exists S \in ES \text{ such that } \nexists E' \in EXTS, E \cap S \subset E' \cap S\}.$$

Let us highlight that the notion of $\subseteq$-maximal monotonic extensions is purposefully different from the cardinality-based monotony measure and optimization approach [24] 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 $AF' = (\{a,b,c,d,e\}, \{(d,a),(d,e),(e,b),(e,c),(e,d)\})$ (Figure 7) and preferred semantics.

$\sigma_{pr}(AF) = \{\{a,b,c\}\}$ and $\sigma_{pr}(AF') = \{\{a,e\},\{b,c,d\},\}$; the only cardinality-maximal monotonic extension of $\sigma_{pr}(AF')$ w.r.t. $\sigma_{pr}(AF)$ is $\{b,c,d\}$, whereas we have two $\subseteq$-maximal monotonic extensions of $\sigma_{pr}(AF')$ w.r.t. $\{\{a,b,c\}\}$, i.e. $\{a,e\}$ and $\{b,c,d\}$. 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 abstract 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 9
Consider $EXTS = \{\{a,b\},\{\}\}$ and $ES = \{\{a\},\{c\}\}$. $EXTS^{\subseteq-max}_{mon}(EXTS, ES) = \{\{a,b\},\{\}\}$:

- For $E := \{a,b\}$, for every $S \in ES$ it holds that $\{a,b\} \cap S \not\subset \{\} \cap S$ and hence $\{a,b\} \in EXTS^{\subseteq-max}_{mon}(EXTS, ES)$.
- For $E := \{\}$, for $S := \{c\}$ it holds that $\{\} \cap S \not\subset \{a,b\} \cap S$ and hence $\{\} \in EXTS^{\subseteq-max}_{mon}(EXTS, ES)$.

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* ⊆-maximal monotonic extensions.

DEFINITION 29 (Pareto Optimal ⊆-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 ⊆-maximal monotonic extensions of *EXTS* w.r.t. *ES*, denoted by $EXTS^{\subseteq-max}_{po-mon}(EXTS, ES)$, as follows:

$$EXTS^{\subseteq-max}_{po-mon}(EXTS, ES) = \Big\{E \mid E \in EXTS \text{ and } \nexists E' \in EXTS \text{ such that } \underbrace{\big(\forall S \in ES, S \cap E \subseteq S \cap E'}_{\text{For every } S \text{ in } ES,\ E' \text{entails all arguments in } S \text{ that } E \text{ entails.}} \underbrace{\text{and } \exists S' \in ES \text{ such that } S' \cap E \subset S' \cap E'}_{\text{But for at least one } S' \text{ in } ES,\ E' \text{ entails strictly more arguments in } S' \text{than } E.}\big)\Big\}$$

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

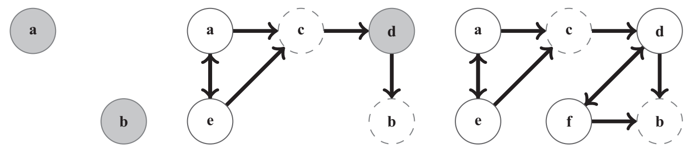
EXAMPLE 10
Consider again $EXTS = \{\{a,b\},\{\}\}$ and $ES = \{\{a\},\{c\}\}$. $EXTS^{\subseteq-max}_{po-mon}(EXTS, ES) = \{\{a,b\}\}$:

- For $E := \{\}$, we have $\{a,b\} \in EXTS$ such that $\forall S \in ES$ it holds that $S \cap E \subseteq S \cap \{a,b\}$ and also for $\{a\} \in ES$ it holds that $\{a\} \cap \{\} \subset \{a\} \cap \{a,b\}$. Hence, $\{\} \notin EXTS^{\subseteq-max}_{po-mon}(EXTS, ES)$.
- For $E := \{a,b\}$, we have—in contrast—no $E' \in EXTS$ such that $\exists S' \in ES$ such that $S' \cap E \subset S' \cap E'$ and hence $\{a,b\} \in EXTS^{\subseteq-max}_{po-mon}(EXTS, ES)$.

For further illustration purposes, let us showcase the approach using another example (where $EXTS^{\subseteq-max}_{po-mon}$ and $EXTS^{\subseteq-max}_{mon}$ happen to coincide).

EXAMPLE 11
Let us again consider the initial extensions $ES = \{\{a,b,c\}\}$ and the inference update after expansion $EXT = \{\{a,e\},\{b,c,d\}\}$ (analogous to the example we have in Figure 7). $EXTS^{\subseteq-max}_{po-mon}(EXTS, ES) = \{\{a,e\},\{b,c,d\}\}$; we observe that the only $S \in ES$ is $\{a,b,c\}$. Note that $\{a,b,c\} \cap \{a,e\} = \{a\}$ and $\{a,b,c\} \cap \{b,c,d\} = \{b,c\}$. Now consider:

- $E := \{a,e\}$. We have no $E' \in \{\{b,c,d\},\{a,e\}\}$ such that $\{a,b,c\} \cap \{a,e\} \subset \{a,b,c\} \cap E'$ and hence $E \in EXTS^{\subseteq-max}_{po-mon}(EXTS, ES)$;



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

FIGURE 8. Example: given $AF_{BP} = (\langle\{a,b\},\{c,d,e\},\{f\}\rangle, \{(a,c),(a,e),(c,d),(d,b),(d,f),(e,a),(e,c),(f,b),(f,d)\})$, the figure depicts AF_0, AF_1 and AF_2.

- $E := \{b,c,d\}$. We have no $E' \in \{\{b,c,d\},\{a,e\}\}$ such that $\{a,b,c\} \cap \{b,c,d\} \subset \{a,b,c\} \cap E'$ and hence $E \in EXTS^{\subseteq-max}_{po-mon}(EXTS, ES)$.

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

DEFINITION 30 (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:

$$\sigma^{BP}(AF_{BP}) =$$

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

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

EXAMPLE 12
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[6] and first provide an intuition that strays from the recursive definition (Definition 30). 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 8 depicts AF_0, AF_1 and AF_2. Then, we determine the CF2 extensions of AF_2 and AF_0:

- $\sigma_{SCF2}(AF_2) = \{\{a,d\},\{a,f\},\{e,d\},\{e,f\}\}$;
- $\sigma_{SCF2}(AF_0) = \{\{a,b\}\}$.

[6]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.

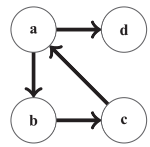


FIGURE 9. Differences between semantics families, by example.

$EXTS^{\subseteq-max}_{po-mon}(\sigma_{SCF2}(AF_2), \sigma_{SCF2}(AF_0)) = \{\{a,d\},\{a,f\}\}$. 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 30), we proceed as follows.

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. $\sigma^{BP}_{SCF2}(AF_{BP}) = EXTS^{\subseteq-max}_{po-mon}(\sigma^{BP}_{SCF2}(AF_{BP-1}), \sigma_{SCF2}(AF_0) \cup \sigma_{SCF2}(AF_1))$ $= EXTS^{\subseteq-max}_{po-mon}(\{\{a,d\},\{a,f\}\}, \{\{a,b\}\} \cup \{\{a,d\},\{e,d\}\}) = \{\{a,d\}\}$.

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. From the proof of Proposition 3 it can be verified 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.

Let us highlight that we have implemented the formal framework that is presented in this section as part of a Java-based argumentation reasoning library that is based on the Tweety project's abstract argumentation capabilities [27]. To keep the scope of this paper as general as possible, we abstain from providing any implementation details. Instead, we make the source code of the library, alongside with Application Programming Interface (API) documentation and comprehensive examples, available at https://git.io/JGueN.

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 3).[7] 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 $AF = (\{a, b, c, d\}, \{(a, b), (b, c), (c, a), (a, d)\})$ (depicted by Figure 9) as an example that highlights key differences.[8]

Admissibility-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 22): 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, admissibility-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 admissibility-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 admissibility-based semantics (see below) may be more suitable.

Weak admissibility-based semantics. Baumann *et al.* introduce the weak admissibility-based semantics family [10] to address a long-standing problem with admissibility-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 admissibility-based semantics achieve this behavior by systematically relaxing admissibility. The application of weak admissibility-based semantics may be useful in the context of burden of persuasion frameworks, given we want to ensure skepticism in face of odd cycles.

Naive set-based semantics. Naive set-based semantics, as initially introduced by Verheij [29] form the most credulous of the three semantics families; the naivety principle (see Definition 22) merely requires that every extension a semantics infers is a $\subseteq$-maximal conflict-free (*naive*) set.

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

[8] 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.* [4] (argumentation semantics overview) and Van der Torre and Vesic [28] (overview of argumentation principles).

By definition, every extension that an admissibility-based or weak admissibility-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 constraints 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[9] 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 these well-known limitations (see Example 4 and also Dvorak and Gaggl [20], as well as Cramer and Van der Torre [17]), 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:

- 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?
- 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 [30]); 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 BPFs and Argumentation Dynamics

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 [18] for a survey). At first glance, this connection may not be obvious. To formally establish the connection, let us first introduce the notion of strict (normal) expansions.

[9] In Example 4, consider AF^* and, in the case of stage semantics, also AF': the addition of a self-attacking 'dummy' argument leads to the credulous acceptance of its attacker, which one intuitively would want to reject, roughly speaking.

DEFINITION 31 (Strict (Normal) Expansions and Strict Expansion Chains).
Let $AF = (AR, AT)$ and $AF' = (AR', AT')$ be argumentation frameworks.

- AF' is a *strict expansion* of AF, denoted by $AF \preceq^S AF'$, iff $AF \preceq AF'$ and ($AR \subset AR'$ or $AT \subset AT'$).
- AF' is a *strict normal expansion* of AF, denoted by $AF \preceq_N^S AF'$, iff $AF \preceq_N AF'$ and $AR \subset AR'$.

An argumentation framework sequence $\langle AF_0, ..., AF_n \rangle$ is a *Strict Expansion Chain* (SEC) iff for $0 \leq i < n$ it holds true that $AF_i \preceq_N^S AF_{i+1}$.

Let us show that we can map every BPF to exactly one SEC and every SEC to exactly one BPF. For this, we first introduce two mapping functions. The first mapping function maps a BPF to a SEC.[10]

DEFINITION 32 (BPF-to-SEC Mapping).
The BPF-to-SEC mapping is a function that takes a burden of persuasion framework $AF_{BP} = (\langle S_0, ..., S_n \rangle, AT)$ as its input and returns a SEC AFS, denoted by $bs(AF_{BP})$, such that $AFS = \langle AF'_0, ..., AF'_n \rangle = \langle (AR'_0, AT'_0), ..., (AR'_n, AT'_n) \rangle$ and for $0 \leq i \leq n$, the following statement holds true:

$$AR'_i = AR_i \text{ and } AT'_i = AT \cap (AR_i \times AR_i).$$

The second mapping function maps a SEC to a BPF.

DEFINITION 33 (SEC-to-BPF-Mapping).
The SEC-to-BPF-mapping is a function that takes a SEC $AFS = \langle AF'_0, ..., AF'_n \rangle = \langle (AR'_0, AT'_0), ..., (AR'_n, AT'_n) \rangle$ as its input and returns a BPF AF_{BP}, denoted by $sb(AFS)$, such that $AF_{BP} = (\langle S_0, ..., S_n \rangle, AT)$, $S_0 = AR'_0$, $AT = AT'_n$, and for $0 < i \leq n$, the following statement holds true:

$$S_i = AR'_i \setminus AR'_{i-1}.$$

Given the *bs* and *sb* functions, we can now introduce and prove a representation theorem.

THEOREM 1
For every BPF AF_{BP}, it holds true that $sb(bs(AF_{BP})) = AF_{BP}$ and for every SEC AFS, it holds true that $bs(sb(AFS)) = AFS$.

Let us provide an example to illustrate this result.

EXAMPLE 13
Consider the BPF $AF_{BP} = (\langle \{a\}, \{b\}, \{c\} \rangle, \{(a,b), (b,a), (b,c), (c,b)\})$. $bs(AF_{BP})$ gives us the SEC $\langle AF_0, AF_1, AF_2 \rangle = \langle (\{a\}, \{\}), (\{a,b\}, \{(a,b), (b,a)\}), (\{a,b,c\}, \{(a,b), (b,a), (b,c), (c,b)\}) \rangle$. Roughly speaking, given this sequence (and an argumentation semantics σ), BP semantics applies an abstract argumentation semantics and returns $EXTS^{\subseteq -max}_{po-mon}(EXTS^{\subseteq -max}_{po-mon}(\sigma(AF_2), \sigma(AF_0)), \sigma(AF_0) \cup \sigma(AF_1))$.

Given these results, our burden of persuasion framework and argumentation semantics can be considered an application of dynamic argumentation approaches to preference-based argumentation (see the following section). Hence, we abstain—for the sake of conciseness—from a more elaborate

[10]Note that we use BPF short-hand notation (Definition 27) in this and the following definitions.

discussion of our results in the context of argumentation dynamics and instead focus on a formal comparison to preference-based argumentation approaches.

6 BPFs and Preference-based Argumentation

The formal framework we provide is fundamentally different from traditional approaches to modeling preferences in formal argumentation, such as preference-based [1] and value-based [13] argumentation (where value-based argumentation is a generalization of preference-based argumentation[11]). While the sequence of sets of arguments in a BPF can be considered 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*.

However, 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. To enable a precise comparison, let us generalize our approach so that it can be applied to any preference-based argumentation semantics.

DEFINITION 34 (BP Semantics for PAFs).
Let $AF_P = (AR', AT', \mathit{Pref})$ be a preference-based argumentation framework and let σ be an argumentation semantics. We define the burden of persuasion extensions of AF_P, denoted by $\sigma^{P,BP}(AF_P)$, as follows:

$$\sigma^{P,BP}(AF_P) = \{E \mid E \in \sigma^{BP}(AF_{BP}), AF_{BP} \in AFS_{BP}\},$$

where AFS_{BP} is the set of all BPFs $AF_{BP} = (\langle S_0, ..., S_n\rangle, AT)$ for which the following statement holds true:

$$AR_n = AR', AT = AT' \text{ and } \forall a, b \in AR' \text{ it holds true that}$$

$$\underbrace{(\text{if } a \succ b \text{ then } a \in S_i, b \in S_j, \text{ such that } 0 \leq i < j \leq n)}_{\text{If } a \text{ is strictly preferred over } b,\ a \text{ has a lower-level burden than } b.} \text{ and}$$

$$\underbrace{(\text{if } a \succeq b \text{ and } b \succeq a \text{ then } a, b \in S_k, \text{ such that } 0 \leq k \leq n)}_{\text{If } a \text{ is preferred over } b \text{ and vice versa, } a \text{ and } b \text{ have the same level of burden.}} \quad \text{and}$$

$$\underbrace{\begin{array}{l}(\text{for } S_i, S_j, \text{ such that } 0 \leq i < j \leq n, \\ \quad \exists c, d \in AR', \text{ such that } c \succ d \text{ (in } AF_p) \text{ and } c \in S_i, d \in S_j)\end{array}}_{\text{We have a minimal number of burdens, given the first condition as a constraint.}}$$

[11] For the sake of conciseness, we do not formally cover value-based argumentation in this paper. However, considering that value-based argumentation is applicable to legal reasoning [3], as well as the well-known close relation between value-based and preference-based argumentation, we speculate that the results presented in this work are relevant for value-based argumentation.

Intuitively, the BP semantics for preference-based argumentation frameworks considers, given a PAF and its set of arguments, all burden of persuasion frameworks whose total preorder on the set of arguments is a superset of (or equal to) the (partial or total) preorder on the set of arguments as established by the PAF's preference relation and that have a minimal number of levels of burdens.[12] The example below illustrates how the generalization works.

EXAMPLE 14
Consider the PAF $AF_P = (AR, AT, Pref) = (\{a,b,c\}, \{(a,b),(b,a),$ $(a,c)\}, \{c \succeq a\})$. Given AF_P, we can construct the following BPFs:

1. $AF_{BP} = (\langle\{b,c\},\{a\}\rangle, AT)$;
2. $AF'_{BP} = (\langle\{c\},\{a,b\}\rangle, AT)$.

Let us apply preferred semantics. $\sigma_{pr}^{P,BP}(AF_P) = \sigma_{pr}^{BP}(AF_{BP}) \cup \sigma_{pr}^{BP}(AF'_{BP}) = \{\{b,c\}\}$.

Let us argue that the formal notion of a burden of persuasion framework is still useful, as it is better aligned with our practical motivation, allows for a more concise notation of BPFs and covers the dynamic nature of BP semantics more intuitively. In contrast, the generalization to preference-based argumentation frameworks/semantics allows for a better comparison with related research. In particular, it is crucial to highlight and motivate the differences to other approaches that allow for the preference-based selection of extensions an argumentation semantics yields given an abstract argumentation framework, as introduced by Kaci *et al.* [22], as well as by Amgoud and Vesic [2] (see Definition 25). We can achieve this by introducing another example.

EXAMPLE 15
Let us consider the following legal reasoning scenario, in the context of infringement proceedings in a patent dispute case. In the proceedings, different experts provide testimonies. Expert e_c claims that the method applied by the case's defendant has already been released into the public domain in a publication that predates the patent (argument c). Expert e_d claims that the publication describes a method that is substantially different with respect to its application domains (argument d; c and d attack each other). Expert e_a claims that the pattern does not cover the technology domain in question and that hence, there is no conflict between the patent and the method that has been released into the public domain (argument a, a attacks c). Expert e_b claims the opposite (argument b, b attacks d and a and b attack each other).

As the claimant has to establish the facts supporting that infringing acts have been executed by the defendant, the burden of persuasion rests on the arguments b and d. We can model this scenario by creating the PAF $AF_P = (AR, AT, Pref) = (\{a,b,c,d\}, \{(a,b),(a,c),(b,a),(b,d),(c,d),(d,c)\},$ $\{a \succeq b, a \succeq c, a \succeq d, b \succeq d, c \succeq a, c \succeq b, c \succeq d, d \succeq b\})$ and preferred semantics (see Figure 10 for the argumentation graph). $\sigma_{pr}((AR, AT)) = \{\{a,d\},\{b,c\}\}$. The different extension-selecting approaches to preference-based argumentation yield the following extensions:

- $\sigma_{pr}^{P,best} = \sigma_{pr}^{P,dem} = \sigma_{pr}^{P,el} = \{\{a,d\},\{b,c\}\}$;
- $\sigma_{pr}^{P,BP} = \{\{a,d\}\}$.

[12] Let us note that the preference relation of a PAF is by definition transitive, which is important for this last requirement of Definition 34.

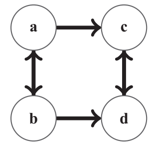


FIGURE 10. Example: assuming (e.g.) preferred semantics and the preferences $a \succeq b$, $a \succeq c$, $a \succeq d$, $b \succeq d$, $c \succeq a$, $c \succeq b$, $c \succeq d$, $d \succeq b$, one would intuitively expect that a successfully defeats b and hence it also defeats c; consequently, the only possible inference (extension) we can draw from this argumentation framework is $\{a, d\}$.

From the perspective of the burden of persuasion, we can say that intuitively, a is unburdened and hence the burdened argument b is defeated by a; consequently, a can defeat the unburdened argument c and hence defend d. In our scenario, the claim that an infringing act was executed by the defendant cannot be established.

To allow for a precise, principle-based comparison, let us define *extension-selecting preference-based argumentation semantics* to then introduce a principle for such semantics.

DEFINITION 35 (Extension-Selecting Preference-based Argumentation Semantics).
Let σ be an argumentation semantics. An extension-selecting preference-based argumentation semantics is a function σ^P that takes a PAF $AF_P = (AR, AT, Pref)$ and returns $ES \subseteq \sigma((AR, AT))$.

Now, we can define the *preference directionality* principle for extension-selecting preference-based argumentation semantics that reflects the behaviors illustrated by Example 15.

DEFINITION 36 (Preference Directionality).
An extension-selecting preference-based argumentation semantics σ^P satisfies the *preference directionality* principle iff the following statement holds true for every universally defined (abstract) argumentation semantics σ, for every PAF $AF_P = (AR, AT, Pref)$, $\forall E, E' \in \sigma((AR, AT))$:

$$\text{if } \forall E'' \in \sigma((AR, AT) \downarrow_{AR^*}), E'' \cap E \supset E'' \cap E' \text{ holds true}$$
$$\text{then } E \in \sigma^P(AF_P) \text{ and } E' \notin \sigma^P(AF_P),$$

where $\subseteq AR^* = \{c | c \in AR, \forall d \in AR, c \succeq d \text{ (in } AF_P) \}$.

Intuitively, the principle stipulates that given a PAF, an extension-selecting preference-based argumentation semantics must infer only extensions that contain a maximal subset of the arguments (w.r.t. other extensions that could potentially be selected) that the semantics may infer from the restriction of the abstract argumentation framework to all most preferred arguments. It is worth highlighting that the principle can potentially be tightened to better deal with *partial* preference preorders: consider e.g. the PAF in the previous example but assume that no preferences between a and c, as well as d and c, have been established; then, the intuition that a should successfully defeat b still applies, but is no longer enforced by the principle, as a is no longer most preferred. Defining a tighter principle can be considered relevant future research; here, we abstain from it for the sake of conciseness, because the principle above is sufficient as a motivator of burden-of-persuasion semantics for preference-based argumentation, distinguishing these semantics from previously established approaches.

Note that the principle must be satisfied given any *universally defined* argumentation semantics σ that an extension-selecting preference-based argumentation semantics may apply. We can show that the burden of persuasion semantics for PAFs satisfies the preference directionality principle.

THEOREM 2
The extension-selecting semantics for PAFs $\sigma^{P,BP}$ satisfies the preference directionality principle.

In contrast, we can show that the extension-selecting preference-based argumentation semantics $\sigma^{P,best}$, $\sigma^{P,dem}$ and $\sigma^{P,el}$ that are based on the works of Kaci *et al.* [22] and of Amgoud and Vesic [2], respectively, do not satisfy the preference directionality principle.

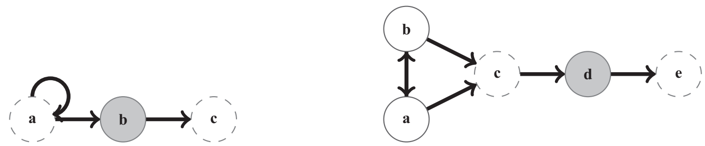
PROPOSITION 3
The extension-selecting semantics for PAFs $\sigma^{P,best}$, $\sigma^{P,dem}$ and $\sigma^{P,el}$ do not satisfy the preference directionality principle.

7 Burden of Persuasion Frameworks in the Context of Related Legal Reasoning Approaches

From a legal perspective, let us note that the burden of persuasion is related to, but different from, the *standard of persuasion* [21], which, from a formal argumentation perspective, relates more directly to the strength that is required for the defeat of an argument by one or several attackers. While the exact meaning of the standard of persuasion and its relation to the burden of persuasion depends on the exact jurisdiction and the legal application domain, we argue that from a formal argumentation perspective the distinction between the burden of persuasion as a tool for managing extension-selection and the standard of persuasion as a tool for managing argument strength is pragmatic and useful, given related existing work on the burden of persuasion in formal argumentation [15]. Modeling standards of persuasion in formal argumentation considering the large body of research on argumentation strength (see, e.g. Beirlaen *et al.* [11]) is certainly interesting future work—and can e.g. take the legal reasoning perspective of argumentation-based standards of *proofs* as provided by Calegari and Sartor [16] as a point of departure—but is 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 [14]. 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 say that a labeling of AF and S is a total function $L_S : AR \rightarrow \{IN, OUT, UND\}$ and for $a \in AR$, $X \in \{IN, OUT, UND\}$ we say that 'a is (labeled) X' iff $L_S(a) = X$ and define $X_S(AF) := \{a | L_S(a) = X\}$.
3. We say that a labeling of AF and S is a *Burden of Persuasion Labeling* (BPL) iff $\forall a \in AR$ the following holds true:
   a **is** IN **iff** $\forall b \in \{a\}^-$ it holds that b is OUT;
   a **is** OUT **iff** *i)* $a \in S$ and $\exists b \in \{a\}^-$ such that (b is IN or b is UND) or *ii)* $a \notin S$ and $\exists b \in \{a\}^-$ such that b is IN;
   a **is** UND, otherwise.



(a) AF. **(b)** AF'.

FIGURE 11. Examples: comparison to Calegari *et al.*'s approach.

4. A BPL is a grounded BPL if $UND_S(AF)$ is $\subseteq$-maximal.

Below, we consider the grounded BPL.

The approach is fundamentally different from the one introduced in our paper. Below we provide two examples, highlighting nuanced differences. Then, we demonstrate fundamental shortcomings of the approach by Calegari *et al.* regarding the handling of odd cycles.

Self-attacking arguments. Take 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*: allowing a to be *IN* would mean that an argument that is *IN* attacks a; (ii) because b carries the burden of persuasion and is attacked by the undecided and unburdened argument a, it is *OUT*; (iii) hence, c is *IN*. This is problematic, because a as a self-attacking 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: $AF_{BP} = (\langle\{a, c\}, \{b\}\rangle, AT)$. $\sigma^{BP}_{SCF2}(AF_{BP}) = \{\{b\}\}$; 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. Intuitively, we argue that a self-attacking argument is nonsensical and hence should never defeat another argument, even if this argument is burdened.

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 *UND*, but both arguments *consistently* attack c. Still, because c is unburdened, it is *UND* as well; hence, the burdened argument d is out and the unburdened argument e is *IN*. However, we may 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. There, 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 $\sigma^{BP}_{SCF2}(AF'_{BP}) = \{\{a, d\}, \{b, d\}\}$.

One could, of course, argue that in some scenarios, contradicting statements that attack each other—and are hence *inconsistent*—but *consistently* attack a burdened argument are not sufficient for rejecting the latter argument. Let us informally claim that given the semantics we introduce in the *Preliminaries* section, our burden of persuasion approach does not support this behavior. However, this behavior can be achieved using an SCC-recursively defined semantics that uses grounded semantics on SCC-level; i.e. achieving such behavior is not a limitation of our approach to drawing inferences from argumentation frameworks with burdened arguments, but rather a limitation of the abstract argumentation semantics on which we rely in this paper. In contrast, the approach by Calegari *et al.* has a limitation with respect to its generalizability to argumentation frameworks with odd cycles,[13] on which we elaborate below. For the sake of conciseness, we abstain from a rigid formalization of a potential generalization and merely use a semi-formal example. Consider the argumentation framework $AF = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$ (a simple three-cycle of arguments with uni-directional attacks) and assume the burden of persuasion is placed on two of the arguments, let us say on a and b. Now, c is unburdened and hence can only be OUT if b is IN (see Point 3, OUT, (ii) in our description of the approach by Calegari *et al.*); this is not possible because if b is IN then a must be OUT and hence c must be UND or IN (see Point 3, OUT, (i) above). Assuming c is UND or IN means that a must be OUT. But from this it follows that b must be IN (see Point 3, IN) and hence c must be OUT, which contradicts that c is UND or IN. Thus, we can conclude that the definition provided by Calegari *et al.* leads to a contradiction, i.e. this case cannot be handled properly.

8 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 burden, 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 and can be generalized to form an extension-selecting approach to preference-based argumentation.

Acknowledgements

We thank the anonymous reviewers for their thoughtful and immensely useful feedback that has greatly improved the technical presentation of this paper. 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] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Journal of Automated Reasoning*, **29**, 125–169, 2002.
[2] Leila Amgoud and Srdjan Vesic. Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning*, **55**, 585–606, 2014.
[3] Katie Atkinson and Trevor JM Bench-Capon. Value-based argumentation. *IfCoLog Journal of Logics and Their Applications*, **8**, 1543–1588, 2021.
[4] P. Baroni, M. Caminada and M. Giacomin. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation. College Publications, chapter 4*, P. Baroni,

[13] Note that the issue highlighted by the first example we have provided above exists with any admissibility-based semantics; in contrast, the issue highlighted by the second example could be solved by applying an admissibility-based semantics that is more credulous than grounded semantics.

D. Gabbay, G. Massimiliano and L. van der Torre., eds, pp. 159–236. College Publications, 2018.
[5] P. Baroni, D. M. Gabbay, M. Giacomin and L. van der Torre. *Handbook of Formal Argumentation*. College Publications, 2018.
[6] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, **171**, 675–700, 2007. Argumentation in Artificial Intelligence.
[7] Pietro Baroni, Massimiliano Giacomin and Giovanni Guida. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, **168**, 162–210, 2005.
[8] Ringo Baumann. On the nature of argumentation semantics: existence and uniqueness, expressibility, and replaceability. *Journal of Applied Logics*, **4**, 2779–2886, 2017.
[9] R. Baumann and G. Brewka. Expanding argumentation frameworks: enforcing and monotonicity results. *COMMA*, **10**, 75–86, 2010.
[10] R. Baumann, G. Brewka and M. Ulbricht. Revisiting the foundations of abstract argumentation-semantics based on weak admissibility and weak defense. In *AAAI*, vol. 34, pp. 2742–2749, 2020.
[11] Mathieu Beirlaen, Jesse Heyninck, Pere Pardo and Christian Straßer. Argument strength in formal argumentation. *IfCoLog Journal of Logics and Their Applications*, **5**, 629–676, 2018.
[12] T. Bench-Capon, H. Prakken and G. Sartor. Argumentation in legal reasoning. In *Argumentation in artificial intelligence*, pp. 363–382. Springer, 2009.
[13] Trevor JM Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, **13**, 429–448, 2003.
[14] R. Calegari, R. Riveret and G. Sartor. The burden of persuasion in structured argumentation. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*. ICAIL '21, New York, NY, USA, 2021. Association for Computing Machinery.
[15] Roberta Calegari and Giovanni Sartor. Burden of persuasion in argumentation. arXiv preprint arXiv:2009.10244, 2020.
[16] R. Calegari and G. Sartor. Burdens of persuasion and standards ofÂ proof in structured argumentation. In *Logic and Argumentation*, P. Baroni, C. Benzmüller and Y. N. Wáng., eds, pp. 40–59. Springer International Publishing, Cham, 2021.
[17] M. Cramer and L. van der Torre. 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.
[18] S. Doutre and J.-G. Mailly. Constraints and changes: a survey of abstract argumentation dynamics. *Argument & Computation*, **9**, 223–248, 2018.
[19] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, **77**, 321–357, 1995.
[20] Wolfgang Dvořák and Sarah Alice Gaggl. Stage semantics and the SCC-recursive schema for argumentation semantics. *Journal of Logic and Computation*, **26**, 1149–1202, 022014.
[21] R. D. Friedman. *Standards of Persuasion and the Distinction Between Fact and Law*, vol. 86, p. 916. Northwestern University Law Review, 1991.
[22] S. Kaci, L. van der Torre and S. Villata. Preference in abstract argumentation. In *Computational Models of Argument*, pp. 405–412. IOS Press, 2018.
[23] A.C. Kakas and P. Mancarella. On the semantics of abstract argumentation. *Journal of Logic and Computation*, **23**, 991–1015, 2013.

[24] Timotheus Kampik and Dov Gabbay. The 'Degrees of Monotony'-dilemma in abstract argumentation. In Jiřina Vejnarová and Nic Wilson., eds, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty 2021*, to appear. Cham, 2021. Springer International Publishing.
[25] T. Kampik, D. Gabbay and G. Sartor. The burden of persuasion in abstract argumentation. In *Logic and Argumentation*, P. Baroni, C. Benzmüller and Y. N. Wáng., eds, pp. 224–243. Springer International Publishing, Cham, 2021.
[26] Henry Prakken and Giovanni Sartor. A logical analysis of burdens of proof. *Legal Evidence and Proof: Statistics, Stories, Logic*, pp. 223–253, 2009.
[27] M. Thimm. 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*, KR'14, pp. 528–537. AAAI Press, 2014.
[28] Leon van der Torre and Srdjan Vesic. The principle-based approach to abstract argumentation semantics. *IfCoLog Journal of Logics and Their Applications*, **4**, 2735–2778, 2017.
[29] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. *Proc. NAIC*, **96**, 357–368, 1996.
[30] Wu Yining and Martin Caminada. A labelling-based justification status of arguments. *Studies in Logic*, **3**, 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 30) 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 29), 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. □

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 30) 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 29), 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. □

THEOREM 1
For every BPF AF_{BP}, it holds true that $sb(bs(AF_{BP})) = AF_{BP}$ and for every SEC AFS, it holds true that $bs(sb(AFS)) = AFS$.

PROOF. We can split the proof into two cases:

Case 1: the proof that for every BPF AF_{BP}, $sb(bs(AF_{BP})) = AF_{BP}$ holds true. Let $AF_{BP} = (\langle S_0, ..., S_n\rangle, AT)$ be a BPF. By definition of sb (Definition 32), it holds true that $sb(AF_{BP}) = AFS$, where AFS is a SEC $\langle AF'_0, ..., AF'_n\rangle = \langle (AR'_0, AT'_0), ..., (AR'_n, AT'_n)\rangle$, such that for $0 \leq i \leq n$, the following statement holds true:

$$AR'_i = AR_i \text{ and } AT'_i = AT \cap (AR_i \times AR_i).$$

Because by definition of bs (Definition 33), it holds true that $bs(AFS) = AF'_{BP} = \langle S'_0, ..., S'_n\rangle, AT')$, such that $S'_0 = AR'_0$, $AT' = AT_n$, and for $0 < i \leq n$, $S'_i = AR_i \setminus AR_{i-1}$ holds true, it follows that $sb(bs(AF_{BP})) = AF_{BP}$ holds true, which proves the proposition for this case.

Case 2: the proof that for every SEC AFS, $bs(sb(AFS)) = AFS$ holds true.

Case 2: Let $AFS = \langle AF'_0, ..., AF'_n\rangle = \langle (AR'_0, AT'_0), ..., (AR'_n, AT'_n)\rangle$ be a SEC. By definition of bs (Definition 33), it holds true that $bs(AFS) = AF_{BP}$, where AF_{BP} is a BPF, such that $AF_{BP} = (\langle S_0, ..., S_n\rangle, AT)$, $S_0 = AR_0$, $AT = AT_n$, and for $0 < i \leq n$, $S_i = AR_i \setminus AR_{i-1}$ holds true. Because by definition of sb (Definition 32), it holds true that $sb(AF_{BP}) = AFS'$, where AFS' is a SEC $\langle AF''_0, ..., AF''_n\rangle = \langle (AR''_0, AT''_0), ..., (AR''_n, AT''_n)\rangle$, such that for $0 \leq i \leq n$, $AR''_i = AR_i$ and $AT''_i = AT \cap (AR_i \times AR_i)$ hold true, $bs(sb(AFS)) = AFS$ holds true, which proves the proposition. □

THEOREM 2
The extension-selecting semantics for PAFs $\sigma^{P,BP}$ satisfies the preference directionality principle.

PROOF. We provide a proof by contradiction.

1. Let σ be a universally defined argumentation semantics. By definition of the preference directionality principle (Definition 36), preference directionality is violated by $\sigma^{P,BP}$ iff the following statement—which we assume for the contradiction—holds true for a PAF $AF_P = (AR, AT, Pref)$:

$$\exists E, E' \in \sigma((AR, AT)), \text{ such that}$$
$$\forall E'' \in \sigma((AR, AT) \downarrow_{AR^*}), E'' \cap E \supset E'' \cap E' \text{ holds true}$$
$$\text{and } (E \notin \sigma^{P,BP}(AF_P) \text{ or } E' \in \sigma^{P,BP}(AF_P)),$$

where $AR^* = \{c | c \in AR, \forall d \in AR, c \succeq d \text{ (in } AF_P) \}$.

2. It follows from 1. that given $AR^* = \{c|c \in AR, \forall d \in AR, c \succeq d\}$, it holds true that $\exists E, E' \in \sigma((AR, AT))$, such that $E \notin EXTS^{\subseteq-max}_{po-mon}(\sigma((AR, AT)), \sigma((AR, AT) \downarrow_{AR^*}))$ or $E' \in EXTS^{\subseteq-max}_{po-mon}(\sigma((AR, AT)), \sigma((AR, AT) \downarrow_{AR^*}))$ (consider the definitions of a burden of persuasion semantics, as well as of a burden of persuasion semantics for preference-based argumentation frameworks, i.e. Definitions 30 and 34), although it holds true that $\forall E'' \in \sigma((AR, AT) \downarrow_{AR^*}) E'' \cap E \supset E'' \cap E'$.
3. The statement in 2. is impossible considering the definition of Pareto-optimal ⊆-maximal monotonic extensions (Definition 29), which establishes the contradiction and hence proves the proposition. □

PROPOSITION 3
The extension-selecting semantics for PAFs $\sigma^{P,best}$, $\sigma^{P,dem}$ and $\sigma^{P,el}$ do not satisfy the preference directionality principle.

PROOF. We provide a proof by counter-example. Consider the PAF $AF_P = (AR, AT, Pref) = (\{a,b,c,d\}, \{(a,b),(a,c),(b,a),(b,d),(c,d),(d,c)\}, \{a \succeq b, a \succeq c, a \succeq d, b \succeq d, c \succeq a, c \succeq b, c \succeq d, d \succeq b\})$ (see Example 15), preferred semantics σ_{pr}, and $\sigma_{pr}^{P,x}$, where $x \in \{best, dem, el\}$. Note that σ_{pr} is universally defined [8]. $\sigma_{pr}^{P,x} = \{\{a,d\},\{b,c\}\}$. It follows that given $AR^* = \{c'|c' \in AR, \forall d' \in AR, c' \succeq d' \text{ (in } AF_P)\} = \{a,c\}$ and $\sigma_{pr}((AR, AT) \downarrow_{AR^*}) = \{\{a\}\}$, for $\{a,d\},\{b,c\} \in \sigma_{pr}((AR, AT))$, the following statement holds true:

$$\forall E'' \in \sigma((AR, AT) \downarrow_{AR^*}), E'' \cap \{a,d\} \supset E'' \cap \{b,c\}$$

$$\text{and } E \in \sigma_{pr}^{P,x}(AF_P) \text{ and } E' \in \sigma_{pr}^{P,x}(AF_P).$$

Consequently, by definition (Definition 36), the preference directionality principle is violated. This proves the proposition. □

Received 16 January 2022