A General Account of Argumentation with Preferences

Sanjay Modgil\(^1\) and Henry Prakken\(^2\)

1. Department of Informatics, King’s College London (sanjay.modgil@kcl.ac.uk)
2. Department of Information and Computing Sciences, Utrecht University, and Faculty of Law, University of Groningen (henry@cs.uu.nl)

April 3, 2012

Abstract

This paper builds on the recently proposed \textit{ASPIC}^+ formalism, to develop a general framework for argumentation with preferences. We motivate a revised definition of conflict free sets of arguments for \textit{ASPIC}^+, adapt \textit{ASPIC}^+ to accommodate a broader range of instantiating logics, and show that the resulting framework satisfies key properties and rationality postulates. We then show that the generalised framework accommodates any Tarskian logic extended with preferences, and then study instantiations of the framework by classical logic approaches to argumentation. We conclude by arguing that \textit{ASPIC}^+’s modelling of defeasible inference rules further testifies to the generality of the framework, and then examine and counter recent critiques of Dung’s framework and its extensions to accommodate preferences.

1 Introduction

Argumentation is a key topic in the logical study of nonmonotonic reasoning and the dialogical study of inter-agent communication [12, 41]. Argumentation is a form of reasoning that makes explicit the reasons for the conclusions that are drawn and how conflicts between reasons are resolved. This provides a natural mechanism to handle inconsistent and uncertain information and to resolve conflicts of opinion between intelligent agents. In logical models of nonmonotonic reasoning, the argumentation metaphor has proved to overcome some drawbacks of other formalisms. Many of these have a mathematical nature that is remote from how people actually reason, which makes it difficult to understand and trust the behaviour of an intelligent system. The argumentation approach bridges this gap by providing logical formalisms that are rigid enough to be formally studied and implemented, while at the same time being close enough to informal reasoning to be understood by designers and users.

Many theoretical and practical developments build on Dung’s seminal theory of abstract argumentation [23]. A Dung argumentation framework (AF) consists of a conflict-based binary \textit{attack} relation \( C \) over a set of arguments \( A \). The justified arguments are then evaluated based on subsets of \( A \) (extensions) defined under a range of semantics. The arguments in an extension are required to not attack each other (extensions are \textit{conflict free}), and attack any argument that in turn attacks an argument
in the extension (extensions reinstate/defend their contained arguments). Dung’s theory has been developed in many directions, including argument game proof theories [21, 24, 33, 44] to determine extension membership of a given argument. Also, several works augment AFs with preferences and/or values [5, 11, 32, 39], so that the conflict-free extensions, and so justified arguments, are evaluated only with respect to the successful attacks (defeats), where an argument $X$ is said to defeat an argument $Y$ iff $X$ attacks $Y$ and $Y$ is not preferred to $X$.

The widespread impact of Dung’s work can partly be attributed to its level of abstraction. AFs can be instantiated by a wide range of logical formalisms; one is free to choose a logical language $\mathcal{L}$ and define what constitutes an argument and attack between arguments defined by a theory in $\mathcal{L}$. The theory’s inferences can then be defined in terms of the conclusions of the theory’s justified arguments. Indeed, the inference relations of existing logics, including logic programming and non-monotonic logics (e.g. default, auto-epistemic and defeasible logic) have been given argumentation based characterisations [15, 22, 23, 27]. Dung’s theory thus provides a dialectical semantics for these logics, and the above-mentioned argument games can be viewed as alternative dialectical proof theories for these logics. The fact that reasoning in existing non-monotonic logics can thus be characterised, testifies to the generality of the dialectical principles of attack and reinstatement; principles that are also both intuitive and familiar in human modes of reasoning, debate and dialogue. Argumentation theory thus provides a characterisation of both human and logic-based reasoning in the presence of uncertainty and conflict, through the abstract dialectical modelling of the process whereby arguments can be moved to attack and reinstate/defend other arguments. The theory’s value can therefore in large part be attributed to its explanatory potential for making non-monotonic reasoning processes inspectable and readily understandable for human users, and its underpinning of dialogical and more general communicative interactions that may involve heterogeneous (human and software) agents reasoning in the presence of uncertainty and conflict.

More recently, the ASPIC framework [3] was developed in response to the fact that the abstract nature of Dung’s theory precludes giving guidance so as to what kinds of instantiations satisfy intuitively rational properties. ASPIC was not designed from scratch but was meant to integrate, generalise and further develop existing work on structured argumentation, partly originating from before Dung’s seminal paper (e.g.[37, 40, 42, 45]). ASPIC adopts an intermediate level of abstraction between Dung’s fully abstract level and concrete instantiating logics, by making some minimal assumptions on the nature of the logical language and the inference rules and then providing abstract accounts of the structure of arguments, the nature of attack, and the use of preferences. [18] then formulated consistency and closure postulates that cannot be formulated at the abstract level, and showed these postulates to hold for a special case of ASPIC; one in which preferences were not accounted for. In [39], ASPIC$^+$ then generalised ASPIC to accommodate a broader range of instantiations (including assumption-based argumentation [15] and systems using argument schemes), and showed that the postulates were satisfied when applying preferences and evaluating the justified arguments on the basis of defeats.

In this paper we build on and modify [39]’s ASPIC$^+$ framework, to develop a more general structured framework for argumentation with preferences. We make three main contributions. Firstly, we motivate a revised definition of conflict free sets of arguments for ASPIC$^+$, adapt ASPIC$^+$ to accommodate a broader range of instantiating logics, and then show that the resulting framework satisfies the key properties and postulates in [23] and [18]. Secondly, we formalise instantiation of the new framework by
Tarskian (and in particular classical) logics extended with preferences, and so demonstrate that such instantiations satisfy [18]’s rationality postulates. Thirdly, we examine and counter recent critiques of Dung’s framework and its extensions to accommodate preferences.

With regard to the first contribution, Section 2 presents the conceptual foundations for our framework in the context of the above value proposition of argumentation as providing a bridging role between formal logic and human modes of reasoning. Specifically, we: i) posit criteria for defining attack relations, given their dual role in declaratively denoting the mutual incompatibility of the information contained in the attacking arguments, and their dialectical use; ii) motivate the distinction between preference dependent and preference independent attacks, where only the former’s use in a dialectical context (as defeats) should be contingent upon preferences; iii) argue that unlike current approaches [5, 11, 32], including [39]’s ASPIC+, it is conceptually more intuitive to define conflict-free sets in terms of those that do not contain attacking arguments, so that defeats are only deployed dialectically. Section 3 then revisits and generalises [39]’s ASPIC+ framework in light of Section 2’s conceptual foundations. Firstly, the new notion of conflict-free is adopted. Secondly, we extend [39]’s ASPIC+ framework to accommodate instantiation by arguments with consistent premises, thus generalising the framework to accommodate a broader range of instantiations. Section 4 then present key technical results. We show that Section 3’s revised and generalised ASPIC+ satisfies properties of Dung’s theory and [18]’s rationality postulates.

Section 5 then presents the second main contribution, and in so doing testifies to the generality of the framework proposed in this paper. Firstly, [39] defines preferences over arguments, based on pre-orderings over the arguments’ constituent rules and premises, and then shows that the thus defined preferences satisfy intuitive properties that ensure satisfaction of rationality postulates. In Section 5 we show that these properties are additionally satisfied by ways of defining preference that are not considered in [39]. Secondly, Amgoud & Besnard [1, 2] recently defined a structured ‘abstract logic’ approach to argumentation, whereby they consider instantiations of Dung’s framework by Tarskian logics. We reconstruct this approach as an instance of Section 3’s ASPIC+ framework. Furthermore, we extend [1, 2]’s abstract logic approach with preferences, and reconstruct this extension in ASPIC+. Given Section 4’s results, these reconstructions imply that we are the first to show satisfaction of [18]’s rationality postulates for Tarskian logic instantiations with and without preferences. Thirdly, we reconstruct classical logic approaches to argumentation [13, 14, 26], including those that additionally accommodate preferences [5]. To the best of our knowledge, we are the first to prove [18]’s postulates for classical logic approaches with preferences. Fourthly, we show a correspondence between a particular classical logic instantiation of Section 3’s ASPIC+ framework and Brewka’s preferred subtheories [16].

Section 6 concludes with a detailed discussion of related work, and so presents the third main contribution of our work. Specifically, we compare the generality of ASPIC+ with that of the abstract logic proposal for structured argumentation, and argue that the latter’s generality is compromised in that it only applies to deductive (e.g., classical logic) approaches to argumentation. In particular, it does not apply to mixed deductive and defeasible argumentation, where the latter requires modelling of defeasible inference rules. We additionally argue that the inclusion of defeasible inference rules in models of argumentation is required by the value proposition of argumentation as bridging the gap between formalisms and human reasoning, as defeasible reasons are an essential ingredient of human reasoning. Section 6 also counters a number of recent criticisms of Dung’s abstract approach, as well as recent critiques of Dung’s approach...
extended with preferences. We claim that our paper shows that a proper modelling of the use of preferences requires making the structure of arguments explicit.

## 2 Logic, Argumentation and Preferences

### 2.1 Background

A *Dung argumentation framework* (AF) [23] is a tuple $(\mathcal{A}, C)$, where $C \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation on the arguments $\mathcal{A}$. $S \subseteq \mathcal{A}$ is then said to be *conflict free* iff $\forall X, Y \in S, (X, Y) \notin C$. The status of arguments is then evaluated as follows:

**Definition 1** Let $(\mathcal{A}, C)$ be a AF. For any $X \in \mathcal{A}$, $X$ is acceptable with respect to some $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $(Y, X) \in C$ implies $\exists Z \in S$ s.t. $(Z, Y) \in C$. Let $S \subseteq \mathcal{A}$ be *conflict free*. Then:

- $S$ is an *admissible* extension iff $X \in S$ implies $X$ is acceptable w.r.t. $S$;
- $S$ is a *complete* extension iff $X \in S$ iff $X$ is acceptable w.r.t. $S$;
- $S$ is a preferred extension iff it is a set inclusion maximal complete extension;
- $S$ is the grounded extension iff it is the set inclusion minimal complete extension;
- $S$ is a stable extension iff it is preferred and $\forall Y \notin S, \exists X \in S$ s.t. $(X, Y) \in C$.

For $T \in \{\text{complete, preferred, grounded, stable}\}$, $X$ is *sceptically or credulously* justified under the $T$ semantics if $X$ belongs to all, respectively at least one, $T$ extension.

A number of works [5, 11, 32] augment AFs to formalise the role of the relative strengths of arguments at the *abstract* level. The basic idea in all these works is that an attack by $X$ on $Y$ succeeds as a *defeat* only if $Y$ is not stronger than $X$. For example, preference-based AFs (PAFs) [5] are tuples $(\mathcal{A}, C, \preceq)$, where given the preordering $\preceq \subseteq \mathcal{A} \times \mathcal{A}$, $Y$ is stronger than $X$ iff $Y$ is strictly preferred to $X$ ($X \prec Y$ iff $X \preceq Y$ and $Y \preceq X$). In [32], preferences between arguments are not based on a given preordering, but rather are themselves defeasible and possibly conflicting, and so are themselves the conclusions of arguments. In [39]’s *ASPIC*+ framework, arguments are defined by strict and defeasible rules and premises expressed in some abstract language. Attacks between arguments are defined, and a preference relation over arguments is used to derive a defeat relation. Unlike PAFs and [11]’s value based AFs, ASPIC+’s use of preferences to define defeat takes the structure of arguments into account.

In all the above approaches, the justified arguments are then evaluated on the basis of the derived defeat relation, rather than the original attack relation. In other words, a conflict free set is one that contains no two defeating arguments, and the defeat relation replaces the attack relation $C$ in Definition 1.

In Section 1 we discussed how abstract argumentation theory, and argument game proof theories: a) provide dialectical semantics, respectively proof theories, for non-monotonic reasoning, where; b) the abstract modelling of the process whereby arguments are submitted to attack and defend, comports with intuitive human modes of reasoning and debate. Thus, the added value of argumentation is in large part due to its potential for facilitating dynamic, interactive and heterogenous (both automated and human) reasoning in the presence of uncertain and conflicting knowledge.

It is in this context that we motivate criteria for defining attack relations, the role of preferences, and a new approach to defining the extensions of a framework in terms...
of both defeat and attack relations. In what follows we will assume that arguments are built from strict and defeasible knowledge (we defend the need for such a distinction in Section 6), and refer to an argument’s conclusion following from its constituent premises and rules (referred to collectively as the argument’s support).

2.2 The Two Roles of Attacks

Attacks play two roles. Firstly, that $X$ attacks $Y$, is an abstract, declarative representation of the mutual incompatibility of the information contained in the attacking arguments. Secondly, the attack abstractly characterises the dialectical use of $X$ as a counter-argument to $Y$. The former role suggests a necessary condition for specifying an attack between $X$ and $Y$, namely, that they contain mutually incompatible information. However the second role suggests that this condition is not sufficient; attacks should also be defined in such a way as to reflect their use in debate and discussion. Intuitively, if $Y$ is proposed as an argument, then in seeking a counter-argument to $Y$, one seeks to construct an argument $X$ whose conclusion is in conflict with some supporting element of $Y$. This precludes defining an attack from $X$ to $Y$, where despite the arguments containing incompatible information, $X$’s conclusion is not incompatible with some supporting element of $Y$. For example, suppose $Y$ concluding Tweety flies, supported by the premise Tweety is a bird and the defeasible rule that birds fly. Suppose also $X$ concluding Tweety does not fly, supported by the premise defeasible rule Tweety is a penguin and penguins don’t fly. $X$ and $Y$ attack each other, but if $X$ is extended with the defeasible rule that non-flying animals do not have wings, to obtain $X'$ claiming Tweety does not have wings, then $X'$ should not attack $Y$. Intuitively, $X'$ would not be moved as a counter-argument to $Y$; rather it is the sub-argument $X$ of $X'$ that would be moved. Notice also that making any continuation of $X$ also attack $Y$, dramatically increases the number of attacks defined by a theory (and thus the computational expense incurred in evaluating the justified arguments).

A final requirement for attacks is that they should only be targeted at fallible elements of the attacked argument. In particular, conclusions of deductive inferences in an argument cannot be attacked. This should be obvious since the very meaning of deductive (that is, standard logical) inference is that the truth of the premises of a deductive inference guarantees the truth of its conclusion. Any disagreement with the conclusion of a strict inference should therefore be expressed as an attack on fallible elements of the attacked argument. Furthermore, [18] show that allowing attacks on deductive inference steps leads to violation of rationality postulates.

2.3 Distinguishing Preference Dependent and Independent Attacks

We now motivate the distinction between preference dependent and preference independent attacks. Consider the above symmetrically attacking arguments $X$ and $Y$ respectively concluding Tweety does not fly and Tweety flies. Based on the specificity principle’s prioritisation of properties of sub-classes over super-classes, one preferentially concludes Tweety does not fly. The use of the specificity principle can be modelled at the meta-level (i.e., meta to the object-level logic in which arguments $X$ and $Y$ are constructed), as a preference for $X$ over $Y$, so that $X$ asymmetrically defeats $Y$.

Assuming sufficient expressive power, one could in principle encode this metalevel arbitration of the conflict in the object level logic. A standard way of doing this is with undercuts [35]. The inferential step licensed by the rule $b/f = birds fly$, is blocked by a
rule that states that if the bird is a penguin, then the rule holds. This suggests the use of undercuts for yielding the same results as those obtained through the use of preferences, in a way that makes the rationale for preference application more explicit. Undercuts also yield effects that cannot be exclusively effected through preferences. Consider Pollock’s classic example in which there is a red light shining undercuts the rule that if an object looks red then it is red, so blocking the inference from there is an object that looks red, to the conclusion the object is red. Here, the undercut effectively expresses a preference for not drawing the inference over drawing the inference; something that cannot be expressed as a preference ordering over arguments.

We conclude that when specifying an attack by Z on Y, based on Z’s conclusion undercutting a rule in Y, the attacking argument is first and foremost expressing reasons for preferring not to infer Y’s conclusion, rather than to infer Y’s conclusion. Such attacks should therefore be ‘preference independent’, since qualifying the success of such an attack (as a defeat) as being contingent on Y not being preferred to Z, would be to contradict the preference that is effectively expressed by the attack itself. Thus, we argue for a distinction between preference dependent and preference independent attacks, where undercuts fall into the latter category. Note that this does not preclude that a third argument Z’ attacks Z’s conclusion that Y’s conclusion should not be inferred, where Z’’s attack is preference dependent.

We also advocate the preference independence of attacks on assumptions (e.g., as described in [15]). In general, ‘ordinary’ premises used in the construction of arguments may be facts considered true given some evidential basis (e.g., based on perceptions). ‘Assumption’ premises may be assumed without evidential basis, and so their very use in the construction of arguments is contingent on the absence of evidence to the contrary. Hence if Y makes use of an assumption a, and an argument Z concludes that a does not hold, then Z preference independent attacks Y, since the very use of the assumption in Y is contingent on the absence of any such Z. For example, if the antecedent of birds fly is augmented with an assumption that the bird to which the rule is applied, is assumed normal with respect to flying, then the validity of Y concluding Tweety flies, is contingent on the absence of an argument Z claiming that Tweety is not normal with respect to flying given that Tweety is a penguin and the rule penguins are abnormal with respect to flying. Hence Z preference independent attacks Y.

### 2.4 The Distinct Uses of Attacks and Defeats

To recap, attacks encode the mutual incompatibility of the information contained in the attacking and attacked arguments, in a way that accounts for their dialectical use. In turn, the dialectical use of attacks as defeats may or may not be contingent on the preferences defined over the arguments.

As described in Section 2.1, existing works that account for preferences and/or values [5, 11, 32], including [39]’s ASPIC+ framework, define conflict-free and acceptable sets of arguments with respect to the defeats. However, we argue that defining conflict free sets in terms of defeats is conceptually wrong. Since attacks indicate the mutual incompatibility of the information contained in the attacking and attacked arguments, then intuitively one should continue to define conflict-free sets in terms of those that do not contain attacking arguments. Defeats only encode the preference dependent use of attacks in the dialectical evaluation of the acceptability of arguments. They have no bearing on whether one argument can be said to be logically incompatible with another, but rather whether the attack can be validly employed in a dialectical setting.

In the following section, we therefore re-define [39]’s ASPIC+ notion of a conflict
free set, as one in which no two arguments attack rather than defeat. We then examine the implications of this in Section 4.2.

3 The ASPIC+ Framework

In this section we review [39]'s ASPIC+ framework in light of the criteria and requirements enumerated in Sections 2.2 and 2.3. Specifically, we present the ASPIC+ notions of arguments and preference dependent and independent attacks. We also modify [39]'s ASPIC+ framework in two ways: 1) we change the definition of conflict free, as proposed above; 2) we further generalise ASPIC+ so as to capture deductive approaches to argumentation [1, 2, 5, 14].

3.1 ASPIC+ Arguments

ASPIC+ further develops [3]'s instantiation of [23]'s abstract frameworks with accounts of the structure of arguments, the nature of attack and the use of preferences. The framework assumes an unspecified logical language \( L \), and defines arguments as inference trees formed by applying strict or defeasible inference rules to premises that are well formed formulae (wff) in \( L \). A strict rule means that if one accepts the antecedents, then one must accept the consequent no matter what. A defeasible rule means that if one accepts all antecedents, then one must accept the consequent if there is insufficient reason to reject it. There are two ways to use these rules: they could encode domain-specific information (as in e.g. default logic) but could also express general laws of reasoning. For example, the defeasible rules could express argument schemes and the strict rules could consist of all classically valid inferences or more generally conform to any Tarskian consequence notion (cf. [1]). Notice that inclusion of defeasible rules requires some explanation, given that much current work formalises construction of arguments as deductive [1, 2], and in particular classical [14, 26] inference. We justify the need for inclusion of defeasible inference rules in Section 6.

In order to define attacks in the context of a general language \( L \), one needs an appropriately general notion of conflict (i.e., one that does not commit to specific forms of negation). Thus, some minimal assumptions on \( L \) are made; namely that certain wff are a contrary or contradictory of certain other wff. Apart from this, the framework is still abstract: it applies to any set of strict and defeasible inference rules, and to any logical language with a defined contrary relation.

The basic notion of ASPIC+ is that of an argumentation system. Arguments are then constructed with respect to a knowledge base:

Definition 2 [ASPIC+ argumentation system] An argumentation system is a tuple \( AS = (L, -, R, \leq) \) where:

- \( L \) is a logical language.
- \( - \) is a contrariness function from \( L \) to \( 2^L \), such that:
  - \( \varphi \) is a contrary of \( \psi \) if \( \varphi \in \overline{\psi}, \psi \notin \overline{\varphi} \);
  - \( \varphi \) is a contradictory of \( \psi \) (denoted by \( \varphi = \neg \psi \)), if \( \varphi \in \overline{\psi}, \psi \in \overline{\varphi} \).
- \( R = R_s \cup R_d \) is a set of strict \( (R_s) \) and defeasible \( (R_d) \) inference rules of the form \( \varphi_1, \ldots, \varphi_n \rightarrow \varphi \) and \( \varphi_1, \ldots, \varphi_n \Rightarrow \varphi \) respectively (where \( \varphi_i, \varphi \) are metavariables ranging over wff in \( L \)), and such that \( R_s \cap R_d = \emptyset \).
- \( \leq \) is a preordering on \( R_d \).
Definition 3 [ASPLIC+ knowledge base] A knowledge base in an argumentation system \((\mathcal{L}, -\cdot, \mathcal{R}, \leq)\) is a pair \((\mathcal{K}, \leq')\) where:

- \(\mathcal{K} \subseteq \mathcal{L}\), and \(\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a\) where these subsets of \(\mathcal{K}\) are disjoint, and: \(\mathcal{K}_n\) is the (necessary) axioms; \(\mathcal{K}_p\) is the ordinary premises; \(\mathcal{K}_a\) is the assumptions.
- \(\leq'\) is a preordering on \(\mathcal{K} \setminus \mathcal{K}_n\)

Intuitively, the axiom premises denote strict knowledge and thus cannot be attacked, whereas the ordinary and assumption premises (the ‘non-axiom’ premises) are defeasible, and are distinguished by the fact that attacks on the former are preference dependent, whereas attacks on the latter are preference independent (as discussed in Section 2.3). Notice also the orderings on defeasible rules and non-axiom premises (we assume their strict counterparts defined in the usual way, i.e., \(l < l'\) iff \(l \leq l'\) and \(l' \not< l\)). These orderings can be used in defining an ordering on the constructed arguments (as will be shown in Section 5.1).

Example 1 Let \((\mathcal{L}, -\cdot, \mathcal{R}, \leq)\) be an argumentation system where:

- \(\mathcal{L}\) is a language of propositional literals, composed from a set of propositional atoms \(\{a, b, c, \ldots\}\) and the symbols \(\sim\) and \(\sim\) respectively denoting strong and weak negation (i.e., negation as failure). \(\alpha\) is a strong literal if \(\alpha\) is a propositional atom or of the form \(\sim\beta\) where \(\beta\) is a propositional atom. \(\alpha\) is a wff of \(\mathcal{L}\), if \(\alpha\) is a strong literal or of the form \(\sim\beta\) where \(\beta\) is a strong literal.
- For any wff \(\alpha\), \(\alpha\) and \(\sim\alpha\) are contradictories and \(\alpha\) is a contrary of \(\sim\alpha\).
- \(\mathcal{R}_s = \{t, q \Rightarrow \sim p\}, \mathcal{R}_d = \{\sim s \Rightarrow t; r \Rightarrow q; a \Rightarrow p\}\)
- \(\leq = \{r \Rightarrow q < a \Rightarrow p\}\) (for simplicity we will leave \(l \leq l'\) and \(l' \not< l\) relations implicit when specifying these orderings)

\((\mathcal{K}, \leq')\) is the knowledge base such that \(\mathcal{K}_n = \{a\}, \mathcal{K}_p = \{r, \sim r\}, \mathcal{K}_a = \{\sim s\}\), and \(\leq = \{\sim r \not< r\}\).

Arguments are now defined, where for any argument \(A\), \(\text{Prem}\) returns all the formulas of \(\mathcal{K}\) (premises) used to build \(A\). \(\text{Conc}\) returns \(A\)’s conclusion, \(\text{Sub}\) returns all of \(A\)’s sub-arguments, \(\text{DefRules}\) and \(\text{StRules}\) respectively return all defeasible and all strict rules in \(A\), and \(\text{TopRule}(A)\) returns the last rule applied in \(A\).

Definition 4 [ASPLIC+ arguments] An argument \(A\) on the basis of a knowledge base \((\mathcal{K}, \leq')\) in an argumentation system \((\mathcal{L}, -\cdot, \mathcal{R}, \leq)\) is:

1. \(\varphi\) if \(\varphi \in \mathcal{K}\) with: \(\text{Prem}(A) = \{\varphi\}; \text{Conc}(A) = \varphi; \text{Sub}(A) = \{\varphi\}; \text{Rules}(A) = \emptyset; \text{TopRule}(A) = \text{undefined}.
2. \(A_1, \ldots, A_n \rightarrow l' \Rightarrow \psi\) if \(A_1, \ldots, A_n\) are finite arguments such that there exists a strict/defeasible rule \(\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow l' \Rightarrow \psi\) in \(\mathcal{R}_s/\mathcal{R}_d\).  
   \(\text{Prem}(A) = \text{Prem}(A_1) \cup \ldots \cup \text{Prem}(A_n),\)
   \(\text{Conc}(A) = \psi,\)
   \(\text{Sub}(A) = \text{Sub}(A_1) \cup \ldots \cup \text{Sub}(A_n) \cup \{A\}.\)
   Note that \(A_1 \ldots A_n\) are referred to as the proper sub-arguments of \(A\)
   \(\text{Rules}(A) = \text{Rules}(A_1) \cup \ldots \cup \text{Rules}(A_n) \cup \{\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow l' \Rightarrow \psi\}\)
Definition 5 [c-consistent] A set $S \subseteq \Lambda$ is c-consistent if for no $\varphi$ it holds that $S \vdash \varphi, \neg \varphi$. Otherwise $S$ is c-inconsistent. We say that $S \subseteq \Lambda$ is minimally c-inconsistent iff $S$ is c-inconsistent and $\forall S' \subseteq S$, $S'$ is c-consistent.

Note that we use the term ‘c-consistent’ to distinguish the notion of consistency in Definition 2. Also note that if $S \vdash \varphi, \phi$, where $\phi \in \overline{\varphi}$, then $S$ can still be c-consistent. As we will see later, such situations do not arise when capturing deductive approaches in ASPIC+ as in these approaches there are no contraries, only contradictories.

Definition 6 [c-consistent argument] An argument $A$ on the basis of a knowledge base $(\mathcal{K}, \preceq')$ in an argumentation system $(\mathcal{L}, \neg, \mathcal{R}, \preceq)$, is c-consistent iff $\text{Prem}(A)$ is c-consistent.

3.2 Attacks and Defeats

We now review [39]'s definition of attacks amongst arguments. An argument $A$ attacks an argument $A'$ if the conclusion of $A$ (i.e., $\text{Conc}(A)$) is a contrary or contradictory of: a non-axiom premise in $A'$; the consequent of a defeasible rule in $A'$, or; a defeasible inference step in $A'$. These three kinds of attack are respectively called undermining, rebutting and undercutting attacks. To model the latter, it is assumed that applications of inference rules can be expressed in the object language; in our examples below we simply assume that rules are named with wff’s $r$ and that undercutters conclude $\varphi$.
Definition 7 [ASPIC\(^+\) attacks] \(A\) attacks \(B\) iff \(A\) undercuts, rebuts or undermines \(B\), where:

- \(A\) undercuts argument \(B\) (on \(B'\)) iff \(\text{conc}(A) \in \mathcal{T}\) for some \(B' \in \text{Sub}(B)\) such that \(B'\)'s top rule is defeasible and is named by \(r\) in \(\mathcal{L}\).\(^1\)
- \(A\) rebuts argument \(B\) (on \(B'\)) iff \(\text{conc}(A) \in \mathcal{T}\) for some \(B' \in \text{Sub}(B)\) of the form \(B'_1, \ldots, B'_n \models \varphi\). In such a case \(A\) contrary-rebuts \(B\) iff \(\text{conc}(A)\) is a contrary of \(\varphi\).
- Argument \(A\) undermines \(B\) (on \(B'\)) iff \(\text{conc}(A) \in \mathcal{T}\) for some \(B' = \varphi, \varphi \in \text{Prem}_{a/p}(B)\). In such a case \(A\) contrary-undersmnes \(B\) iff \(\text{conc}(A)\) is a contrary of \(\varphi\) or if \(\varphi \in \mathcal{K}_a\).

Notice that an attack on an assumption is categorised as a special case of a contrary undermining attack.

Example 3 [Example 1 continued] The arguments defined on the basis of the knowledge base and argumentation system in Example 1 are: \(A' = [a], A = [A' \Rightarrow p], B_1 = [\sim s], B_1' = [B_1 \Rightarrow t], B_2 = [r], B_2' = [B_2 \Rightarrow q], B = [B_1', B_2' \Rightarrow \sim p], C = [\sim r]\).

Then, \(B\) rebuts \(A\) on \(A\), \(C\) undermines \(B\) and \(B_2'\) on \(B_2\), and \(C\) and \(B_2\) undermine each other.

Notice that if in addition one had the argument \(D = [\sim d; \sim d \Rightarrow s]\), then \(D\) would contrary-underline \(B\) on \(B_1\) and \(B_1'\).

The above arguments and attacks are depicted in Figure 1a) in which we first show the arguments with their internal structure, and then the abstract attack graph.

Note that Definition 7 complies with Section 2.2's rationale for defining attacks. An attack originating from an argument \(A\) requires that its conclusion \(\text{conc}(A)\) (and not the conclusion of any sub-argument of \(A\)) be in conflict with some fallible element – i.e., some non-axiom premise, or defeasible rule or conclusion of a defeasible rule – in the attacked argument. Thus, while \(B2\) rebut-attacks \(C\) in Example 3, the argument \(B\), that contains \(B2\) as a sub-argument, does not attack \(C\). Also, although \(A\) and \(B\) have contradictory conclusions, only \(B\) rebuts \(A\); \(A\) does not rebut \(B\) as \(B\)'s conclusion is the consequent of a strict rule. [18] refer to this as a restricted rebut, and show for a special case of ASPIC\(^+\) that if the restriction is lifted so as to allow \(A\) to rebut \(B\), then this would lead to violation of [18]'s rationality postulates.

We assume any partial preordering \(\preceq\) on the constructed arguments. In Section 5.1 we will utilise the orderings on non-axiom premises and defeasible rules, to give example definitions of \(\preceq\). As is usual, its strict counterpart \(\prec\) is defined as \(X \prec Y\) iff \(X \preceq Y\) and \(Y \not\preceq X\). The notion of a successful attack (defeat) can then be defined:

Definition 8 [ASPIC\(^+\) defeats] \(A\) defeats \(B\) iff \(A\) undercuts \(B\), or; \(A\) rebuts/undermines \(B\) on \(B'\), and either \(A\) contrary rebuts/undermines \(B\) or \(A \neq B'\). \(A\) strictly defeats \(B\) iff \(A\) defeats \(B\) and \(B\) does not defeat \(A\).

Notation 4 Henceforth, \(\Rightarrow\) may denote the attack relation, and \(\Rightarrow\) the defeat relation.

The definition of defeats complies with Section 2.3’s rationale for distinguishing between preference dependent and preference independent attacks. Undercuts always succeed as defeats, and so are preference independent. The preference independence of attacks on assumptions is ensured by declaring attacks on assumptions to be a special case of contrary-undermining attacks. As discussed in Section 2.3, both undercutting

\(^1\)This definition assumes that defeasible inference rules are named in \(\mathcal{L}\); the precise nature of this naming convention will be left implicit.
and assumption attacks encode an asymmetry: the use of the attacked rule/assumption $\psi$ in an argument $B$ is contingent on the absence of an acceptable attacking argument $A$ with an undercutting, respectively undermining, conclusion $\varphi$. However, the reverse does not hold, in that the use of $\varphi$ in an argument is not contingent on the absence of an argument concluding $\psi$. The notion of a contrary relation generalises the above cases of asymmetric preference independent attacks. The notion provides for greater flexibility in declaring other such $\psi$s and $\varphi$s, where attacks from $\varphi$ to $\psi$ are not undercuts or undermining attacks on assumptions, but are still preference independent. Notice that in such cases it would be counter-intuitive to allow $\psi$ to be an axiom premise or the conclusion of a strict rule. Prohibiting such cases is formalised as an assumed property of any $\text{ASPIC}^+$ framework in the following section.

Example 5 [Example 3 continued] Given the orderings on defeasible rules and premises in Example 1, let $B_2' \prec A$, $B \prec A$ (given $r \Rightarrow q < a \Rightarrow p$), $C \prec B_2$, $C \prec B_2'$, $C \prec B$ (given $\neg r < r'$). Then $B$ does not defeat $A (B \rightarrow A)$, $C \rightarrow B$, $C \rightarrow B_2'$ and $B_2 \rightarrow C$ (the arguments and defeats are depicted in Figure 1b)).

Notice that if one had the additional argument $D$ described in Example 3, then $D$ would defeat $B_1$ and so $B_1'$ and $B$. Note that the preference independence of $D \rightarrow B_1$ effectively implements the negation as failure semantics of $\sim$. The validity of $B_1, B_1'$ and $B$ is contingent on $s$ not being provable (i.e., there being no acceptable argument claiming $s$). Also note that the preference independence of $D \rightarrow B_1'$ is based both on $s$ being a contrary of $\sim s$, and $\sim s$ being an assumption premise, given that one is
assuming that \( s \) is not provable. Of course, in general, attacks by contraries can also be on ordinary premises or the consequents of defeasible rules.

### 3.3 Structuring Argumentation Frameworks

We now define two notions of a structured argumentation framework instantiated by an argumentation theory. The first is defined as in [39]. The second accounts for this paper’s definition of c-consistent arguments.

**Definition 9 [Argumentation theory]** An argumentation theory is a tuple \( AT = (AS, KB) \) where \( AS \) is an argumentation system and \( KB \) is a knowledge base in \( AS \).

**Definition 10 [(c-)Structured Argumentation Frameworks]** Let \( AT \) be an argumentation theory \( (AS, KB) \).

- A structured argumentation framework \( (SAF) \) defined by \( AT \), is a triple \( \langle A, C, \preceq \rangle \) where \( A \) is the set of all arguments constructed from \( KB \) in \( AS \) (henceforth called the set of arguments on the basis of \( AT \)), \( \preceq \) is a pre-ordering on \( A \), and \((X, Y) \in C\) iff \( X \) attacks \( Y \).

- A c-structured argumentation framework \( (c-SAF) \) defined by \( AT \), is a triple \( \langle A, C, \preceq \rangle \) where \( A \) is the set of all c-consistent arguments constructed from \( KB \) in \( AS \), \( \preceq \) is a pre-ordering on \( A \), and \((X, Y) \in C\) iff \( X \) attacks \( Y \).

Henceforth, we may write ‘\((c-)SAF\)’ instead of writing ‘\( SAF \) or \( c-SAF \)’.

In [39], it is assumed that any argumentation theory satisfies a number of properties. We repeat these here, and add an additional ‘c-classicality’ property for \( c-SAF \)s. Prior to this, recall that for any \( P \subseteq L \), the closure of \( P \) under strict rules, denoted \( \text{Cl}_{\mathcal{R}_s}(P) \), is the smallest set containing \( P \) and the consequent of any strict rule in \( \mathcal{R}_s \) whose antecedents are in \( \text{Cl}_{\mathcal{R}_s}(P) \).

**Definition 11 [Well defined (c-)SAFs]** Let \( AT = (AS, KB) \) be an argumentation theory, where \( KB = (K_n, \preceq') \) and \( AS = (\mathcal{L}, \neg, \mathcal{R}, \preceq) \). We say that \( AT \) is:

- **closed under contraposition** iff for all \( S \subseteq \mathcal{L} \), \( s \in S \) and \( \phi \), if \( S \vdash \phi \), then \( S \setminus \{s\} \cup \{-\phi\} \vdash -s \).

- **closed under transposition**\(^2\) iff if \( \phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{R}_s \), then for \( i = 1 \ldots n \), \( \phi_1, \phi_{i+1}, \ldots, \phi_n \rightarrow -\phi_i \in \mathcal{R}_s \).

- **axiom consistent** iff \( \text{Cl}_{\mathcal{R}_s}(K_n) \) is consistent.

- **c-classical** iff for any minimal c-inconsistent \( S \subseteq \mathcal{L} \) and for any \( \varphi \in S \), it holds that \( S \setminus \{\varphi\} \vdash -\varphi \) (i.e., amongst all arguments defined there exists a strict argument with conclusion \( -\varphi \) with all premises taken from \( S \setminus \{\varphi\} \)).

- **well formed** if whenever \( \varphi \) is a contrary of \( \psi \) then \( \psi \notin K_n \) and \( \psi \) is not the consequent of a strict rule.\(^3\)

If a \( c-SAF \) is defined by an \( AT \) that is c-classical, axiom consistent, well formed and closed under contraposition or closed under transposition, then the \( c-SAF \) is said to be well defined.

If a \( SAF \) is defined by an \( AT \) that is axiom consistent, well formed and closed under contraposition or closed under transposition, then the \( SAF \) is said to be well defined.

---

\(^2\)The notion of closure under contraposition is taken from [18].

\(^3\)This formulation repairs an error in the one of [39].
Henceforth, we will assume that any (c-)SAF is well defined. The intuitions underlying the first four properties are self-evident. The rationale for the well-formed assumption is discussed in Section 3.2.

(c-)SAFs can now be linked to Dung frameworks. Firstly, note that as with existing approaches [5, 11, 32], [39]’s notion of a conflict free set of arguments is defined with respect to the derived defeat relation.

**Definition 12** [Defeat conflict free for (c-)SAFs] Let \( \Delta = (A, C, \preceq) \) be a (c-)SAF, and \( D \subseteq A \times A \), where \((X, Y) \in D \) iff \( X \) defeats \( Y \). Then \( S \subseteq A \) is defeat conflict free iff \( \forall X, Y \in S, (X, Y) \notin D \).

However, we have in Section 2.4 argued that conflict free sets should be defined with respect to the attack relation, and defeats reserved for the dialectical use of attacks:

**Definition 13** [Attack conflict free for (c-)SAFs] Let \( \Delta = (A, C, \preceq) \) be a (c-)SAF. Then \( S \subseteq A \) is attack conflict free iff \( \forall X, Y \in S, (X, Y) \notin C \).

In either case, the justified arguments are then evaluated on the basis of the extensions of a Dung framework instantiated by the arguments and derived defeat relation:

**Definition 14** [Extensions and justified arguments/conclusions of (c-)SAFs] Let \( \Delta = (A, C, \preceq) \) be a (c-)SAF, and \( D \subseteq A \times A \), where \((X, Y) \in D \) iff \( X \) defeats \( Y \). Let \( S \subseteq A \) be defeat or attack conflict free. The extensions and justified arguments of \( \Delta \) are the extensions of the Dung framework \((A, D)\), as defined in Definition 1.

For \( T \in \{\text{admissible, complete, preferred, grounded, stable}\} \), we say that:

- \( \varphi \) is a \( T \) credulously justified conclusion of \( \Delta \) iff there exists an argument \( A \) such that \( \text{Conc}(A) = \varphi \), and \( A \) is credulously justified under the \( T \) semantics.

- \( \varphi \) is a \( T \) sceptically justified conclusion of \( \Delta \) iff for every \( T \) extension \( E \), there exists an argument \( A \in E \) such that \( \text{Conc}(A) = \varphi \).

\( S \) is a \( d-T \) extension if \( S \) is defined as defeat conflict free, and an \( a-T \) extension if \( S \) is defined as attack conflict free.

We now recall some notation from [39], and then in this paper we define the notion of an argument \( A \) being a strict extension of a set of arguments \( \{A_1, \ldots, A_n\} \).

**Notation 6** Let \( M(B) \) denotes the maximal fallible sub-arguments of \( B \), where for any \( B' \in \text{Sub}(B), B'' \in M(B) \) iff:

1. \( B' \) final inference is defeasible or \( B' \) is a non-axiom premise, and;
2. there is no \( B'' \in \text{Sub}(B) \) s.t. \( B'' \neq B \) and \( B' \in \text{Sub}(B'') \), and \( B'' \) satisfies 1).

**Definition 15** [Strict Extensions of Arguments] For any set of arguments \( \{A_1, \ldots, A_n\} \), the argument \( A \) is a strict extension of \( \{A_1, \ldots, A_n\} \) iff:

- \( \text{Prem}_{a/p}(A) \supseteq \bigcup_{i=1}^n \text{Prem}_{a/p}(A_i) \) (i.e., the ordinary and assumption premises in \( A \) are exactly those in \( \{A_1, \ldots, A_n\} \));
- \( \text{DefRules}(A) \supseteq \bigcup_{i=1}^n \text{DefRules}(A_i) \) (i.e., the defeasible rules in \( A \) are exactly those in \( \{A_1, \ldots, A_n\} \));
- \( \text{StRules}(A) \supseteq \bigcup_{i=1}^n \text{StRules}(A_i) \) and \( \text{Prem}_a(A) \supseteq \bigcup_{i=1}^n \text{Prem}_a(A_i) \) (i.e., the strict rules and axiom premises of \( A \) are a superset of the strict rules and axiom premises in \( \{A_1, \ldots, A_n\} \)).
Example 7 [Example 5 continued] Referring to the arguments in Example 3, we have
\( M(A) = \{ A \} \), \( M(B) = \{ B'_1, B'_2 \} \), and \( B \) is a strict extension of \( B'_1 \) and \( B'_2 \). Now notice that the argumentation theory in Example 1 is not well defined, since it is neither closed under contraposition or transposition. Closure under transposition augments \( R_s \) with rules \( t, p \to \neg q \) and \( p, q \to \neg t \), so obtaining the additional arguments \( A^+_1 = [B'_1, A \Rightarrow \neg q] \) that rebut-attacks \( B'_2 \) (and so \( B \)), and \( A^+_2 = [A, B'_2 \Rightarrow \neg t] \) that rebut-attacks \( B'_1 \) (and so \( B \)). Figure 1c) shows these additional arguments and attacks.

4 Properties and Postulates

In this section we examine the implications of the attack definition of conflict free sets. We will show that under some intuitive assumptions on the preference ordering over arguments both SAFs and c-SAFs satisfy the key properties of Dung frameworks. We also show that [18]'s rationality postulates can straightforwardly be shown to hold. On the other hand, we will show that under [39]'s ‘defeat definition’ of conflict free, key properties of Dung frameworks straightforwardly hold, whereas satisfaction of [18]'s rationality postulates requires assumptions on preference orderings. Note that for the defeat definition, [39] has already shown satisfaction of the rationality postulates for SAFs. This paper extends [39]'s results to c-SAFs. Finally, we will show equivalence of admissible and complete extensions under the attack and defeat definitions of conflict free.

4.1 Properties of SAFs and c-SAFs under the Attack Definition of Conflict Free

Retaining attacks when defining conflict-free sets potentially undermines some key results shown for Dung frameworks. To illustrate, consider Example 5’s SAF, with the arguments and attacks shown in Figure 1a). Since \( B \prec A \), then \( A \not\rightarrow B \) (as shown in Figure 1b)). So \( B \) is acceptable w.r.t. \( \{ B \} \), and \( \{ B \} \) is a-admissible. Now \( A \) is acceptable w.r.t. \( \{ B \} \), but \( \{ A, B \} \) is not attack conflict free and so not a-admissible. This violates Dung’s fundamental lemma [23], which states that if \( S \) is admissible and \( A \) is acceptable w.r.t. \( S \) then \( S \cup \{ A \} \) is admissible. However, if the SAF is well defined (Definition 11), then under the assumption that an argument ordering is reasonable, we can show that the fundamental lemma holds.

A preordering \( \preceq \) is reasonable if it satisfies properties that one would intuitively expect of any ordering over arguments composed from from fallible and infallible elements. Firstly, whenever an argument \( A \) contains no fallible elements (i.e., is strict and firm), then it is strictly stronger than all arguments with fallible elements, and not weaker than any other argument. Also, extending an argument with only axiom premises and strict inferences does not change its strength. The intuitions here are self-evident. The second property is essentially a strengthening of the requirement that the strict counter-part \( \prec \) of \( \preceq \) is asymmetric, by stating that for any set \( A' = \{ C_1, \ldots, C_n \} \) of arguments, it cannot be that for all \( i, C' \prec C_i \) where \( C' \) is a strict extension of \( A' \setminus C_i \).

Definition 16 [Reasonable Argument Orderings] \( \preceq \) is said to be reasonable iff:

1. i) \( \forall A, B, \) if \( A \) is strict and firm and \( B \) is plausible or defeasible, then \( B \prec A \); ii) \( \forall A, B, \) if \( B \) is strict and firm then \( B \not\prec A \); iii) \( \forall A, A', B, C \) such that \( A \prec B, C \prec A, \) and \( A' \) is a strict extension of \( \{ A \}, \)
then \( A' \prec B, C \prec A' \) (i.e., applying strict rules to a single argument’s conclusion and possibly adding new axiom premises does not change the strength of arguments).

2. Let \( \{C_1, \ldots, C_n\} \) be a finite subset of \( \mathcal{A} \), and for \( i = 1 \ldots n \), let \( C^{+/i} \) be some strict extension of \( \{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\} \). Then it is not the case that: \( \forall i, C^{+/i} \prec C_i \).

In Section 5.1 we give example definitions of argument orderings; specifically, the preorderings on the defeasible premises and rules are used to define argument orderings under the commonly used weakest and last link principles. We will then show that these argument orderings are reasonable. Indeed, we will show that for these argument orderings, Definition 16-2 is equivalent to requiring that the strict counterpart \( \prec \) of \( \preceq \) is a strict partial ordering (transitive, irreflexive, and so asymmetric). Henceforth, we will assume that the preorder \( \preceq \) of any (c-)SAF is reasonable.

We now examine the implications of an argument ordering being reasonable. Under the assumption that Example 5’s SAF is well-defined, we additionally have the arguments and attacks described in Example 7, and shown in Figure 1c). Recall that we have the maximal fallible sub-arguments \( \{A, B'_1, B'_2\} \) of \( A \) and \( B \), where:

- \( B \) is a strict extension of \( \{B'_1, B'_2\} \);
- \( A_1^+ \) a strict extension of \( B'_1, A \);
- \( A_2^+ \) a strict extension of \( B'_2, A \).

Assuming \( \preceq \) is reasonable, then by Definition 16-2 it cannot be that \( B \prec A, A_1^+ \prec B'_2 \) and \( A_2^+ \prec B'_1 \). Since by assumption \( B \prec A \) then it must be that either \( A_1^+ \not\prec B'_2 \) or \( A_2^+ \not\prec B'_1 \), and so \( A_1^+ \) defeats \( B'_2 \) or \( A_2^+ \) defeats \( B'_1 \). Intuitively, since \( B \prec A \) is based on \( A \) being strictly preferred to some maximal fallible sub-argument in \( B \) (i.e., \( B'_1 \) or \( B'_2 \)), then if both \( A_1^+ \prec B'_2 \) (\( B'_2 \) strictly preferred to \( B'_1 \) or \( A \)) and \( A_2^+ \prec B'_1 \) (\( B'_1 \) strictly preferred to \( B'_2 \) or \( A \)), then this would imply a cycle in the preference ordering over \( \{A, B'_1, B'_2\} \). Hence, \( A_1^+ \not\prec B'_2 \) given that \( B'_2 \prec A \) (given \( r \Rightarrow q < a \Rightarrow p \)). In fact, the following general result can be shown:

**Proposition 8** Let \( A \) and \( B \) be arguments where \( B \) is plausible or defeasible and \( A \) and \( B \) have contradictory conclusions, and if \( A \) and \( B \) are defined as in Definition 6, then \( \text{Prem}(A) \cup \text{Prem}(B) \) is c-consistent. Then:

1. For all \( B' \in M(B) \), there exists a strict extension \( A^+_{B'} \) of \( (M(B) \setminus \{B'\}) \cup M(A) \) such that \( A^+_{B'} \) rebuts or undermines \( B \) on \( B' \).
2. If \( B \prec A \), and \( \preceq \) is reasonable, then for some \( B' \in M(B) \), \( A^+_{B'} \) defeats \( B \).

Let us now generalise the earlier suggested counter-example to the fundamental lemma. Assume \( B \in S, S \) is admissible, and either: 1) \( B \) attacks \( A \) on \( A' \), \( B \prec A' \), and so \( B \) does not defeat \( A \), or; 2) \( A \) attacks \( B \) on \( B' \), \( A \prec B' \), and so \( A \) does not defeat \( B' \). The proof of Proposition 10 below then makes use of Proposition 8 to show that in neither case can \( A \) be acceptable w.r.t. \( S \). This means that the result that if \( A \) is acceptable w.r.t. an admissible \( S \) then \( S \cup \{A\} \) is conflict free, and hence Dung’s fundamental lemma, is not under threat. Prior to Proposition 10, we state a key result for c-SAFs in order to show that Dung’s fundamental lemma and the rationality postulates can be shown when arguments are restricted to those with consistent premises:
Proposition 9 Let $(\mathcal{A}, C, \preceq)$ be a c-SAF. If $A_1, \ldots, A_n$ are acceptable w.r.t. some conflict-free $E \subseteq \mathcal{A}$, then $\bigcup_{i=1}^{n} \text{Prem}(A_i)$ is c-consistent.

Proposition 10 Let $A$ be acceptable w.r.t an admissible extension $S$ of a (c-)SAF $(\mathcal{A}, C, \preceq)$. Then $S' = S \cup \{A\}$ is conflict free.

Proposition 11 Let $A, A'$ be acceptable w.r.t an admissible extension $S$ of a (c-)SAF $(\mathcal{A}, C, \preceq)$. Then:
1. $S' = S \cup \{A\}$ is admissible
2. $A'$ is acceptable w.r.t. $S'$.

We have shown that given reasonable argument orderings, a well defined (c-)SAF satisfies Dung’s fundamental lemma. This implies that the admissible extensions of a (c-)SAF form a complete partial order w.r.t. set inclusion, and that every admissible extension is contained in a preferred extension. Also, given the definitions of defeat and acceptability, it is easy to see (in the same way as for Dung frameworks) that if $A$ is acceptable w.r.t. $S$, then $A$ is acceptable w.r.t. any superset of $S$ (this result is stated as Lemma 33-1 in the Appendix). Thus, a (c-)SAF’s characteristic function is monotonic, implying that the grounded extension can be identified by the function’s least fixed point. It is also easy to see that every stable extension is a preferred extension.

4.2 Rationality Postulates for SAFs and c-SAFs under the Attack Definition of Conflict Free

As discussed in Section 1, the intermediate level of abstraction (between concrete instantiating logics and Dung’s fully abstract theory) adopted by ASPIC [3] and ASPIC+ [39] frameworks, allows for the formulation and evaluation of postulates [18] whose satisfaction ensure that any concrete instantiations of the frameworks fulfil some rational criteria. We now show that under the attack definition of conflict free, SAFs and c-SAFs satisfy [18]’s rationality postulates for the complete (and so by implication the grounded, preferred and stable) semantics defined in Definition 1.

Theorem 12 below states that for any argument $A$ in a complete extension $E$, all sub-arguments of $A$ are in $E$. Theorem 13 then states that the conclusions of arguments in a complete extension are closed under strict inference.

Theorem 12 (Sub-argument Closure) Let $\Delta = (\mathcal{A}, C, \preceq)$ be a (c-)SAF and $E$ an a-complete extension of $\Delta$. Then for all $A \in E$: if $A' \in \text{Sub}(A)$ then $A' \in E$.

Theorem 13 (Closure under Strict Rules) Let $\Delta = (\mathcal{A}, C, \preceq)$ be a (c-)SAF and $E$ an a-complete extension of $\Delta$. Then $\{\text{Conc}(A) \mid A \in E\} = \text{Cl}_{R_e}(\{\text{Conc}(A) \mid A \in E\})$.

Theorem 14 below, states that the conclusions of arguments in an admissible extension (and so by implication complete extension) are mutually consistent. Theorem 15 then states the mutual consistency of the strict closure of conclusions of arguments in a complete extension.

Theorem 14 (Direct Consistency) Let $\Delta = (\mathcal{A}, C, \preceq)$ be a (c-)SAF and $E$ an a-admissible extension of $\Delta$. Then $\{\text{Conc}(A) \mid A \in E\}$ is consistent.
**Theorem 15 (Indirect Consistency)** Let $\Delta = (A, C, \succeq)$ be a (c-)SAF and $E$ an a-complete extension of $\Delta$. Then $\text{Cl}_{R_s}(\{\text{Conc}(A) \mid A \in E\})$ is consistent.

Note that the task of showing that [3]'s consistency postulates are satisfied is simplified by the fact that a conflict free set excludes attacking arguments. The proof that 'direct consistency' holds essentially amounts to isolating the case where two arguments $A, B \in E$ have contrary conclusions ($\text{Conc}(A) \in \overline{\text{Conc}(B)}$), but neither attacks the other. But then one can show that these arguments are of the type described in Proposition 8, and so by the proposition’s first result, there is a strict extension $A^+$ of fallible sub-arguments in $A$ and $B$ such that $A^+$ attacks $B$. Since $A$ and $B$ are by assumption acceptable w.r.t. $E$, then $A^+$ must be acceptable w.r.t. $E$. But then this contradicts Proposition 10 which states that $E \cup \{A^+\}$ must be conflict free.

4.3 Comparison of Attack and Defeat Definition of Conflict Free

Under the defeat definition of conflict free, the properties discussed in Section 4.1 are of course shown to hold for (c-)SAFs in the same way as for Dung frameworks. We now state the equivalence of extensions of (c-)SAFs under the attack and defeat definitions of conflict free.

**Proposition 16** Let $\Delta$ be a (c-)SAF. For $T \in \{$admissible, complete, grounded, preferred, stable$\}$, $E$ is an a-$T$ extension of $\Delta$ iff $E$ is a d-$T$-admissible extension of $\Delta$.

Given the previous section’s results, Proposition 16 implies that [18]'s rationality postulates not only hold for SAFs under the defeat definition (as already shown in [39]), but also for c-SAFs under the defeat definition.

**Corollary 17** Let $\Delta$ be a (c-)SAF. Then Theorems 12-15 hold for the d-admissible and d-complete extensions of $\Delta$.

Notice that directly proving satisfaction of the consistency postulates for the defeat definition of conflict free is more involved. One needs to show that an admissible extension contains arguments that do not defeat each other. Proof of this crucially relies on Proposition 8. The trade off is that with the attack definition, proof of the fundamental lemma is more involved since one needs to first show that any argument acceptable w.r.t. an admissible extension is conflict free when included in that extension. It is the proof of this result that crucially depends on Proposition 8. Notice that in both cases, one needs to consider the internal structure of arguments and assume that the preference ordering satisfies some intuitive properties.

Proposition 16’s equivalence begs the question as to why one should advocate the attack definition of conflict free. Firstly, a result that shows that the two different notions of conflict-freeness are (under certain assumptions) equivalent in the extensions they produce is theoretically valuable in itself. Apart from this, we have argued in Section 2 that the attack definition is conceptually more well justified. In Example 5, neither $B$ or $A$ defeat each other, and neither $B$ or $C$ defeat each other. Under the defeat definition, $\{B, A\}$ and $\{B, C\}$ are ‘conflict free’. But in what meaningful sense can these sets be said to be conflict free, when they contain elements that are mutually inconsistent? Consider then [7]'s example that purports to illustrate violation of the consistency postulates by approaches augmenting Dung frameworks with preferences. An expert argues $(A)$ that a given violin is a Stradivarius and therefore expensive. A three-year old child's argument $B$ then states that it is not a Stradivarius. According to
attacks A but A does not attack B, and A is preferred over B since the expert is more reliable than the child. [7] observe that the unique PAF-extension \{A, B\} violates the consistency postulate. In Section 6.2 we demonstrate that the problem does not arise if all arguments that can be constructed are taken into account (the expert can use a sub-argument \(A'\) of A that defeats B so that \{A, B\} is not admissible), illustrating that the problem has more to do with imperfect reasoners. However, we also note that the attack definition of conflict free is more tolerant of imperfect reasoning. Without taking into account all constructible arguments, \{A, B\} is of course not conflict free and so not a PAF-extension, and so consistency is not violated.

Finally, the provision of a more conceptually well justified definition of conflict free has implications for other extensions of Dung’s framework. Consider Extended Argumentation Frameworks (EAFs) [32] in which arguments expressing preferences undermine the success of attacks as defeats, by attacking attacks. [32] illustrates (pg.9) that EAFs’ characteristic functions are not in general monotonic; monotonicity is shown only for EAFs that are effectively hierarchical stratifications of Dung frameworks in which arguments expressing preferences only attack attacks between arguments in lower levels of the hierarchy. Also, [34] links ASPIC+ arguments (extended to express preferences over other arguments) and attacks to EAFs. [34] shows that the rationality postulates only hold for hierarchical EAFs. Both the lack of monotonicity and failure of postulates for non-hierarchical EAFs is attributable to the fact that conflict free and admissible sets under the defeat definition may contain arguments that attack each other. Preliminary results, to be developed in future work, show that monotonicity and rationality postulates are satisfied by non-hierarchical EAFs under the attack definition of conflict free.

To conclude, given our advocacy of the attack definition of conflict free, we henceforth assume any extension of a (c-)SAF to be attack conflict free, and thus will henceforth refer to a ‘T’ extension rather than an ‘a-T’ extension.

## 5 Instantiating Structured Argumentation Frameworks

We have modified ASPIC+ in two ways: we have additionally defined c-SAFs whose arguments must be built on mutually consistent premises, and motivated an alternative attack definition of conflict free sets. We have shown that properties and postulates hold for well defined SAFs and c-SAFs with argument preference preordering that are reasonable in that they satisfy assumptions one would intuitively expect to hold. In this section we study various ways to instantiate the ASPIC+ framework. In Section 5.1 we consider ways of ‘instantiating’ preference preorderings over arguments; that is, specific ways of defining preorderings based on the preorderings over defeasible rules and non-axiom premises. We will show that the defined preorderings are reasonable. In Section 5.2 we extend with preferences Amgoud & Besnard’s approach to structured argumentation [1, 2] based on Tarski’s notion of an abstract logic. We then the extended abstract logic approach in ASPIC+. In Section 5.3 we define classical logic instantiations of c-SAFs, and show an equivalence between one such instantiation and Brewka’s Preferred Subtheories [16].

### 5.1 Weakest and Last Link Preference Relations

In [39], argument preorderings \(\preceq\) are defined, by defining the conditions under which \(B \prec A\) under the well known weakest- and last-link principles. Intuitively, \(B \prec A\) by
the weakest-link principle if the set of defeasible rules in $B$ is $\prec_s$ the set of defeasible rules in $A$, and the set of non-axiom premises in $B$ is $\prec_s$ the set of non-axiom premises in $A$, where $\prec_s$ is a set ordering based on the preorders $\leq$ and $\leq'$ over defeasible rules and non-axiom premises. Intuitively, $B \prec A$ by the last-link principle if the set of last defeasible rules applied in constructing $B$ is $\prec_s$ the set of last defeasible rules in $A$. If these sets are empty the sets of non-axiom premises are compared. [39] uses the Elitist1 definition of $\prec_s$ (see below). In this paper we broaden the range of instantiations of $\leq$, by allowing for alternative definitions of $\prec_s$.

**Definition 17** [Ordering $\prec_s$] Let $\Gamma$ be a finite set $\{l_1, \ldots, l_n\}$, $\Gamma' \subseteq \Gamma$, $\Gamma'' \subseteq \Gamma$, and $\leq$ a pre-ordering on $\Gamma$, where $l < l'$ iff $l \leq l'$ and $l' \nleq l$. Then,

1. Elitist1: $\Gamma' \prec_{Elit1} \Gamma''$ iff $\exists l' \in \Gamma' \text{ s.t. } \forall l'' \in \Gamma'' \implies l'' < l'$.
2. Elitist2: $\Gamma' \prec_{Elit2} \Gamma''$ iff $\exists l' \in \Gamma' \setminus \Gamma'' \text{ s.t. } \forall l'' \in \Gamma'' \implies l < l''$.
3. Democratic1: $\Gamma' \prec_{Dem1} \Gamma''$ iff $\forall l' \in \Gamma'$, $\exists l'' \in \Gamma'' \text{ s.t. } l' < l''$.
4. Democratic2: $\Gamma' \prec_{Dem2} \Gamma''$ iff $\forall l' \in \Gamma' \setminus \Gamma'', \exists l'' \in \Gamma'' \setminus \Gamma' \text{ s.t. } l' < l''$.

**Proposition 18** For $s \in \{Elitist1, Elitist2, Democratic1, Democratic2\}$, $\prec_s$ is a strict partial order.

Elitist2 applies the principle Elitist1 (used in [39]) only to the elements that are not in the intersection of the sets $\Gamma'$ and $\Gamma''$ of defeasible rules (non-axiom premises). To see the difference, consider $\Gamma' = \{l_1, l_2\}$, $\Gamma'' = \{l_1, l_3, l_4\}$ where $l_2 < l_3, l_2 < l_4$. Then $\Gamma' \prec_{Elit1} \Gamma''$ but $\Gamma' \not\prec_{Elit1} \Gamma''$. Similarly, the Democratic2 principle applies the Democratic1 principle to elements that not contained in the intersection of the sets under comparison (the Democratic2 principle is used in [9]).

We then define a set ordering $\prec$ in terms of $\prec_s$. This is because we require that when comparing the sets $\Gamma$, $\Gamma'$ of defeasible rules (non-axiom premises) of arguments $B$ and $A$ respectively, then intuitively, $B$ is weaker than $A$ when $B$ is defeasible (plausible) and $A$ is strict (firm). In other words, we want to constrain the ordering so that for any non-empty $\Gamma$, if $\Gamma' = \emptyset$ then $\Gamma$ is strictly less than $\Gamma'$. We define $\prec$ because one cannot guarantee that this holds for every conceivable way of defining $\prec_s$ (it is easy to verify that $\prec_{Dem1}$ and $\prec_{Dem2}$ do not satisfy this constraint).

**Definition 18** [Ordering $\prec$] Let $\prec_s$ be a strict partial ordering. Then $\forall \Gamma', \Gamma''$, if $\Gamma' \neq \emptyset$, then:

1. If $\Gamma' = \emptyset$ then $\Gamma' \prec \Gamma''$, else
2. If $\Gamma' \prec_s \Gamma'$ then $\Gamma' \prec \Gamma''$

**Definition 19** [Last-link principle] $B \prec A$ under the last-link principle iff

1. $\text{LastDefRules}(B) \prec \text{LastDefRules}(A)$; or
2. $\text{LastDefRules}(B) = \emptyset, \text{LastDefRules}(A) = \emptyset$, and $\text{Premap}(B) \prec \text{Premap}(A)$

\[\text{Notice that it suffices to restrict } \prec_s \text{ to finite sets since } \text{ASPIC}^+ \text{ arguments are defined as finite (in Definition 4) and so their non-axiom premises/defeasible rules must be finite.}\]
When defining the weakest-link principle, we need to compare both the weakest non-axiom premises and weakest defeasible rules of the arguments. Notice that [39]'s definition of the weakest link principle implies an anomaly that is corrected here. [39]'s definition implies that if both $B$ and $A$ are strict (contain no defeasible rules), then $B ≺ A$ if there are non-axiom premises in $B$ that are $≺$ the non-axiom premises in $A$. However, if both $B$ and $A$ are firm (contain no non-axiom premises), then it is not the case that $B ≺ A$ if there are defeasible rules in $B$ that are $≺$ the defeasible rules in $A$. Thus there is an asymmetry in the way that premises and defeasible rules are compared. In what follows, the weakest link is defined so that the defeasible rules and non-axiom premises are treated in the same way.

**Definition 20 [Weakest-link principle]** $B ≺ A$ under the weakest-link principle iff:

1. If both $B$ and $A$ are strict, then $\text{Prem}_{ap}(B) ≺ \text{Prem}_{ap}(A)$, else;
2. If both $B$ and $A$ are firm, then $\text{DefRules}(B) ≺ \text{DefRules}(A)$, else;
3. $\text{Prem}_{ap}(B) ≺ \text{Prem}_{ap}(A)$ and $\text{DefRules}(B) ≺ \text{DefRules}(A)$

**Example 19 [Example 3 continued]** Given the preorderings $r \Rightarrow q \leq a \Rightarrow p$ and $\neg r \leq \neg q$ on defeasible rules and non-axiom premises in Example 1, and employing the abbreviations DR for DefRules and LDR for LastDefRules, we have:

- $\text{DR}(A) = \text{LDR}(A) = \{a \Rightarrow p\}$, $\text{Prem}_{ap}(A) = \{a\}$;
- $\text{DR}(B) = \text{LDR}(B) = \{\neg s \Rightarrow t, t \Rightarrow q\}$, $\text{Prem}_{ap}(B) = \{\neg s, r\}$;
- $\text{DR}(B_2) = \text{LDR}(B_2) = \emptyset$, $\text{Prem}_{ap}(B_2) = \{r\}$;
- $\text{DR}(C) = \text{LDR}(C) = \emptyset$, $\text{Prem}_{ap}(C) = \{\neg r\}$.

Then:

- $LDR(B) \prec_{E_{11}(2)} LDR(A)$ and so $B \prec A$ under the last-link principle.
- $\text{DR}(B) \prec_{E_{11}(2)} \text{DR}(A)$, but $\text{Prem}_{ap}(B) \not\prec_{E_{11}(2)} \text{Prem}_{ap}(A)$, so $B \not\prec A$ under the weakest-link principle.
- $LDR(B) \not\prec_{D_{12}(2)} LDR(A)$ and $\text{DR}(B) \not\prec_{D_{12}(2)} \text{DR}(A)$, so $B \not\prec A$ under the last or weakest-link principle.
- $\text{Prem}_{ap}(C) \prec_{E_{11}(2)} \text{Prem}_{ap}(B_2)$, so $C \prec B_2$ under the last or weakest-link principle.
- $\text{Prem}_{ap}(C) \prec_{D_{12}(2)} \text{Prem}_{ap}(B_2)$, so $C \prec B_2$ under the last or weakest-link principle.

We now show that the argument orderings defined in Definitions 19 and 20 are reasonable. Firstly, we note that in general, it does not hold that:

(*) $\leq$ satisfies condition 2 of a reasonable argument ordering iff $\leq$'s strict counterpart $\prec$ is a strict partial order.

However, under the assumption that $\prec$ is transitive, the left to right half of (*) does follow given that condition 2 implies that $\prec$ is irreflexive. To see this, it suffices to consider $\{C_1\} \subseteq A$, in which case condition 2 implies $C_1 \not\prec C_1$ (note that trivially, any argument is a strict extension of itself). Also, we note here that in the case that $\prec$ is defined under the last- or weakest-link principles, based on the orderings $\prec$, in Definition 17, $\prec$ is a strict partial order (Propositions 20 and 21 below). The right to left half of (*) then holds. This can be seen by inspection of the proofs of Propositions 22 and 23 below, which state that the last and weakest link argument orderings are reasonable. In these proofs, the contrapositive is shown: assuming that condition 2 does not hold implies that $\prec$ is not a strict partial order.
Proposition 20 Let $\preceq$ be defined according to the last-link principle. Then $\prec$ is a strict partial order.

Proposition 21 Let $\preceq$ be defined according to the weakest-link principle. Then $\prec$ is a strict partial order.

Proposition 22 Let $\preceq$ be defined according to the last-link principle. Then $\preceq$ is reasonable.

Proposition 23 Let $\preceq$ be defined according to the weakest-link principle. Then $\preceq$ is reasonable.

5.2 Reconstructing and extending the abstract logic approach as an instance of $\text{ASPIC}^+$

In [1, 2], Amgoud & Besnard present an abstract approach to defining the structure of arguments and attacks, based on Tarski’s notion of an abstract logic.

Definition 21 [Abstract Logic] An abstract logic is a pair $(\mathcal{L}, Cn)$, where $\mathcal{L}$ is a language and the consequence operator $Cn$ is a function from $2^{\mathcal{L}}$ to $2^{\mathcal{L}}$ satisfying the following conditions for all $X \subseteq \mathcal{L}$:

1. $X \subseteq Cn(X)$
2. $Cn(Cn(X)) = Cn(X)$
3. $Cn(X) = \bigcup_{Y \subseteq f X} Cn(Y)$
4. $Cn(\{p\}) = \mathcal{L}$ for some $p \in \mathcal{L}$
5. $Cn(\emptyset) \neq \mathcal{L}$

Here $Y \subseteq f X$ means that $Y$ is a finite subset of $X$. A set $X \subseteq \mathcal{L}$ is defined as consistent if $Cn(X) \neq \mathcal{L}$.

Amgoud & Besnard [1] note that the following properties hold:

6. If $X \subseteq X'$ then $Cn(X) \subseteq Cn(X')$ (monotonicity)
7. If $Cn(X) = Cn(X')$ then $Cn(X \cup Y) = Cn(X' \cup Y)$

[1] also restricts its focus to so-called adjunctive abstract logics:

8. $\forall x, y \in \mathcal{L}$ such that $Cn(\{x, y\}) \neq Cn(\{x\}), Cn(\{x, y\}) \neq Cn(\{y\}), \exists z$ such that $z \neq x, z \neq y$ and $Cn(\{z\}) = Cn(\{x, y\})$.

They then define arguments and various kinds of attack relations, and then investigate consistency properties of various types of attack relations when instantiating Dung’s framework with arguments and attacks. We discuss this part of their work in Section 6. We repeat here [2]’s notion of an undermining attack[6]. We also extend their abstract logic approach to accommodate a pre-ordering over the formulae in an abstract logic theory.

[5]For example, classical logic is adjunctive because of the conjunction connective

[6][2] call undermining “undercutting” but to be consistent with $\text{ASPIC}^+$’s terminology we rename it to ‘undermining’.
Definition 22 [Arguments and attacks in abstract logics] Let \((\mathcal{L}, Cn)\) be an abstract logic and \((\Sigma, \preceq)\) a theory in \((\mathcal{L}, Cn)\), where \(\Sigma \subseteq \mathcal{L}\) and \(\preceq\) a preordering over \(\Sigma\):

- an **AL-argument** is a pair \((X, p)\) such that: 1) \(X \subseteq \Sigma\); 2) \(X\) is consistent; 3) \(p \in Cn(X)\); 4) no proper subset of \(X\) satisfies (1-3).

- \((X, p)\) **AL-undermines** \((Y, q)\) if there exists a \(q' \in Y\) such that \(\{p, q'\}\) is inconsistent.

We formally define the notion of an **ASPIC\(^+\)** argumentation theory based on an abstract logic with preferences. This involves defining the set of strict rules in terms of the abstract-logic’s consequence notion but also relating the contrariness relation to the Tarskian notion of consistency. Note that the latter does not allow for defining asymmetric contrary relations, and so we have to assume that **ASPIC\(^+\)**’s contrariness relation is symmetric. Next, two conditions are needed to relate the contrariness relation to the Tarskian notion of consistency. Firstly, if two formulas are contradictories of each other then they are jointly inconsistent. Secondly, if two formulas are jointly inconsistent, then each of them has a consequence that is a contradictory of the other. Also, a knowledge base will represent the elements of an abstract logic theory as ordinary premises with the associated preordering. In what follows, note the slight abuse of notation, whereby if \(S = \{s_1, \ldots, s_n\}\), we write \(S \rightarrow p \in R_s\) to indicate any or some \(s_1, \ldots, s_n\) (letting the context disambiguate). Also, to avoid confusion we henceforth refer to the abstract logic notion of consistency as ‘AL-consistency’.

Definition 23 [AT and c-SAF based on abstract logic with preferences] Let \((\mathcal{L}', Cn)\) be an abstract logic and \((\Sigma, \preceq')\) a theory in \((\mathcal{L}', Cn)\). An abstract logic (AL) argumentation theory based on \((\mathcal{L}', Cn)\) and \((\Sigma, \preceq')\), is a pair \((AS, KB)\) such that \(AS\) is an argumentation system \((\mathcal{L}, \neg, R, \preceq)\) based on \((\mathcal{L}', Cn)\), where:

1. \(\mathcal{L} = \mathcal{L}'\); 
2. \(R_d = \emptyset, \leq = \emptyset\), and for all \(S \subseteq \mathcal{L}\) and \(p \in \mathcal{L}\), \(S \rightarrow p \in R_s\) iff: 1) \(p \in Cn(S)\), and 2) \(p \notin Cn(S')\) for all \(S' \subset S\)
3. \(\neg\) is defined such that:
   - (a) if \(\varphi \in \overline{\psi}\) then \(\psi \in \overline{\varphi}\);
   - (b) if \(\varphi \in \overline{\psi}\) then \(\{\varphi, \psi\}\) is AL-inconsistent;
   - (c) if \(\{\varphi, \psi\}\) is AL-inconsistent then there exists a \(\varphi' \in Cn(\varphi)\) such that \(\varphi' \in \overline{\psi}\).
   - (d) \(\overline{\psi}\) is nonempty for all \(\varphi\).

\(KB\) is a knowledge base \((K_n, \preceq')\) such that \(K_n = K_o = \emptyset, K_p = \Sigma\). 
\((A, C, \preceq)\) is the c-SAF based on an \((AS, KB)\), as defined in Definition 10.

We can then show that a c-SAF based on an abstract logic with preferences is well defined:

**Proposition 24** A c-SAF based on an AL argumentation theory is closed under contraposition, axiom consistent, c-classical, and well-formed.
Assuming \( \preceq \) is reasonable, Proposition 24 implies that all the results and rationality postulates in Sections 4.1 and 4.2 hold for c-SAFs based on an abstract logic with preferences. However, note that these c-SAFs are instantiated by ASPIC\(^+\) arguments and undermining attacks (since rebuts and undercuts only apply to defeasible rules). The question naturally arises as to whether they are equivalent to the AL arguments and undermining attacks in Definition 22. We first show that the attacks are indeed equivalent. To do so, we define the notion of an AL-c-SAF:

**Definition 24** Let an ASPIC\(^+\)-AL-undermining attack be defined in the same way as an undermining attack in Definition 7, with ‘\( \text{Conc}(A) \vdash \neg \varphi \)’ replacing ‘\( \text{Conc}(A) \notin \varphi \)’. Then an AL-c-SAF defined by \( (AS, KB) \) is defined in the same way as a c-SAF, with ‘\( (X, Y) \in C \) iff \( X \) ASPIC\(^+\)-AL-undermines \( Y \)’ replacing ‘\( (X, Y) \in C \) iff \( X \) attacks \( Y \)’ in Definition 10.

Given Lemma 41 in Section 8.5, which shows that \( \text{Conc}(A) \vdash \neg \varphi \) iff \{Conc\( (A) \), \( \varphi \)\} is AL-inconsistent, then the following result shows that ASPIC\(^+\)’s undermining attacks faithfully reconstruct abstract logic undermining attacks:

**Proposition 25** Let \( (AS, KB) \) be based on \( (\mathcal{L}', Cn) \) and \( (\Sigma, \preceq') \). Let \( \Delta_1 \) be the c-SAF defined by \( (AS, KB) \), and \( \Delta_2 \) the AL-c-SAF defined by \( (AS, KB) \). Then, for \( T \in \{\text{complete, grounded, preferred, stable}\} \), \( E \) is a \( T \) extension of \( \Delta_1 \) iff \( E \) is a \( T \) extension of \( \Delta_2 \).

Now note that we do not have an equivalence between ASPIC\(^+\) arguments and AL arguments, because the latter imposes a subset minimality condition on the premises. This condition is not imposed on ASPIC\(^+\) arguments in Definition 4, and neither is it implied by the definition of \( R_\subset \) in Definition 23\(^3\). Consider the following counterexample. Given \( q \in Cn(\{p\}) \), \( s \in Cn(\{p, r\}) \) and \( q \in Cn(\{s\}) \), we obtain \( R_\subset = \{p \rightarrow q; p, r \rightarrow s; s \rightarrow q\} \). Then we have the strict arguments \( \{p \vdash q \} \) and \( \{p, r \vdash q \} \) where the latter is not subset minimal.

In general, minimality of premise sets is undesirable. Suppose a defeasible rule \( p \Rightarrow q \) and a strict rule \( p, r \Rightarrow q \); then we clearly do not want to rule out an argument for \( q \) with premises \( p \) and \( r \), since it could well be stronger than the defeasible argument. However, since the ASPIC\(^+\) arguments defined by an AL argumentation theory are strict, we define here the notion of premise minimal ASPIC\(^+\) arguments and show an equivalence with AL arguments.

**Definition 25** [Premise minimal ASPIC\(^+\) arguments] Let for any argument \( A, A_\sim \) be any argument such that \( \text{Prem}(A_\sim) \subseteq \text{Prem}(A) \) and \( \text{Conc}(A_\sim) = \text{Conc}(A) \). Given a set of ASPIC\(^+\) arguments \( A \), let \( A^- = \{A \in A | \text{there is no } A_\sim \in A \text{ such that } \text{Prem}(A_\sim) \subset \text{Prem}(A)\} \) be the premise minimal arguments in \( A \).

**Proposition 26** Let \( (AS, KB) \) be based on \( (\mathcal{L}', Cn) \) and \( (\Sigma, \preceq') \). Then \( A \) is a c-consistent premise minimal argument on the basis of \( (AS, KB) \) iff \( (\text{Prem}(A), \text{Conc}(A)) \) is an abstract logic argument on the basis of \( (\Sigma, \preceq') \).

We can then show that for c-SAFs and SAFs, when restricting consideration to arguments with minimal premise sets, the conclusions of arguments in complete extensions remains unchanged, under the assumption that an argument cannot be strengthened by just adding premises.

---

\(^3\)Note that the subset minimality condition enforces the requirement that an argument contains only the premises relevant to proving the argument’s conclusion, and that this requirement is met by ASPIC\(^+\) arguments, since all premises of an argument are used in deriving its conclusion.
Let $\Delta$ be the $(c)$-SAF $(A, C, \preceq)$ defined on the basis of an $AT$ for which $\preceq$ is defined such that for any $A \in A$, $A \preceq A_\infty$.

Let $\Delta_-$ be the premise minimal $(c)$-SAF $(A^-, C^-, \preceq^-)$ where:

- $A^-$ is the set of premise minimal arguments in $A$.
- $C^- = \{(X,Y) | (X,Y) \in C, X, Y \in A^- \}$.
- $\preceq^- = \{(X,Y) | (X,Y) \in \preceq, X, Y \in A^- \}$.

Then for $T \in \{\text{complete, grounded, preferred, stable}\}$, $E$ is a $T$ extension of $\Delta$ iff $E'$ is a $T$ extension of $\Delta^-$, where $E' \subseteq E$ and $\bigcup_{X \in E} \text{Conc}(X) = \bigcup_{Y \in E'} \text{Conc}(Y)$.

**Corollary 28** Given $\Delta$ and $\Delta^-$ as defined in Proposition 27:

1. $\varphi$ is a $T$ credulously (sceptically) justified conclusion of $\Delta$ iff $\varphi$ is a $T$ credulously (sceptically) justified conclusion of $\Delta^-$.  
2. $\Delta^-$ satisfies the postulates direct consistency, indirect consistency and sub-argument closure.

The assumption that arguments are not strengthened by adding premises is not entirely innocent, since it is not satisfied by all ways of defining $\preceq$. Consider a $(c)$-SAF $(A, C, \preceq)$ defined by an AL argumentation theory, where $\preceq$ is defined on the basis of the democratic relations $\preceq_{\text{dem1}}$ or $\preceq_{\text{dem2}}$, and suppose arguments $A_\infty$ and $A$ with $\text{Prem}(A_\infty) = \{p\}$, $\text{Prem}(A) = \{p, q\}$ and $p \prec q$. Then $\{p\} \prec_{\text{dem1}} \{p, q\}$ and $\{p\} \prec_{\text{dem2}} \{p, q\}$, and so $A_\infty \prec A$, violating the above assumption. However, the assumption is satisfied by the elitist orderings in Definition 17, as implied by the following:

**Proposition 29** Let $(A, C, \preceq)$ be defined by an AL argumentation theory, where $\preceq$ is defined under the weakest or last link principles, based on the set orderings $\preceq_{\text{el1}}$ or $\preceq_{\text{el2}}$. Then $\forall A, B \in A$, if $\text{Prem}(A) \subseteq \text{Prem}(B)$, then $A \neq B$.

In conclusion, we have shown that all results and rationality postulates hold for $(c)$-SAF’s based on abstract logics with preferences. We have also shown that for the $(c)$-SAF $\Delta = (A, C, \preceq)$ defined by $(\Sigma^\prime, C^n)$ and $(\Sigma, \preceq^\prime)$, the AL undermining attacks and $C$ are equivalent in the complete extensions that they generate, and the AL arguments and premise minimal ASPIC$^+$ arguments in $A$ are equivalent. Proposition 27 and Corollary 28 then imply that:

**Remark 30** 1) We have reconstructed [1, 2]’s abstract logic approach to argumentation (which assumes no preference relation over $\Sigma$) in the ASPIC$^+$ framework, in that the justified conclusions of a Dung framework instantiated by AL arguments and AL undermining attacks, are exactly those of $\Delta$. We have also shown that Section 4.2’s rationality postulates hold for Amgoud & Besnard’s approach.

2) Given that we have extended the abstract logic approach to accommodate preferences, consider a preference-based argumentation framework (see Section 2.1) $\Gamma = (A', C', \preceq')$ defined by $(\Sigma, \preceq)$ and $(\Sigma', C_n)$, where $A'$ and $C'$ are the AL arguments and AL undermining attacks. Then, under the assumption that $\preceq'$ does not strengthen arguments when adding premises, the justified conclusions of $\Gamma$ (specifically the Dung framework instantiated by arguments and defeats defined by $\Gamma$) are exactly those of $\Delta$. This assumption is satisfied when defining $\preceq'$ under the last or weakest link principles, based on Definition 17’s elitist set orderings that utilise the preordering $\preceq$ over formulae in $\Sigma$. We have also shown that Section 4.2’s rationality postulates hold for Amgoud & Besnard’s approach extended with preferences.
We conclude by observing that Amgoud & Besnard investigate the consistency of extensions of a Dung framework instantiated by the arguments and attacks defined by an abstract logic. Specifically, they consider whether for any extension the union of the premises of the extension’s arguments is AL consistent. We now show that this is equivalent to indirect consistency, and then refer to this result in Section 6 in which we compare ASPIC\(^+\) and the abstract logic approach.

**Proposition 31** Let \(\Delta\) be the c-SAF defined by an AL argumentation theory. Then for any complete extension \(E\) of \(\Delta\):
\[
S = \{ \phi | \phi \in \text{Prem}(A), A \in E \}
\]
is AL-inconsistent iff
\[
S' = Cl_{R_c}(\{ \text{conc}(A) | A \in E \})
\]
is inconsistent.

### 5.3 Classical logic instances of the ASPIC\(^+\) framework

The previous section’s results allow us to reconstruct classical logic approaches to argumentation as a special case of ASPIC\(^+\), and in so doing extend the classical approaches of [13, 14] with preferences. We also prove a relation with Brewka’s preferred subtheories.

#### 5.3.1 Defining classical logic instantiations of ASPIC\(^+\)

Much recent work on structured argumentation formalises arguments as minimal classical consequences from consistent and finite premise sets in standard propositional or first-order logic [5, 13, 14, 26]. Writing ‘\(\vdash_c\)’ to denote classical consequence, then:

**Definition 26** Given a propositional or first order theory \(\Sigma\), an argument is of the form \((\Gamma', \alpha)\) such that \(\Gamma'\) is finite, \(\Gamma' \subseteq \Sigma\), \(\Gamma' \vdash_c \alpha\) and \(\Gamma'' \not\vdash_c \alpha\) for all \(\Gamma'' \subset \Gamma'\)

Since classical logic can be specified as a special case of a Tarskian abstract logic, a classical argumentation theory and its c-SAF is defined as in Definition 23, given a standard propositional or first-order language \(L'\) and a theory \(\Sigma \subseteq L'\).

Amongst the above-cited works on classical argumentation, [5] and [26] adopt a Dung-style semantics, where only [5] considers preferences. Let us first consider [26]. They define several alternative notions of attack and investigate their properties, including the rationality postulates of [18] studied in this paper. They show that the only two attack relations that are ‘well behaved’, in the sense that they satisfy consistency postulate for all the semantics, are the so called ‘direct undercut’ and ‘direct defeater’:

- \(Y\) directly undercuts \(X\) if \(\text{conc}(Y) \equiv \neg p\) for some \(p \in \text{Prem}(X)\)
- \(Y\) directly defeats \(X\) if \(\text{conc}(Y) \vdash_c \neg p\) for some \(p \in \text{Prem}(X)\)

Although [26] do not study this paper’s undermining attack, their direct undercut and defeats are equivalent to undermining attacks in the sense that the complete extensions generated are the same. For direct defeats, this result is shown by Proposition 25. For direct undercut, it suffices to adapt the proof of Proposition 25, showing that: 1) if \(Y\) undermines \(X\), then letting \(\text{conc}(Y) = q\), by the symmetry of classical negation \(q = \neg p\) for some \(p \in \text{Prem}(X)\) and so \(\text{conc}(Y')) \equiv \neg p\); 2) if \(Y\) directly undercuts \(X\) then \(Y\) directly defeats \(X\), and so as already shown, \(Y\) undermines \(X\).

It follows from the above and the results and discussion in Section 5.2, that we have reconstructed and extended with preferences, [26]’s variants with direct undercut and
direct defeat, and shown that [18]’s postulates are satisfied for these extensions (recall that [26]’s other variants violate the consistency postulate even without preferences). Indeed, we argue that extending classical logic approaches with preferences is of particular importance, given that (as shown in [20, 26]) the preferred/stable extensions generated from a Dung framework instantiated by arguments and direct undercuts or defeats, simply correspond to the maximal consistent subsets of the theory $\Sigma$ from which the arguments are defined\(^8\). Intuitively, one would expect this correspondence given that classical logic does not provide any logical machinery for arbitrating conflicts (in contrast with the use of undercuts, assumptions and negation as failure in defeasible and non-monotonic logics (as discussed in Section 2.3)). One must therefore resort to some meta-logical mechanism, such as preferences, if argumentation is to usefully be deployed in resolving inconsistencies in a classical-logic setting.

We conclude by noting that [5] make use of preferences to determine the success of two of [26]’s variants of attack, and show that this leads to violation of the consistency postulates. We will discuss this in detail in Section 6.

5.3.2 Brewka’s Preferred Subtheories as an instance of the ASPIC\(^+\) framework

Brewka’s preferred subtheories [16] models the use of an ordering over a classical propositional or first order theory $\Gamma$, in order to resolve inconsistencies. It has therefore been used to both formalise default reasoning and belief revision [17].

**Definition 27** A default theory $\Gamma$ is a tuple $(\Gamma_1, \ldots, \Gamma_n)$, where each $\Gamma_i$ is a set of classical first order formulae. A preferred subtheory is a set $\Sigma = \Sigma_1 \cup \ldots \cup \Sigma_n$ such that for $i = 1 \ldots n$, $\Sigma_1 \cup \ldots \cup \Sigma_i$ is a maximal (under set inclusion) consistent subset of $\Gamma_1, \ldots, \Gamma_i$

Intuitively, a preferred subtheory is obtained by taking a maximal under set inclusion consistent subset of $\Gamma_1$, extending this with a maximal consistent subset of $\Gamma_2$, extending this with a maximal consistent subset of $\Gamma_3$, and so on. We can reconstruct preferred subtheories as an instance of the ASPIC\(^+\) framework.

**Definition 28** Let $\Gamma$ be a default theory $(\Gamma_1, \ldots, \Gamma_n)$, and $\forall \alpha, \beta \in \Gamma$, $(\alpha, \beta) \in \leq'$ iff $\alpha \in \Gamma_i, \beta \in \Gamma_j, i > j$. Let $(AS, KB)$ be a classical argumentation theory defined with respect to $(\Gamma, \leq')$ as described in Section 5.3.1 (with $\Gamma$ replacing $\Sigma$).

Let $\Delta$ be the c-SAF $(A, C, \preceq)$ defined on the basis of $(AS, KB)$, where $\preceq$ is defined under the weakest or last link link principle, and on the basis of the $\preceq_{Eli1}$ ordering on sets. We say that $\Delta$ is the c-SAF corresponding to $\Gamma$.

**Theorem 32** Let $(A, C, \preceq)$ be a c-SAF corresponding to a default theory $\Gamma$, and for any $\Sigma \subseteq \Gamma$, let $\text{Args}(\Sigma) \subseteq A$ be the set of all arguments with premises taken from $\Sigma$. Then:

1) If $\Sigma$ is a preferred subtheory of $\Gamma$, then $\text{Args}(\Sigma)$ is a stable extension of $(A, C, \preceq)$.

2) If $E$ is a stable extension of $(A, C, \preceq)$, then $\bigcup_{A \in E} \text{Prem}(A)$ is a preferred subtheory of $\Gamma$.

The above theorem paves the way for applying argument-game proof theories and labelling algorithms for the stable semantics [33], to preferred subtheories, as well as studying the preferred subtheories approach under the full range of semantics defined for Dung frameworks.

\(^8\)In the sense that the union of formulae in the supports of arguments in each preferred/stable extension is a maximal consistent subset of $\Sigma$. 

26
6 A Critical Assessment of Some Related Work

6.1 Comparison with General Frameworks for Argumentation

In this section we compare ASPIC$^+$ to related work. To start with, the inclusion of defeasible rules in ASPIC$^+$ requires some explanation, given that much current work formalises the construction of arguments as deductive [1, 2], and in particular classical [14, 26] inference. These approaches regard argumentation-based inference as a form of inconsistency handling in deductive logic; the alleged advantage being that the logic of deductive inference is well-understood [14, p. 16]. This raises the question of whether defeasible inference rules are needed at all. Our answer is that the research history in our field shows that at best only part of argumentation can be formalised as inconsistency handling in deductive logic. To start with, the distinction between strict and defeasible inference rules has a long history in AI research on argumentation [29, 30, 35, 36, 37, 38, 40, 42, 43], so a truly general framework for structured argumentation must include this distinction. Pollock in particular provides philosophical arguments that appeal to epistemological accounts of human reasoning, so that the modelling of defeasible rules is a particularly salient requirement in light of the bridging role (discussed in Sections 1 and 2) that argumentation plays between human and formal logic-based models of reasoning.

Moreover, conceptually, defeasible reasoning is not about handling inconsistent information but about making deductively unsound but still rational ‘jumps’ to conclusions on the basis of consistent but deductively inconclusive information. Consider the following well-known example, with the given information that quakers are normally pacifists, that republicans are normally not pacifists and that Richard Nixon was both a quaker and a republican. A defeasible reasoner is then interested in what can be concluded about whether Nixon was a pacifist while consistently accepting all the given information. The reason that they are jointly consistent is that ‘If $q$ then normally $p$’ and ‘$q$’ does not deductively imply $p$ since things could be abnormal: Nixon could be an abnormal quaker (or republican). A defeasible reasoner therefore does not want to reject any of the above statements, but rather wants to assume whenever possible that things are normal, in order to jump to conclusions about Nixon in the absence of evidence to the contrary. In other words, defeasible reasoning is not about inconsistency handling but about making uncertain inferences from consistent (though deductively inconclusive) premises. Therefore, attempts to formalise defeasible reasoning as inconsistency handling are at least unnatural. Moreover, the literature on nonmonotonic logic contain arguments that such attempts are prone to validating counterintuitive inferences (see e.g. [17, 25]). We can therefore conclude that, given the research literature, it makes sense to include defeasible inferences in models of argumentation, and therefore any account of argumentation that claims to be general should leave room for them.

A number of works purport to be general approaches to argumentation. A well-known and established framework is that of assumption based argumentation (ABA) [15], which is shown to be a special case of the ASPIC$^+$ framework in [39] in which arguments are built from assumption premises $K_a$ and strict inference rules. Our work is relevant for ABA in that we provide conditions under which ABA satisfies [18]’s consistency postulates (ABA does not in general satisfy these postulates).

More recently, Amgoud & Besnard [1, 2] proposed the abstract-logic approach (AL) to defining structured argumentation. ASPIC$^+$ is considerably more complex than AL. Firstly because ASPIC$^+$ models the use of preferences to resolve attacks, so it has to distinguish between attack and defeat, and secondly because ASPIC$^+$ combines
deductive and defeasible argumentation, which means that not only the premises, but also the defeasible inferences of an argument can be attacked. This requires that the arguments’ structure be made explicit in order to know which parts of an argument can be attacked. By contrast, if all inferences are certain, then arguments can only be attacked on their premises so their internal structure is irrelevant for their evaluation.

In Section 5.2 we straightforwardly extended AL with preferences and then reconstructed the extended AL in ASPIC+. However, AL cannot accommodate and so (when reconstructing in ASPIC+) generate defeasible inferential rules. This is because the inferential reasoning from premises to conclusion is not rendered explicit, but rather is encoded in AL’s single consequence operator, which cannot distinguish between strict and defeasible inference rules: there is no way to distinguish p’ and S’s for which \( p \in Cn(S) \) implies that \( S \rightarrow p \) should be in \( R_s \) or \( S \Rightarrow p \) should be in \( R_d \). Furthermore, recall that in Section 5.2 we argued that AL’s subset minimality condition on premises is not appropriate when accounting for defeasible inference rules.

The inappropriateness of accommodating defeasible argumentation in AL, is further illustrated when considering whether ASPIC+’s notion of an argument generates an abstract logic. If this is the case for a given instance of ASPIC+, then all of [1, 2]’s results hold for this instance. Suppose an ASPIC+ AT and \( Cn \) defined as follows:

- \( p \in Cn(X) \) iff there exists an ASPIC+ argument \( A \), with \( Conc(A) = p \) and \( Prem(A) = X \).

It can be shown that conditions (1), (2) and (3) in the definition of an abstract logic (Definition 21) are satisfied. However (4) is in general not satisfied. Consider an AT with \( K = \{ p \} \), \( R_s = \emptyset \) and \( R_d = \{ p \Rightarrow q \} \). Also, (5) is not in general satisfied.

Consider any AT with \( K = \emptyset \) and \( R_d = \{ \Rightarrow p \mid p \in \mathcal{L} \} \).

Of course, many instances of ASPIC+ will satisfy (4) and (5), and thus generate abstract logics. But then we should interpret [1, 2]’s results, as they apply to these instances, with care. In particular, the notion of consistency of an abstract logic behaves in an unexpected way. Recall that Amgoud & Besnard investigate whether for any Dung-extension \( E \), the set \( \bigcup_{(X,p) \in E} X \) is AL consistent (see end of Section 5.2).

Now, consider an ASPIC+ AT formalising the above Nixon example in a language \( \mathcal{L} \) including atoms \( p, r \) and \( q \) respectively denoting ‘Nixon is a pacifist’, ‘Nixon is a republican’ and ‘Nixon is a quaker’ and a connective \( \Rightarrow \) for default conditionals. Let the contrariness relation correspond to classical negation, \( \mathcal{R}_s \) contain all propositionally valid inferences (including \( p, \neg p \Rightarrow \varphi \) for any \( \varphi \in \mathcal{L} \) and \( \mathcal{R}_d \) contain a defeasible modus ponens scheme \( \varphi, \varphi \Rightarrow \psi \Rightarrow \psi \). Then if \( K = \{ q, r, q \Rightarrow p, r \Rightarrow \neg p \} \), any Dung-extension contains all elements of \( K \) as arguments but does not contain arguments for both \( p \) and \( \neg p \) so any such extension satisfies indirect consistency (the closure under strict rules is consistent). However, in the abstract logic generated by equation •, the set \( \{ q, r, q \Rightarrow p, r \Rightarrow \neg p \} \) is AL inconsistent, since there exists an ASPIC+ argument for every \( \varphi \) by combining the defeasible arguments for \( p \) and \( \neg p \) (even though this argument is not in any extension).

This discrepancy is caused by the fact that an abstract logic’s consequence operator cannot distinguish between strict and defeasible inferences, and so regards a set \( S \) as inconsistent if the closure of \( S \) under defeasible rules is directly inconsistent. But this is far too strong a consistency requirement, since the very idea of defeasible reasoning

\[9\] Since for AL, consistency requirements on arguments are added on top of a given consequence notion \( Cn \), we cannot incorporate consistency requirements into the definition of \( Cn \), and so assume ASPIC+ without the restriction to c-consistent arguments.
is that one’s knowledge need not be closed under defeasible inference, since defeasible inference rules can be defeated even if all their antecedents hold.

Concluding our comparison, the abstract logic approach provides a very interesting and insightful generalisation of earlier work on classical argumentation, but does not apply to mixed strict and defeasible argumentation, such as modelled in ASPIC+ and earlier by many others. We should here emphasise our view that deductive approaches certainly do have their place in the study and application of of argumentation. However, we argue that a truly general account of argumentation should also accommodate the use of defeasible inference rules.

Amgoud & Besnard [1, 2] also use the abstract logic approach to make some informal negative claims about the suitability of Dung-style semantics. First of all, they informally claim [2] that to satisfy the consistency postulates, an attack relation should be valid in the sense that when two arguments have jointly AL inconsistent premises, they should attack each other. However, this informal claim should be read with care: what they formally show is that validity is a sufficient condition for consistency. Their results do not preclude “invalid” attack relations, such as undermining attacks, from satisfying consistency. Indeed, in [39] and this paper we have identified alternative sufficient conditions for consistency. Furthermore, Amgoud & Besnard themselves show that, under the assumption that a Dung framework contains all arguments that can be logically constructed, their notion of consistency of extensions is satisfied assuming the AL undermining attacks in Definition 22, which is, of course, consistent with our more general result showing that AL satisfies all of [18] postulates (Remark 30 and Proposition 31 in Section 5.2). Amgoud & Besnard regard this assumption as problematic. However, we regard this not as a problem of the attack relation, but of the reasoner: if an imperfect reasoner is modelled who cannot be relied on to produce all relevant arguments, then perfect results cannot be expected. Furthermore, as argued in Section 2.2, requiring that attacks be ‘valid’ goes against the dialectical role of attacks and has the computational problem in that it can give rise to infinitely many attacks.

Finally, Amgoud & Besnard also claim that their notion of rebutting attack – \((X, p) \text{rebuts} (Y, q)\) if \(\{p, q\}\) is AL inconsistent – is not well-behaved, since it violates consistency. However in this paper we have shown that rebut attacks are appropriate in systems with defeasible inference rules.

### 6.2 Comparison with other works on Preference-based Argumentation

We now consider approaches that accommodate preferences to determine which attacks succeed as defeats. Recently, both Kaci [28] and Amgoud & Vesic [6, 7, 8, 10] have addressed the issue of how consistency can be ensured for instantiations of the preference-based argumentation frameworks (PAFs) [4] reviewed in Section 2.1. They all argue that instantiations of standard PAFs have problems with unsuccessful asymmetric attacks. [28] argues that all attacks should therefore be symmetric. However, [1] show that for classical argumentation this would still lead to inconsistency problems. Nevertheless, [6, 7, 10] also criticise ‘standard’ PAF approaches, arguing that unsuccessful asymmetric attacks may violate consistency. As a solution they propose that unsuccessful asymmetric attacks should result in rejection of the attacker even if it is not attacked by any argument. However, our consistency results obviate the need for reversing unsuccessful attacks. We have shown that by taking into account the structure of arguments, one can show that if \(A\) unsuccessfully attacks \(B\), then either some
sub-argument of $B$ defeats $A$, or $B$ can be extended to an argument that defeats $A$, and that this result is key for showing consistency as discussed in Section 4.3.

But how do our consistency results square with [6, 7, 10]'s examples of inconsistent $PAFs$? [7] give a semiformal example which we described earlier in Section 4.3 (and which is also described in terms of uninstantiated abstract arguments in [10]). Recall that an expert’s argument $A$, that a given violin is a Stradivarius ($s$) and therefore expensive ($e$), is asymmetrically attacked by a child’s argument $B$ that it is not a Stradivarius ($\neg s$). The greater reliability of the expert’s assertion about the violin means that $A$ is preferred to $B$ so that $B$ does not defeat $A$. We observed that inconsistency is not violated under this paper’s attack definition of conflict free, since $\{A, B\}$ is not conflict free and so not admissible. However, even under the defeat definition, we can see that [7]'s suggested problem arises only when failing to take into account all arguments. Formalising the example in $ASPIC^+$, $A = [s; s \Rightarrow e]$ where $s$ is an ordinary premise, and $A' = [s]$ is a sub-argument of $A$. $B = [\neg s]$ where $\neg s$ is an ordinary premise. Hence $B$ attacks $A$ on $A'$, and the expert’s greater reliability means that $A'$ is preferred to $B$ and so $B$ does not defeat $A$. However, one must also then acknowledge that $A'$ rebut attacks and defeats $B$, so that $\{A, B\}$ is not admissible.

In [6, 10], Amgoud & Vesic give a formal classical logic instantiation of a $PAF$ that demonstrates inconsistency. Figure 2-a) shows the equivalent arguments ($\Rightarrow$ denotes material implication) and attacks obtained by the corresponding classical logic $ASPIC^+$ instantiation (see Section 5.3.1), where $\mathcal{K}_p = \{x, \neg y, x \Rightarrow y\}$, $\mathcal{K}_n = \mathcal{K}_a = 0$. Then,
assuming \( x >' ¬y, x >' x ⊃ y \), they state that \( A_1 \) is strictly preferred to the other arguments, all of which are equally preferred. They thus obtain the defeat graph shown in Figure 2-b), and so the single stable extension \( \{ A_1, A_2, A_3, A_5 \} \), which violates consistency.

However, a different outcome is obtained in ASPIC+ . Here it is crucial that [6]'s use of the premise ordering to resolve attacks (which is taken from [4]), differs from our Definition 8 in which if \( A \) undermines \( B \) on premise \( p \), then \( A \) defeats \( B \) if \( A \neq p \). However, in [6], \( A \) defeats \( B \) if \( A \neq B \) based on a comparison of all premises of \( B \). This makes a crucial difference. Since both \( A_4 \) and \( A_5 \) have \( ¬y \) as weakest premise, \( A_4 \) and \( A_5 \) are equally preferred in [6]. However, we have that \( A_5 \)'s attack on \( A_4 \) is on \( A_4 \)'s subargument \( A_1 \), so the comparison is between \( A_5 \) and \( A_1 \). Now since \( x >' ¬y \), we have that \( A_1 \) is strictly preferred to \( A_5 \), so \( A_5 \) does not defeat \( A_4 \), so \( A_4 \) strictly defeats \( A_5 \) (as visualised in Figure 2-c)). But then a set including \( \{ A_1, A_2, A_3, A_5 \} \) is not a stable extension, since it does not defend \( A_5 \) against \( A_4 \). Instead, \( \{ A_1, A_4, A_2 \} \) is stable and satisfies consistency. We prefer our approach over [5, 6], since we do not see why the preference of the premise \( ¬y \) of \( A_4 \), which is irrelevant to the conflict on \( x \), should be relevant in resolving this conflict. Note here that the crucial point is that the structure of arguments and the nature of attack should be taken into account when applying preferences. In this case it is crucial to see that \( A_5 \)'s attack on \( A_4 \) was a direct attack on \( A_4 \)'s sub-argument \( A_1 \).

The issue also arises in different ways. Consider the ASPIC+ example, with \( \mathcal{K}_p = \{ p, q \}, \mathcal{K}_q = \mathcal{K}_r = \emptyset, \mathcal{R}_a = \emptyset, \mathcal{R}_b = \{ p \Rightarrow r; q \Rightarrow ¬r; ¬r \Rightarrow s \} \). We then have the arguments and attacks in Figure 2-d). Then, assuming \( p \Rightarrow r > q \Rightarrow ¬r, ¬r \Rightarrow s > p \Rightarrow r \), the argument ordering \( B_2 \prec A_2, A_2 \prec B_3 \) is generated by the last link principle. A PAF modelling then generates the defeat graph in Figure 2-e), so obtaining the single extension (in whatever semantics) \( \{ A_1, B_1, A_2, B_3 \} \). So not only \( A_2 \) but also \( B_1 \) is justified. However, not only are \( A_2 \) and \( B_3 \) based on arguments with contradictory conclusions, but the sub-argument closure postulate is violated; \( B_3 \) is justified, but its sub-argument \( B_2 \) is not. The problem arises because the PAF modelling cannot recognise that \( A_2 \) attacks \( B_3 \) on its sub-argument \( B_2 \), so we should compare \( A_2 \) with \( B_2 \), and not \( B_3 \). Now since \( B_2 \prec A_2 \), then \( A_2 \) defeats \( B_3 \), so the single extension (in whatever semantics) is \( \{ A_1, B_1, A_2 \} \) and we have that \( A_2 \) is justified and both \( B_2 \) and \( B_3 \) are overruled, as visualised in Figure 2-e). Note that these problems are not due to the use of defeasible rules or the last-link ordering. Consider a classical logic instantiation of ASPIC+ in which \( \mathcal{K}_n = \mathcal{K}_q = \emptyset, \mathcal{K}_p = \{ p, q, ¬p \} \) and \( q >' ¬p >' p \). The following arguments can be constructed: \( A_1 = p, A_2 = q, A_1, A_2 \rightarrow p \land q \) and \( B = ¬p \). Then, \( A_1 \) and \( B \) attack each other and \( B \) attacks \( A_3 \) (on \( p \)). Suppose arguments are compared based on the weakest link principle, applying either of Section 5.1’s democratic principles to the premise sets. Then \( A_1 \prec B \) and \( B \prec A_3 \). The PAF for this example then generates a stable extension containing \( A_3 \) and \( B \), which again violates sub-argument closure. In ASPIC+ we instead obtain that \( B \) defeats \( A_3 \) on \( A_1 \), so the correct outcome is obtained.

Concluding, [6, 7, 10] are right that PAFs need to be repaired, but the proper repair is not to change definitions at the abstract level but to make the structure of arguments and the nature of attack explicit. We have seen that seeming problems with unsuccessful asymmetric attack at the abstract level disappear if the structure of arguments and the nature of attack are specified, and that seeming violations of postulates do not occur if the success of an attack on an argument \( X \) is based on a preference-based comparison on the sub-argument of \( X \) that is attacked. We have also seen in this paper that there are reasonable notions of attack that result in defeat irrespective of preferences, such
as ASPIC$^+$’s undercutting, assumption and contrary attacks. A framework that does not make the structure of arguments explicit cannot distinguish between preference dependent and independent attacks.

Finally, note that as well as reversing asymmetric attacks, Amgoud & Vesic [6, 7, 8, 10] also propose a solution to the problematic cases they identify, and that we have countered above, by using the preference ordering over arguments to define an ordering over sets of arguments, privileging those that are conflict free under the attack relation. However this precludes the dialectical use of preferences in deciding the success of attacks between individual arguments (as described in Section 2); it is not clear how their use of preferences can be accounted for in dialogues and proof theoretic argument games. Furthermore, they do not show satisfaction of [18]’s postulates, except in the case of stable extensions, where they show that consistency is satisfied, via a correspondence with Brewka’s preferred subtheories [16]. However, we have shown this correspondence without reversing asymmetric attacks, or applying preferences over sets of arguments.

7 Conclusions

A newcomer to the area of abstract argumentation theory might legitimately question its added value above and beyond the conceptual insights yielded by its uniform characterisation of the inference relations of non-monotonic formalisms. This paper began with a response to this rhetorical question. Argumentative characterisations of inference encapsulate the dynamic and dialectical processes of reasoning familiar in everyday debate and discourse$^{10}$. It thus serves to both bridge formal logic and human reasoning in order that the one can inform the other, and support communicative interactions in which heterogeneous agents jointly reason and infer in the presence of uncertainty and conflict. We then discussed the declarative and procedural roles that attacks, preferences and defeats should play in the context of this value proposition, and then reviewed and modified [39]’s ASPIC$^+$ framework in light of this discussion. Specifically, the attack relation’s denotation of the mutual incompatibility of information in arguments determines whether a given set of arguments is conflict free, as distinct from their possibly preference dependent dialectical use as defeats.

ASPIC$^+$ provides an account of argumentation that combines Dung’s argumentation theory with structured arguments, attacks and the use of preferences. The added structure accommodates a range of concrete instantiating logics, to the extent that one can meaningfully study and prove satisfaction of rationality postulates. While the account retains the dialectical apparatus of Dung’s theory, one must additionally show that ASPIC$^+$’s intermediate level of abstraction allows for a broad range of instantiations, if one is to continue to appeal to the above stated value proposition of argumentation. To this end, we have argued that any general account should accommodate both the traditional use of defeasible inference rules as well as deductive approaches that essentially model non-monotonicity as inconsistency handling. [39]’s version of ASPIC$^+$ reconstructed approaches that use defeasible inference rules (e.g., [38, 40]) and showed that assumption-based argumentation [15] and systems using argument schemes can be formalised in ASPIC$^+$. The modelling of defeasible rules inevitably introduced a degree of complexity that exceeds that of other proposals for general frameworks. How-

---

$^{10}$Indeed, recent empirically validated work in cognitive science and psychology claims that the cognitive capacity for human reasoning evolved primarily in order to assess and counter the claims and arguments of interlocutors in social settings [31].
ever we have argued that a truly general framework for structured argumentation must include defeasible rules. In this paper we adapted ASPIC+ to additionally accommodate deductive approaches that require arguments to have consistent premises, and then showed that the adapted ASPIC+, with the revised definition of conflict free, satisfies key properties of Dung frameworks and [18]'s rationality postulates under some intuitive assumptions. We then formalised instantiation of the adapted ASPIC+ by Tarskian (and in particular classical) logic theories extended with preferences, thus demonstrating satisfaction of rationality postulates by these instantiations, and paving the way for the study of other non-classical Tarskian approaches to argumentation. We also considered a broader range of instantiations of the preference orderings over arguments and showed that these satisfied assumptions required for proof of the aforementioned properties and postulates.

Finally, a key rhetorical claim of this paper is that a proper modelling of the use of preferences requires that we take into account the structure of arguments. We believe this claim to be supported by the results in this paper, and our discussion of recent critiques of Dung and preference-based argumentation frameworks.

We conclude by mentioning future research. Firstly, we are currently developing a structured ASPIC+ approach to extended argumentation [32]. As mentioned in Section 4.3 we are building on a preliminary such structuring in [34]. Early results suggest that the attack definition of conflict free will ensure satisfaction of all key properties and postulates by structured extended argumentation theories, and so address some of the limitations of the extended theory described in [32, 34]. Secondly, [19] recently proposed the additional so called ‘non-interference’ and ‘crash resistance’ rationality postulates, which are about whether self-defeating arguments can interfere with the justification status of other arguments in undesired ways. We plan to study the conditions under which these postulates are satisfied by the ASPIC+ framework.

8 Appendix

8.1 Proofs for Section 4.1

**Proposition 8** Let $A$ and $B$ be arguments where $B$ is plausible or defeasible and $A$ and $B$ have contradictory conclusions, and if $A$ and $B$ are defined as in Def. 6, then $\operatorname{Prem}(A) \cup \operatorname{Prem}(B)$ is $c$-consistent. Then:

1. For all $B' \in M(B)$, there exists a strict extension $A_{B'}^+$ of $(M(B) \setminus \{B'\}) \cup M(A)$ such that $A_{B'}^+$ rebuts or undermines $B$ on $B'$.
2. If $B \prec A$, and $\preceq$ is reasonable, then for some $B' \in M(B)$, $A_{B'}^+$ defeats $B$.

**Proof.** 1) Consider first systems closed under contraposition (Def. 11). Observe first that $\operatorname{Conc}(M(B)) \cup \operatorname{Prem}_m(B) \vdash \operatorname{Conc}(B)$ (i.e., one can construct a strict argument concluding $\operatorname{Conc}(B)$ with all premises taken from $\operatorname{Conc}(M(B))$ and the axiom premises in $B$). By contraposition, and since $\operatorname{Conc}(A)$ and $\operatorname{Conc}(B)$ contradict each other, we have that for any $B_i \in M(B)$: $\operatorname{Conc}(M(B) \setminus \{B_i\}) \cup \operatorname{Prem}_m(B) \cup \operatorname{Conc}(A) \vdash \neg \operatorname{Conc}(B_i)$. Hence, one can construct a strict extension $A_{B_i}^+$ that extends $\{A\} \cup M(B) \setminus \{B_i\} \cup \operatorname{Prem}_m(B)$ with strict rules, and that concludes $\neg \operatorname{Conc}(B_i)$.

By construction, $M(B) \setminus \{B_i\}$ and $M(A)$ are the maximal fallible sub-arguments of $A_{B_i}^+$, and $\operatorname{Prem}(A_{B_i}^+) \subseteq \operatorname{Prem}(A) \cup \operatorname{Prem}(B)$.

Since by construction of $M(B)$ either $B_i$ is a non-axiom premise or ends with a defeasible inference, $A_{B_i}^+$ either undermines or rebuts $B_i$. But then $A_{B_i}^+$ also undermines or rebuts $B$.

For systems closed under transposition the existence of arguments $A_{B_i}^+$ and $B_i$ is proven.
by straightforward generalisation of Lemma 6 in [18]. Then the proof can be completed as above.

In the case that \( A \) and \( B \) are defined as in Def. 6, one only need additionally show that \( \text{Prem}(A_{B_i}^+) \) is c-consistent, which follows given \( \text{Prem}(A_{B_i}^+) \subseteq \text{Prem}(A) \cup \text{Prem}(B) \), and \( \text{Prem}(A) \cup \text{Prem}(B) \) is c-consistent by assumption.

2) By construction, each \( B' \) extension \( A_{B_i}^+ \) of \( A \) is a strict extension of \( \{A\} \cup M(B) \setminus \{B'\} \cup \text{Prem}(B) \). Hence, letting \( M(B) = \bigcup_{i=1}^{n} B_i \), we have \( \{B_1, \ldots, B_n, A\} \) where each \( A_{B_i}^+ \) is a strict extension of \( \{B_1, \ldots, B_{i-1}, B_{i+1}, B_n, A\} \). Also, \( B \) is a strict extension of \( \{B_1, \ldots, B_n\} \). Since \( \preceq \) is reasonable, it cannot be that: \( B \not\prec A \) and \( A_{B_1}^+ \not\prec B_1 \) and \( \ldots \) and \( A_{B_n}^+ \not\prec B_n \). Since by assumption \( B \not\prec A \), then for some \( i \), \( A_{B_i}^+ \) rebuts or undermines \( B \) on \( B_i \), \( A_{B_i}^+ \not\prec B_i \), and so \( A_{B_i}^+ \) defeats \( B \).

\[ \text{QED} \]

In what follows, recall Notation 4, in which \( X \rightarrow Y \) denotes \( X \) attacks \( Y \) and \( X \rightarrow Y \) denotes \( X \) defeats \( Y \).

\textbf{Lemma 33} Let \( (A, C, \preceq) \) be a (c-)SAF:

1. If \( A \) is acceptable w.r.t. \( S \subseteq A \) then \( A \) is acceptable w.r.t. any superset of \( S \).
2. If \( A \rightarrow B \), then \( A \rightarrow B' \) for some \( B' \in \text{Sub}(B) \), and if \( A \rightarrow B' \), \( B' \in \text{Sub}(B) \), then \( A \rightarrow B \).  
3. If \( A \) is acceptable w.r.t. \( S \subseteq A \), \( A' \in \text{Sub}(A) \), then \( A' \) is acceptable w.r.t. \( S \).

\textbf{Proof.} Proofs of 33-1 and 33-2 are straightforward given the definitions of acceptability and defeat. For 33-3, suppose \( B \rightarrow A' \). By 33-2, \( B \rightarrow A \), and so \( \exists C \in S \) s.t. \( C \rightarrow A \). Hence \( A' \) is acceptable w.r.t. \( S \). \[ \text{QED} \]

\textbf{Lemma 34} Suppose \( B \rightarrow A \), where \( B \) attacks \( A \) on \( A' \), and if \( A \) and \( B \) are defined as in Def. 6, then \( \text{Prem}(A) \cup \text{Prem}(B) \) is c-consistent. If not \( B \rightarrow A \) then either:

1. \( A' \rightarrow B \), or;
2. There is a strict extension \( A_{B_i}^+ \) of \( (M(B) \setminus \{B'\}) \cup M(A') \) s.t. \( A_{B_i}^+ \rightarrow B \).

\textbf{Proof.} Since \( B \rightarrow A \), then: \( B \) rebuts on the conclusion \( \varphi \) of \( A' \) where \( A' \)'s top rule is defeasible, or \( B \) undermines the ordinary premise \( A' = \varphi \), and \( B \not\prec A' \). Also, \( \text{Conc}(B) \) must be a contradictory of \( \varphi \) since otherwise \( \text{Conc}(B) \) would be a contrary of \( \varphi \) implying that \( B \rightarrow A \). Also, \( B \) must be plausible or defeasible since for \( B \not\prec A' \) to be the case, \( B \) cannot be strict and firm (under the assumption that \( \preceq \) is reasonable (Def. 16)).
1) If \( B \) is an ordinary or assumption premise or has a defeasible top rule, \( A' \rightarrow B \), and since \( B \not\prec A', A' \rightarrow B \).
2) If \( B \) has a strict top rule, then by Proposition 8 there exists a strict extension \( A_{B_i}^+ \) s.t. \( A_{B_i}^+ \rightarrow B \).  

\[ \text{QED} \]

The following lemma follows from the fact that if \( B \) defeats some strict extension \( A \) of \( \{A_1, \ldots, A_n\} \) then the defeat must be on some \( A_i \).

\textbf{Lemma 35} Let \( (A, C, \preceq) \) be a (c-)SAF. Let \( A \in A \) be a strict extension of \( \{A_1, \ldots, A_n\} \subseteq A \), and for \( i = 1, \ldots, n \), \( A_i \) is acceptable w.r.t. \( E \subseteq A \). Then \( A \) is acceptable w.r.t. \( E \).

\textbf{Proof.} Let \( B \) be any argument s.t. \( B \rightarrow A \). By Def. 7, \( B \) attacks \( A \) by undercutting or rebutting on defeasible rules in \( A \) or undermining on a non-axiom premise in \( A \). Hence, by definition of strict extensions (Def. 15), it must be that \( B \rightarrow A \) if \( B \rightarrow A_i \) for some (possibly more than one) \( A_i \in \{A_1, \ldots, A_n\} \). Either:
1) $B$ undercuts or contrary rebuts/undermines some $A_i$, and so by Def. 8, $B$ defeats $A_i$, or;
2) $B$ does not undercut or contrary rebut/undermine some $A_i$. Suppose for all $A_i$, for all sub-arguments $A_i'$ of $A_i$ s.t. $B$ rebuts or undermines $A_i$ on $A_i'$, $B \not\rightarrow A_i'$. This contradicts $B$ defeats $A_i$. Hence, for some $A_i$, $B$ defeats $A_i$.

We have shown that if $B$ defeats $A_i$ then $B$ defeats some $A_i$. By assumption of $A_i$ acceptable w.r.t. $E$, $\exists C \in E$ s.t. $C$ defeats $B$. Hence, $A$ is acceptable w.r.t. $E$. QED

For the following proposition, recall that by assumption, any $c$-SAF is well defined and so satisfies c-classicality (Def. 11).

Proposition 9 Let $(A, C, \preceq)$ be a $c$-SAF. If $A_1, \ldots, A_n$ are acceptable w.r.t. some conflict-free $E \subseteq A$, then $\bigcup_{i=1}^n \text{Prem}(A_i)$ is $c$-consistent.

Proof. Suppose for contradiction otherwise, and let $S$ be any minimally c-inconsistent subset of $\bigcup_{i=1}^n \text{Prem}(A_i)$. By assumption of c-classicality:

for all $\varphi \in S$, $S \setminus \{\varphi\} \vdash \varphi$ and $S \setminus \{\varphi\}$ is c-consistent.

We thus have the set of non-axiom premises $\{\varphi_1, \ldots, \varphi_m\} \subseteq \bigcup_{i=1}^n \text{Prem}(A_i)$ (that must be non-empty given that $K_n$ is c-consistent by assumption of axiom consistency (Def. 11)), such that for $i = 1 \ldots m$, there is a strict extension $B^{+i}$ of $\{\varphi_1, \ldots, \varphi_{i-1}, \varphi_{i+1}, \varphi_m\}$ s.t. $B^{+i} \not\vdash \varphi_i$. Since $\preceq$ is reasonable, for some $i, B^{+i} \not\preceq \varphi_i$ and so $B^{+i} \rightarrow \varphi_i$. Since for $i = 1 \ldots n$, $A_i$ is acceptable w.r.t. $E$, then: since $\varphi_i \in \bigcup_{i=1}^n \text{Prem}(A_i)$, then by Lemma 33-3, $\varphi_i$ is acceptable w.r.t. $E$.

Since $B^{+i}$ is a strict extension of some subset of $\bigcup_{i=1}^n \text{Prem}(A_i)$, then by Lemmas 33-3 and 35, $B^{+i}$ is acceptable w.r.t. $E$. But then since $B^{+i} \rightarrow \varphi_i$, $\exists X, Y \in E$ s.t. $Y \rightarrow B^{+i}$, contradicting $E$ is conflict free. QED

Lemma 36 Let $A$ be acceptable w.r.t an admissible extension $S$ of a $(c)$-SAF $(A, C, \preceq)$. Then $\forall B \in S \cup \{A\}$, neither $A \rightarrow B$ or $B \rightarrow A$.

Proof. Suppose for contradiction that: 1) $A \rightarrow B$, $B \in S \cup \{A\}$. By assumption of $B$’s acceptability, $\exists C \in S$ s.t. $C \rightarrow A$, and by acceptability of $A$, $\exists D \in S$ s.t. $D \rightarrow C$, hence $D \rightarrow C$, contradicting $S$ is conflict free; 2) $B \rightarrow A$, $B \in S$. By acceptability of $A$, $\exists D \in S$ s.t. $D \rightarrow B$, hence $D \rightarrow B$, contradicting $S$ is conflict free. QED

Proposition 10 Let $A$ be acceptable w.r.t an admissible extension $S$ of a $(c)$-SAF $(A, C, \preceq)$. Then $S' = S \cup \{A\}$ is conflict free.

Proof. Firstly, since for any $B \in S$, $B$ is acceptable w.r.t. $S$, then by Lemma 9, $\text{Prem}(A) \cup \text{Prem}(B)$ is c-consistent.

Suppose for contradiction that $S'$ is not conflict free. By assumption, $S$ is conflict free. $A$ cannot attack itself since $A$ must then defeat itself, contradicting Lemma 36. Hence, we have the following two cases:

1) $\exists B \in S$, $B \rightarrow A$, and $B \rightarrow A$ by Lemma 36. By Lemma 34, for some sub-argument $A'$ of $A$, either:

1.1) $A'$ defeats $B$, hence (by acceptability of $B$) $\exists C \in S$ s.t. $C \rightarrow A'$, and so (by Lemma 33-2) $C \rightarrow A$, contradicting Lemma 36, or; 1.2) $\exists A''$, s.t. $A'' \rightarrow B$, hence $\exists C \in S$ s.t. $C \rightarrow A''$. By construction of $A''$ and Lemma 33-2, it must be that $C \rightarrow Z$, $Z \in \text{Sub}(A) \cup \text{Sub}(B)$. Hence, (by Lemma 33-2) either $C \rightarrow B$, contradicting $S$ is conflict free, or $C \rightarrow A$, contradicting Lemma 36.

2) $\exists B \in S$, $A \rightarrow B$, and $A \rightarrow B$ by Lemma 36. By Lemma 34, for some sub-argument $B'$ of $B$, either:

35
2.1) \( B' \) defeats \( A \), hence (by acceptability of \( A \)) \( \exists C \in S \) s.t. \( C \rightarrow B' \) and so (by Lemma 33-2) \( C \rightarrow B \), hence \( C \rightarrow B \), contradicting \( S \) is conflict free, or \( \text{2.2) } \exists B'_A^+ \text{ s.t. } B'_A^+ \rightarrow A, \) hence \( \exists C \in S \) s.t. \( C \rightarrow B'_A^+ \). By construction of \( B'_A^+ \), \( C \rightarrow Z, Z \in \text{Sub}(A) \cup \text{Sub}(B) \), leading to a contradiction as in 1.2).

**Proposition 11** Let \( A, A' \) be acceptable w.r.t an admissible extension \( S \) of a \( (c-)SAF \) \( (A, C, \leq) \). Then:
1. \( S' = S \cup \{ A \} \) is admissible
2. \( A' \) is acceptable w.r.t. \( S' \).

**Proof.** 1) By Lemma 33-1, all arguments in \( S' \) are acceptable w.r.t. \( S' \). By Proposition 10, \( S' \) is conflict free. Hence \( S' \) is admissible. 2) By Lemma 33-1, \( A' \) is acceptable w.r.t. \( S' \).

**QED**

### 8.2 Proofs for Section 4.2

**Theorem 12** [Sub-argument Closure] Let \( \Delta = (A, C, \leq) \) be a \( (c-)SAF \) and \( E \) a complete extension of \( \Delta \). Then for all \( A \in E \): if \( A' \in \text{Sub}(A) \) then \( A' \in E \).

**Proof.** \( A' \) is acceptable w.r.t. \( E \) by Lemma 33-3. \( E \cup \{ A' \} \) is conflict free by Prop.10. Hence, since \( E \) is complete, \( A' \in E \).

**QED**

**Theorem 13** [Closure under Strict Rules] Let \( \Delta = (A, C, \leq) \) be a \( (c-)SAF \) and \( E \) a complete extension of \( \Delta \). Then \( \{ \text{Conc}(A) | A \in E \} = \text{Cl}_{IR}(\{ \text{Conc}(A) | A \in E \}) \).

**Proof.** It suffices to show that any strict extension \( X \) of \( \{ A | A \in E \} \) is in \( E \). By Lemma 35, any such \( X \) is acceptable w.r.t. \( E \). By Proposition 10, \( E \cup \{ X \} \) is conflict free. Hence, since \( E \) is complete, \( X \in E \). Note that if \( \Delta \) is a \( cSAF \), Proposition 9 guarantees that \( X \)'s premises are \( c \)-consistent.

**QED**

**Theorem 14** [Direct Consistency] Let \( \Delta = (A, C, \leq) \) be a \( (c-)SAF \) and \( E \) an admissible extension of \( \Delta \). Then \( \{ \text{Conc}(A) | A \in E \} \) is consistent.

**Proof.** We show that if \( A, B \in E, \text{Conc}(A) \in \text{Conc}(B) \) (i.e., \( E \) is inconsistent (Def. 2)), then this leads to a contradiction:
1. \( A \) is firm and strict, and:
   1.1 if \( B \) is strict and firm, then this contradicts the assumption of axiom consistency (Def. 11); 1.2 if \( B \) is plausible or defeasible, and 1.2.1 \( B \) is an ordinary/assumption premise or has a defeasible top rule, then \( A \rightarrow B \), contradicting \( E \) is conflict free, or 1.2.2 \( B \) has a strict top rule (see 3 below).

2. \( A \) is plausible or defeasible, and:
   2.1 if \( B \) is strict and firm then under the well-formed assumption (Def. 11) \( \text{Conc}(A) \) cannot be a contrary of \( \text{Conc}(B) \), and so they are contradictory (i.e., a contrary of each other), and 2.1.1 \( A \) is an ordinary/assumption premise or has a defeasible top rule, in which case \( B \rightarrow A \), contradicting \( E \) is conflict free, or 2.1.2 \( A \) has a strict top rule (see 3 below);
   2.2 if \( B \) is plausible or defeasible and 2.2.1 \( B \) is an ordinary/assumption premise or has a defeasible top rule then \( A \rightarrow B \), contradicting \( E \) is conflict free, or 2.2.2 \( B \) has a strict top rule (see 3 below).

3. Each of 1.2.2, 2.1.2 and 2.2.2 describes the case where \( X, Y \in E, \text{Conc}(X) \in \text{Conc}(Y), \) \( Y \) is defeasible or plausible and has a strict top rule, and so by the well-formed assumption \( \text{Conc}(X) \) and \( \text{Conc}(Y) \) must be contradictory.

In the case that \( \Delta \) is a \( cSAF \), since \( X, Y \in E \), then \( X, Y \) are acceptable w.r.t. \( E \), and so by Proposition 9, \( \text{Prem}(A) \cup \text{Prem}(B) \) is \( c \)-consistent.

By Prop 8 there is a strict extension \( X^+ \), of \( M(Y) \cup \{ Y \} \cup M(A) \) s.t. \( X^+ \rightarrow Y \). By Lemma 35 \( X^+ \) is acceptable w.r.t. \( E \), and by Prop. 10, \( E \cup \{ X^+ \} \) is conflict free, contradicting \( X^+ \rightarrow Y \).

**QED**
Theorem 15 [Indirect Consistency] Let $\Delta = (A, C, \preceq)$ be a (c-)SAF and $E$ a complete extension of $\Delta$. Then $Cl_\Delta (\{ \text{Conc}(A)| A \in E \})$ is consistent.

**Proof.** Follows from Theorems 13 and 14. QED

8.3 Proofs for Section 4.3

Proposition 16 Let $\Delta$ be a (c-)SAF. For $T \in \{ \text{admissible, complete, grounded, preferred, stable} \}$, $E$ is an a-T extension of $\Delta$ if and only if $E$ is a d-T-admissible extension of $\Delta$.

**Proof.** Since the acceptability of arguments is defined w.r.t. to the defeat relation in both cases, and in both cases arguments outside of a stable extension are required to be defeated by an argument in the stable extension, it suffices to show that $E$ is conflict free under the attack definition if $E$ is conflict free under the defeat definition.

The left to right half is trivial: If no two arguments in $E$ attack each other, then no two arguments in $E$ defeat each other. For the right to left half, suppose $B, A \in E$, $B \rightarrow A$. First note that since $A, B$ are acceptable w.r.t. $E$, then in the case of a c-SAF where $A$ and $B$ are defined as in Def 6, $\text{Prem}(A) \cup \text{Prem}(B)$ is c-consistent by Proposition 9. Then, by Lemma 34, $\exists A' \in \text{Sub}(A)$ s.t. i) $A' \rightarrow B$, or ii) there is a strict extension $A''_{B}$ of $(M(B) \setminus B') \cup M(A')$ s.t. $A''_{B} \rightarrow B$. In case i), by acceptability of $B$ $\exists C \in E$ s.t. $C \rightarrow A'$, and so (by Lemma 33-2) $C \rightarrow A$, contradicting $E$ is conflict free. In case ii), (by acceptability of $B$), $\exists C \in E$ s.t. $C \rightarrow A''_{B}$. By construction of $A''_{B}$ and Lemma 33-2, $C \rightarrow Z$, $Z \in \text{Sub}(A) \cup \text{Sub}(B)$. Hence, (by Lemma 33-2) either $C \rightarrow B$ or $C \rightarrow A$, contradicting $E$ is defeat conflict free. QED

8.4 Proofs for Section 5.1

Proposition 18 For $s \in \{ \text{Elitist1, Elitist2, Democratic1, Democratic2} \}$, $\prec_s$ is a strict partial order.

**Proof.** *Irreflexivity.* It is straightforward to see that $\Gamma' \not\preceq_{\text{Elitist1(Dem2)}} \Gamma'$ given that it cannot be the case that $\emptyset \prec_{\text{Elitist1(Dem2)}} \emptyset$. Since $\prec_s$ is defined over finite sets, we cannot have an infinite $\Gamma' = \{ l_1, l_2, \ldots \}$, where for $i = 1, 2 \ldots, l_i < l_{i+1}$ which would be a counter-example for $\prec_{\text{Dem1}}$. It is then straightforward to see that for any finite $\Gamma'$ if $\Gamma' \prec_{\text{Elitist1(Dem2)}} \Gamma''$, then $\exists \ell', \ell'' \in \Gamma'$ s.t. $\ell' < \ell'' < \ell''$. Contradiction.

*Transitivity:* Suppose $\Gamma \prec_s \Gamma'$ and $\Gamma' \prec_s \Gamma''$.

Let $s = \text{Elitist1(Elitist2)}$. Then by definition of $\prec_s$, $\Gamma (\Gamma \setminus \Gamma')$, $\Gamma' (\Gamma' \setminus \Gamma'')$ must be non-empty. If $\Gamma'' (\Gamma'' \setminus \Gamma'') = \emptyset$, then by definition of $\prec_s$, $\Gamma \prec_s \Gamma''$. Else let $\Gamma'' (\Gamma'' \setminus \Gamma') = \{ l''_1, \ldots, l''_n \}$. By definition of $\prec_s: \exists \ell' \in \Gamma'' (\Gamma'' \setminus \Gamma'')$, for $j = 1 \ldots, n, l' < l''_j$; $\exists l \in \Gamma (\Gamma \setminus \Gamma'$), $l < l'$. Hence, by transitivity of $\prec$, for $j = 1 \ldots n, l < l''_j$, i.e., $\Gamma \prec_s \Gamma''$.

Let $s = \text{Democratic1(Democratic2)}$. Suppose $\Gamma (\Gamma \setminus \Gamma'') = \emptyset$. Then by definition of $\prec_s$, $\Gamma \prec_s \Gamma''$. Suppose $\Gamma (\Gamma \setminus \Gamma'') = \emptyset$. By definition of $\prec_s$, it cannot be that $\Gamma'' (\Gamma'' \setminus \Gamma') = \emptyset$, and: $\forall l \in \Gamma (\Gamma \setminus \Gamma')$, $\exists \ell'' \in \Gamma'' (\Gamma'' \setminus \Gamma'), l < l''$; $\forall l' \in \Gamma' (\Gamma'' \setminus \Gamma''), \exists l''' \in \Gamma'' (\Gamma'' \setminus \Gamma'')$, $l < l'''$, Hence by transitivity of $\prec$, $\forall l \in \Gamma (\Gamma \setminus \Gamma'')$, $\exists l'' \in \Gamma'' (\Gamma'' \setminus \Gamma')$ s.t. $l < l''$. Hence $\Gamma \prec_s \Gamma''$.

QED

Lemma 37 $\prec$ is a strict partial order.

**Proof.** *Irreflexivity:* Suppose $\Gamma \prec \Gamma$. Then by Def. 18, $\Gamma \not\prec \emptyset$, and so $\Gamma \prec_s \Gamma$, contradicting $\prec_s$ is a strict partial ordering. *Transitivity:* Suppose $\Gamma \prec \Gamma'$ and $\Gamma' \prec \Gamma''$. Then by Def. 18, $\Gamma, \Gamma'$ must be non-empty and so $\Gamma \prec_s \Gamma'$. If $\Gamma'' = \emptyset$, then by definition $\Gamma \prec \Gamma''$. Else if $\Gamma'' \not\prec \emptyset$, then $\Gamma' \prec_s \Gamma''$. Hence by transitivity of $\prec_s, \Gamma \prec_s \Gamma''$. Hence $\Gamma \prec \Gamma''$ QED
In the following proofs, we may write \( \text{LDR} \) as an abbreviation for LastDefRules, and \( \text{DR} \) as an abbreviation for DefRules.

**Proposition 20** Let \( \preceq \) be defined according to the last link principle. Then \( \prec \) is a strict partial order.

**Proof.** *Irreflexivity:* If \( A \prec A \), then either \( \text{LDR}(A) \preceq \text{LDR}(A) \) or \( \text{Prem}_\text{ap}(A) \preceq \text{Prem}_\text{ap}(A) \), contradicting the irreflexivity of \( \preceq \) (Lemma 37).

*Transitivity:* Suppose \( C \prec B \) and \( B \prec A \). Consider the following two cases:

1) Suppose \( \text{LDR}(A) \neq \emptyset \). Then \( B \prec A \) by condition 1 of Def. 19; i.e., \( \text{LDR}(B) \preceq \text{LDR}(A) \), and so it must be that \( \text{LDR}(B) \neq \emptyset \). By the same reasoning, \( \text{LDR}(C) \preceq \text{LDR}(B) \), and \( \text{LDR}(C) \neq \emptyset \).

2) Suppose \( \text{LDR}(A) = \emptyset \).
   - If \( \text{LDR}(C) \neq \emptyset \), then \( \text{LDR}(C) \preceq \text{LDR}(A) \) by definition of \( \preceq \), and so \( C \prec A \) by condition 1 of Def. 19.
   - If \( \text{LDR}(C) = \emptyset \) then it must be that \( \text{LDR}(B) = \emptyset \) (if \( \text{LDR}(B) \neq \emptyset \) one could not conclude \( C \prec B \) by condition 1 or 2 of Def. 19). Hence \( B \prec A \) and \( C \prec B \) by condition 2 of Def. 19. That is to say, \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(B) \) and \( \text{Prem}_\text{ap}(B) \preceq \text{Prem}_\text{ap}(A) \). By transitivity of \( \prec \), \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(A) \), and so \( C \prec A \).

QED

**Proposition 21** Let \( \preceq \) be defined according to the weakest link principle. Then \( \prec \) is a strict partial order.

**Proof.** *Irreflexivity:* If \( A \prec A \), then either \( \text{LDR}(A) \preceq \text{LDR}(A) \) and/or \( \text{Prem}_\text{ap}(A) \preceq \text{Prem}_\text{ap}(A) \), contradicting the irreflexivity of \( \preceq \).

*Transitivity:* Suppose \( C \prec B \) and \( B \prec A \). Consider the following three cases:

1) \( C \) and \( B \) are strict. Since \( B \) is strict then \( A \) must be strict, since otherwise, given \( B \prec A \), it must be that \( \text{DR}(B) \preceq \text{DR}(A) \), which cannot hold by virtue of \( \text{DR}(B) = \emptyset \). Hence \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(B) \), \( \text{Prem}_\text{ap}(B) \preceq \text{Prem}_\text{ap}(A) \), and so by transitivity of \( \prec \), \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(A) \). Hence \( C \prec A \) by condition 1 of Def. 20.

2) \( C \) and \( B \) are firm. Since \( B \) is firm then \( A \) must be firm, since otherwise, given \( B \prec A \), it must be that \( \text{Prem}_\text{ap}(B) \preceq \text{Prem}_\text{ap}(A) \), which cannot hold by virtue of \( \text{Prem}_\text{ap}(B) = \emptyset \). Hence \( \text{DR}(C) \preceq \text{DR}(B) \), \( \text{DR}(B) \preceq \text{DR}(A) \), and so by transitivity of \( \prec \), \( \text{DR}(C) \preceq \text{DR}(A) \). Hence \( C \prec A \) by condition 2 of Def. 20.

3) \( C \) and \( B \) fall into neither of the above two cases. Then \( \text{DR}(C) \preceq \text{DR}(B) \) and \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(B) \). Hence, by definition of \( \prec \), it must be that \( \text{DR}(C) \neq \emptyset \), \( \text{Prem}_\text{ap}(C) \neq \emptyset \).
   - If \( \text{DR}(A) = \emptyset \), then by definition of \( \prec \), \( \text{DR}(C) \preceq \text{DR}(A) \). If \( \text{DR}(A) \neq \emptyset \), then given \( B \prec A \), \( \text{DR}(B) \preceq \text{DR}(A) \), and so by transitivity of \( \prec \), \( \text{DR}(C) \preceq \text{DR}(A) \).
   - If \( \text{Prem}_\text{ap}(A) = \emptyset \), then by definition of \( \prec \), \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(A) \). If \( \text{Prem}_\text{ap}(A) \neq \emptyset \), then given \( B \prec A \), \( \text{Prem}_\text{ap}(B) \preceq \text{Prem}_\text{ap}(A) \), and so by transitivity of \( \prec \), \( \text{Prem}_\text{ap}(C) \preceq \text{Prem}_\text{ap}(A) \). Hence \( C \prec A \) by condition 3 of Def. 20.

QED

**Proposition 22** Let \( \preceq \) be defined according to the last link principle. Then \( \preceq \) is reasonable.

**Proof.** *Proof of the first condition of reasonableness:*

1-i) Assume \( A \) is strict and firm, and so \( \text{LDR}(A) = \emptyset \) and \( \text{Prem}_\text{ap}(A) = \emptyset \). If \( \text{LDR}(B) \neq \emptyset \), then by definition of \( \prec \), \( \text{LDR}(B) \preceq \text{LDR}(A) \), and so \( B \prec A \) by the first condition of Def. 19. If \( \text{LDR}(B) = \emptyset \), then by assumption of \( B \) being plausible or defeasible, \( \text{Prem}_\text{ap}(B) = \emptyset \). By definition of \( \prec \), \( \text{Prem}_\text{ap}(B) \preceq \text{Prem}_\text{ap}(A) \), and so \( B \prec A \) by the second condition of Def. 19.

1-ii) Assume \( B \) is strict and firm, and so \( \text{LDR}(B) = \emptyset \), \( \text{Prem}_\text{ap}(B) = \emptyset \). Then by definition of
< it cannot be that $LDR(B) < LDR(A)$ or $Premap(B) < Premap(A)$, and so it cannot be that $B < A$, by the first or second conditions of Def. 19.

1-iii) Follows straightforwardly from Def. 19, given that $A'$ differs from $A$ only in it’s strict rules and/or axiom premises.

Proof of the second condition of reasonableness:
Suppose for contradiction that:

\[ \forall i, \text{there is a strict extension } C^{+i} \text{ of } \{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\} \text{ such that } C^{+i} < C_i \]

1) Suppose for some $i = 1 \ldots n$, $LDR(C_i) \neq \emptyset$. Then, it must be that $C^{+i} < C_i$ based on condition 1 of Def. 19, i.e., $\bigcup_{i=1,j \neq i}^{n} LDR(C_j) < LDR(C_i)$.

If $<$ is defined on the basis of $\preceq_{\text{EL11}(2)}$, there must be some $l \in \bigcup_{i=1,j \neq i}^{n} LDR(C_j)$, and so for some $j = 1 \ldots n, j \neq i$, $l \in \bigcup_{j=1}^{n} LDR(C_j)$, s.t. $l < l' \in LDR(C_i)$.

If $<$ is defined on the basis of $\preceq_{\text{Dem1}(2)}$, $\forall l \in \bigcup_{i=1,j \neq i}^{n} LDR(C_j)$, and so for some $j = 1 \ldots n, j \neq i$, $\forall l \in LDR(C_j)$, s.t. $\exists l' \in LDR(C_i), l < l'$.

Hence in both cases, for some $j = 1 \ldots n, j \neq i$, $LDR(C_j) < LDR(C_i)$. By definition of $<$, it must be that $LDR(C_i) \neq \emptyset$. Then, since $C^{+i} < C_j$, one can reason in the same way, concluding that for some $k = 1 \ldots n, k \neq j$, $LDR(C_k) < LDR(C_j)$ and $LDR(C_k) \neq \emptyset$.

Since $\{C_1, \ldots, C_n\}$ is finite, we can continue reasoning in the same way, obtaining that for some $\{C_1, \ldots, C_{n+m}\} \subseteq \{C_1, \ldots, C_n\}$, $LDR(C_i) < \ldots < LDR(C_{i+m}) < LDR(C_i)$, and so $C_i \lland \ldots \lland C_{i+m} < C_i$, contradicting $<$ is a strict partial order (Proposition 20).

2) Suppose for $i = 1 \ldots n$, $LDR(C_i) = \emptyset$. Then, given i), and given that $<$ is defined based on $\preceq_{\text{EL11}(2)}$ or $\preceq_{\text{Dem1}(2)}$: $\forall i, \exists j, Premap(C_j) < Premap(C_i)$ and so $C_j < C_i$. Since $\{C_1, \ldots, C_n\}$ is finite, this contradicts $<$ is a strict partial order.

QED

Proposition 23 Let $\preceq$ be defined according to the weakest link principle. Then $\preceq$ is reasonable.

Proof. Proof of the first condition of reasonableness:
1-i) Assume $A$ is strict and firm, and so $LDR(A) = \emptyset$ and $Premap(A) = \emptyset$. Suppose $B$ is strict. Then by assumption, $B$ is plausible, i.e., $Premap(B) \neq \emptyset$. Hence $Premap(B) < Premap(A)$ and $B < A$ by Def. 20-1). Suppose $B$ is firm. Then by assumption $B$ is defeasible, i.e., $DR(B) \neq \emptyset$. Hence $DR(B) < DR(A)$ and $B < A$ by Def. 20-2). Suppose $B$ is defeasible and plausible. Then $Premap(B) < Premap(A)$, $DR(B) < DR(A)$, and $B < A$ by Def. 20-3).

1-ii) Assume $B$ is strict and firm, and so $LDR(B) = \emptyset$, $Premap(B) = \emptyset$. Then by definition of $<$ it cannot be that $LDR(B) < LDR(A)$ or $Premap(B) < Premap(A)$, and so it cannot be that $B < A$, by the first, second or third conditions of Def. 20.

1-iii) Follows straightforwardly from Def. 20, given that $A'$ differs from $A$ only in it’s strict rules and/or axiom premises.

Proof of the second condition of reasonableness for the weakest link principle:
Suppose for contradiction that:

\[ \forall i, \text{there is a strict extension } C^{+i} \text{ of } \{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\} \text{ such that } C^{+i} < C_i \]

1) Suppose for some $i = 1 \ldots n$, $DR(C_i) \neq \emptyset$. Then, it must be that $C^{+i} < C_i$ by Def. 20-2) or 20-3). Reasoning in the same way as in the proof of Proposition 22, for some $j = 1 \ldots n, j \neq i$, $DR(C_j) < DR(C_i)$. By definition of $<$, it must be that $DR(C_j) \neq \emptyset$. Then, since $C^{+i} < C_j$, one can reason in the same way, concluding that for some $k = 1 \ldots n,$
$k \neq j$, $\text{DR}(C_k) \triangleleft \text{DR}(C_j)$ and $\text{DR}(C_k) \neq \emptyset$. Since $\{C_1, \ldots, C_n\}$ is finite, we can continue reasoning in the same way, obtaining that for some $\Pi = \{C_1, \ldots, C_{i+m}\} \subseteq \{C_1, \ldots, C_n\}$, $\text{DR}(C_i) \triangleleft \ldots \triangleleft \text{DR}(C_{i+m}) \triangleleft \text{DR}(C_j)$.

2) Suppose for some $i = 1 \ldots n$, $\text{Prem}_\text{ap}(C_i) \neq \emptyset$. Then, it must be that $C_1^{\bot^i} \prec C_i$ by Def. 20-(1) or 20-(3). And so, for some $j = 1 \ldots n$, $j \neq i$, $\text{Prem}_\text{ap}(C_j) \prec \text{Prem}_\text{ap}(C_i)$. By definition of $\prec$, it must be that $\text{Prem}_\text{ap}(C_j) \neq \emptyset$. Then, since $C_1^{\bot^j} \prec C_j$, one can reason in the same way, concluding that for some $k = 1 \ldots n$, $k \neq j$, $\text{Prem}_\text{ap}(C_k) \prec \text{Prem}_\text{ap}(C_j)$ and $\text{Prem}_\text{ap}(C_k) \neq \emptyset$. Since $\{C_1, \ldots, C_n\}$ is finite, we can continue reasoning in the same way, obtaining that for some $\Pi' = \{C_i', \ldots, C_{i+m'}\} \subseteq \{C_1, \ldots, C_n\}$, $\text{Prem}_\text{ap}(C_i') \prec \ldots \prec \text{Prem}_\text{ap}(C_{i+m'}) \prec \text{Prem}_\text{ap}(C_j)$.

Suppose $\Pi \cap \Pi' = \emptyset$ and $\Pi \neq \emptyset$. In other words $\Pi$ consists of defeasible firm arguments (if $\Pi$ contains a plausible argument with a non-empty set of non-axiom premises, then by the reasoning in case 2, such an argument must also be in $\Pi'$). Hence, by condition 2 of Def. 20, $C_i \prec \ldots \prec C_{i+m} \prec C_i$, contradicting $\prec$ is a strict partial order (Proposition 21).

Suppose $\Pi \cap \Pi' = \emptyset$ and $\Pi' \neq \emptyset$. In other words $\Pi'$ consists of strict plausible arguments (if $\Pi'$ contains a defeasible argument with a non-empty set of defeasible rules, then by the reasoning in case 1, such an argument must also be in $\Pi$). Hence, by condition 1 of Def. 20, $C_i \prec \ldots \prec C_{i+m} \prec C_i$, contradicting $\prec$ is a strict partial order.

$\Pi \cap \Pi' \neq \emptyset$. Let $C_i \in \Pi \cap \Pi'$. By transitivity of $\prec$, $\text{DR}(C_i) \triangleleft \text{DR}(C_i')$ and $\text{Prem}_\text{ap}(C_i) \prec \text{Prem}_\text{ap}(C_i')$. Hence, by condition 3 of Def. 20, $C_i \prec C_i'$, contradicting $\prec$ is a strict partial order.

QED

8.5 Proofs for Section 5.2

Lemma 38 Let $(AS, KB)$ be the abstract logic argumentation theory based on $(L', C_\text{n})$ and $(\Sigma, \subseteq')$. Then for all $X \subseteq \Sigma$ it holds that $p \in \text{Cn}(X)$ iff $X \vdash p$.

Proof. From left to right, suppose $p \in \text{Cn}(X)$ for some subset-minimal $X \subseteq \Sigma$. By Def. 21-(3), $X$ is finite, so $X \rightarrow p \in R_s$, so $X \vdash p$ (recall that $X \vdash p$ denotes a strict ASPIC+ argument for $p$ based on premises $X' \subseteq X$).

From right to left is proven by induction on the structure of arguments. Assume $A = X \vdash p$, where by Def. 23, $X \subseteq K_p$, $X \subseteq \Sigma$. Assume first $p \in X$. Then $p \in \text{Cn}(X)$ by Def. 21-(1). Consider next any $A = A_1, \ldots, A_n \rightarrow \varphi$. By inductive hypothesis, $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \in \text{Cn}(X)$. Since $\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \rightarrow \varphi \in R_s$, then $\varphi \in \text{Cn}(\bigcup_{i=1}^{n} \text{Conc}(A_i))$. Since $\bigcup_{i=1}^{n} \text{Conc}(A_i) \subseteq \text{Cn}(X)$, then by monotonicity, $\varphi \in \text{Cn}(\text{Conc}(X))$. But then by Def. 21-(2), $\varphi \in \text{Cn}(X)$.

QED

Lemma 39 Let $(L, C_\text{n})$ be an abstract logic. For any finite $S \subseteq L$ and any $p \in L$, if $S \cup \{p\}$ is AL inconsistent, then there exists an $s \in \text{Cn}(S)$ such that $\{s, p\}$ is AL-inconsistent.

Proof. Since $S$ is finite, we can with repeated application of the definition of adjunction conclude that there exists an $s$ such that $\text{Cn}(\{s\}) = \text{Cn}(S)$. By Def. 21-(1), $s \in \text{Cn}(\{s\})$ and so $s \in \text{Cn}(S)$. By Def. 21-(2), $\text{Cn}(S \cup \{p\}) = \text{Cn}(\{s\} \cup \{p\})$. Then $\text{Cn}(\{s\} \cup \{p\}) = L$ so $\{s, p\}$ is AL-inconsistent.

QED

Lemma 40 If $X \cup Z$ is AL-inconsistent and $X \subseteq \text{Cn}(Y)$ then $Y \cup Z$ is AL-inconsistent.

Proof. We prove the contraposition that if $Y \cup Z$ is AL-consistent and $X \subseteq \text{Cn}(Y)$, then $X \cup Z$ is AL-consistent. This prove this, we first show that:

If $X \subseteq \text{Cn}(Y)$, then $\text{Cn}(X \cup Y) = \text{Cn}(Y)$  \hspace{1cm} (1)

By monotonicity, $\text{Cn}(Y) \subseteq \text{Cn}(X \cup Y)$. We prove that $\text{Cn}(X \cup Y) \subseteq \text{Cn}(Y)$. Let $X \subseteq \text{Cn}(Y)$. By Def.21-(1), $Y \subseteq \text{Cn}(Y)$. But then $X \cup Y \subseteq \text{Cn}(Y)$. Then by monotonicity $\text{Cn}(X \cup Y) \subseteq \text{Cn}(\text{Cn}(Y))$. Since $\text{Cn}(\text{Cn}(Y)) = \text{Cn}(Y)$ (by Def.21-(2)), then
$Cn(X \cup Y) \subseteq Cn(Y)$. So $Cn(X \cup Y) = Cn(Y)$. Given (1), then by property (7) in Section 5.2, $Cn(X \cup Y \cup Z) = Cn(Y \cup Z)$. Now if $Cn(Y \cup Z) \neq \mathcal{L}$ then $Cn(X \cup Y \cup Z) \neq \mathcal{L}$. But then by monotonicity $Cn(X \cup Z) \neq \mathcal{L}$.

**Proposition 24** A c-SAF based on an AL argumentation theory is closed under contraposition, axiom consistent, c-classical, and well-formed.

**Proof.** Well-formedness immediately follows from the fact that the contrariness relation is symmetric. Axiom consistency follows from the fact that $K_n = \emptyset$. To prove satisfaction of contraposition we must prove:

If $S \vdash p$ then for all $s \in S$ it holds that $S \setminus \{ s \} \cup \{ \neg p \} \vdash \neg s$ for any $\neg p$ and $\neg s$.

By definition of $\vdash$, if $S \vdash p$ then $s' \vdash p$ for some finite $s' \subseteq S$. Therefore we can without loss of generality assume that $S$ is finite. Next, if $S \vdash p$ then $p \in Cn(S)$ by Lemma 38. Consider any $\neg p$. Then $\{ p, \neg p \}$ is AL-inconsistent by Def. 23(3b). But then $S \cup \{ \neg p \}$ is AL-inconsistent by Lemma 40. Then by simple rewriting for all $s \in S$ it holds that $(S \setminus \{ s \}) \cup \{ \neg p \} \cup \{ s \}$ is AL-inconsistent. By assumption that $S$ is finite, Lemma 39 then yields that there exists an $s'' \in Cn(S \setminus \{ s \} \cup \{ \neg p \})$ such that $\{ s'', s \}$ is AL-inconsistent. Then by Def. 23(3c) there exists an $s''' \in Cn(s'')$ such that $s''' = \neg s$. By Def. 21-(2) and monotonicity, $s''' \in Cn(S \setminus \{ s \} \cup \{ \neg p \})$. Hence, $\neg s \in Cn(S \setminus \{ s \} \cup \{ \neg p \})$. Hence, $S \setminus \{ s \} \cup \{ \neg p \} \vdash \neg s$ by Lemma 38.

C-classicality is proven as follows. We first prove that if $S \subseteq \mathcal{L}$ is AL-inconsistent, then some finite $S' \subseteq S$ is AL-inconsistent. By Def. 21-(4), $\exists p \in \mathcal{L}$ such that $Cn(p) = \mathcal{L}$. Let $p$ be denoted by $\bot$. Now suppose $S$ is inconsistent. Then $Cn(S) = \mathcal{L}$, so $\bot \in Cn(S)$. By Def. 21-(3), $\bot \in Cn(S')$ for some finite $S' \subseteq S$. But since $Cn(Cn(S')) = Cn(S')$ (Def. 21-(2)), this implies that $Cn(S') = \mathcal{L}$.

Now assume $S \subseteq \mathcal{L}$ is minimally c-inconsistent. Then for some $p$ it holds that $S \vdash p$, $\neg p$. By Lemma 38, $\{ p, \neg p \} \subseteq Cn(S)$. By Def. 23(3b), $\{ p, \neg p \}$ is AL-inconsistent, and so by monotonicity $Cn(S)$ is AL-inconsistent. But then by Def. 21-(2), $S$ is AL-inconsistent, and so some finite $S' \subseteq S$ is AL-inconsistent. But since $S$ is minimally inconsistent, it holds that $S' = S$. Consider any $s \in S$. Then $S \setminus \{ s \} \cup \{ s \}$ is inconsistent. By adjunction and finiteness of $S$ there exists a formula $x \in \mathcal{L}$ that has exactly the same consequences as $S \setminus \{ s \}$. Then by property (7) in Section 5.2, $\{ x, s \}$ is AL-inconsistent, and so by Def. 23(3c) there exists a $y \in Cn(\{ x \})$ such that $y = \neg s$. But then $S \setminus \{ s \} \vdash \neg s$.

**Lemma 41** For any $(AS, KB)$ based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leq')$, it holds that $\{ p, q \}$ is AL-inconsistent iff $p \vdash \neg q$ for some $\neg q$.

**Proof.** Suppose $\{ p, q \}$ is AL-inconsistent. By Def. 23(3) $\neg q \in Cn(\{ p \})$ for some $\neg q$. By Lemma 38, $p \vdash \neg q$. Suppose $p \vdash \neg q$ for some $\neg q$. Then $\neg q \in Cn(\{ p \})$ by Lemma 38. By monotonicity $\neg q \in Cn(\{ p, q \})$. Furthermore, $q \in Cn(\{ p, q \})$ by Def. 21-(1). Since $q$ and $\neg q$ are contradistinctions, then by Def. 23(3)-a&b, $\{ q, \neg q \}$ is AL-inconsistent. Hence $Cn(\{ q, \neg q \}) = \mathcal{L}$. Then by Def. 21-(2) and monotonicity, $Cn(\{ p, q \}) = \mathcal{L}$.

**Proposition 25** Let $(AS, KB)$ be based on $(\mathcal{L}', Cn)$ and $(\Sigma, \leq')$. Let $\Delta_1$ be the c-SAF defined by $(AS, KB)$, and $\Delta_2$ the AL-c-SAF defined by $(AS, KB)$. Then, for $T \in \{ complete, grounded, preferred, stable \}$, $E$ is a $T$ extension of $\Delta_1$ iff $E$ is a $T$ extension of $\Delta_2$.

**Proof.** We first show that:

$X$ is acceptable w.r.t. $E$ in $\Delta_1$ iff $X$ is acceptable w.r.t. $E$ in $\Delta_2$ \hspace{1cm} (1)

1.1) Firstly, if $Y$ undermines $X$, then $p (\equiv Conc(Y)) \in \mathcal{P}$, where $q \in Prem(X)$. By def.23(3)-b, $(p, q)$ is AL-inconsistent. By Lemma 41, $p \vdash \neg q$. Hence $Y \ ASPIC^\pm$-AL-undermines $X$.

1.2) Secondly, if $Y$ ASPIC$^\pm$-AL-undermines $X$, then $p (\equiv Conc(Y)) \vdash \neg q$, where $q \in$
Lemma 42
Consider any \( A \)-inconsistent. Contradiction. QED

Suppose \( Y \) defeats \( X \) in \( \Delta_1 \) (\( Y \rightarrow_{\Delta_1} X \)), and so \( \exists Z \in E, Z \rightarrow_{\Delta_1} Y \). By 1.1), \( Y \rightarrow_{\Delta_2} X \) and \( Z \rightarrow_{\Delta_2} Y \). Hence the left to right half of 1) is shown. Suppose \( Y \rightarrow_{\Delta_3} X \), and so \( \exists Z \in E, Z \rightarrow_{\Delta_3} Y \). By 1.2) there is a strict extension \( Y' \) of \( Y \) s.t. \( Y' \) undermines \( X \). Since \( \preceq \) is reasonable, it remains the case that \( Y' \not\in E \), and so \( Y' \rightarrow_{\Delta_3} X \). By the same reasoning, there is a strict extension \( Z' \) of \( Z \), s.t. \( Z' \rightarrow_{\Delta_3} Y \). By Lemma 35, \( Z' \) is acceptable w.r.t. \( E \), and by Proposition 10, \( E \cup \{Z'\} \) is conflict free. Hence, since \( E \) is complete, \( Z' \in E \). Hence the right to left half of 1) is shown.

Given (1), the main proposition now follows from the following:

\[ E \text{ is conflict free in } \Delta_1 \iff E \text{ is conflict free in } \Delta_2 \] (2)

Suppose \( E \) is conflict free in \( \Delta_1 \), \( E \) is not conflict free in \( \Delta_2 \). Then \( \exists Y, X \in E \) such that \( Y \) does not undermine \( X \), \( X \) ASPIC+-AL-undermines \( Y \). And there is a strict extension \( Y' \) of \( Y \) s.t. \( Y' \) undermines \( X \). Applying the same reasoning as for \( Z' \) above, \( Y' \in E \), contradicting \( E \) is conflict free in \( \Delta_1 \). Suppose \( E \) is conflict free in \( \Delta_2 \), \( E \) is not conflict free in \( \Delta_1 \). Then \( \exists Y, X \in E \) such that \( Y \) does not ASPIC+-AL-undermines \( X \), \( X \) undermines \( Y \), contradicting 1.1. Hence (2) is shown.

QED

Proposition 26
Let \((AS, KB)\) be based on \((L', Cn)\) and \((\Sigma, \preceq')\). Then \( A \) is a c-consistent premise minimal argument on the basis of \((AS, KB)\) iff \((\text{Prem}(A), \text{Conc}(A))\) is an abstract logic argument on the basis of \((\Sigma, \preceq')\).

Proof. From right to left, let \((X, p)\) be an AL argument. Then \( X \rightarrow p \in R_+, \) and so \( A \) is a strict ASPIC+- argument with \( \text{Prem}(A) = X, \text{Conc}(A) = p \). \( A \) must be premise minimal since otherwise there is a strict ASPIC+- argument \( A' \) for \( p \) with \( X' = \text{Prem}(A'), \) \( X' \subseteq X \). But then by Lemma 38, \( p \in Cn(X') \), contradicting the minimality of \((X, p)\). Suppose for contradiction that \( A \) is not c-consistent. Then \( X \vdash p, \neg p \), and so by Lemma 38, \( \{p, \neg p\} \subseteq Cn(X) \). By Def. 23(b), \( \{p, \neg p\} \) is AL-inconsistent, and so \( Cn(\{p, \neg p\}) = L \). Then by Def. 21(b) and monotonicity, \( Cn(X) = L \), so \( X \) is AL-inconsistent. Contradiction.

From left to right, let \( A \) be a c-consistent premise minimal ASPIC+- argument with \( \text{Prem}(A) = X, \text{Conc}(A) = p \) (i.e., \( X \vdash p \)). Then \( X \subseteq \Sigma \) and \( p \in Cn(X) \) (by Lemma 38), satisfying (1) and (3) of Def. 22. If \( p \in Cn(X') \) for some \( X' \subseteq X \), then \( X' \rightarrow p \in R_+ \), and since \( X' \subseteq \Sigma_+ \), there is an \( A' \) s.t. \( \text{Prem}(A') = X', \text{Conc}(A) = p \), contradicting \( A \) is premise minimal. Hence condition (4) of Def. 22 is satisfied. Suppose for contradiction that \( X \) is AL-inconsistent. By repeated application of adjunction to the formulae in \( X \setminus \{\varphi\} \), for some \( \varphi \in X \), we have that \( Cn(X) = Cn(\{\varphi\}) = L' \). By Def. 23(c), \( \exists \varphi' \in Cn(\{\varphi\}) \), s.t. \( \varphi' \not\in P' \), and so \( \varphi' = \neg \varphi \). By monotonicity, \( \neg \varphi \in Cn(\{\varphi\}) \). By Def. 21(1), \( \varphi \in Cn(\{\varphi\}) \). Hence \( \varphi, \neg \varphi \in Cn(X) \). By Lemma 38, \( X \vdash \varphi, X \vdash \neg \varphi \); i.e., \( A \) is c-inconsistent. Contradiction.

QED

In the following lemma we introduce, for any argument \( A \in A \), the additional notation \( A_+ \) to denote any argument such that \( \text{Prem}(A) \subseteq \text{Prem}(A_+) \) and \( \text{Conc}(A) = \text{Conc}(A_+) \).

Lemma 42
Consider any \( AT \) for which \( \preceq \) is defined such that \( A \preceq A_- \).

1. If \( A \) defeats \( B \) then \( A_- \) defeats \( B_+ \) for all \( A_- \) and \( B_+ \).

2. For all complete extensions \( E \):

   (a) if \( A \in E \) then \( A_- \in E \) for all \( A_- \);

   (b) if \( B \not\in E \) then \( B_+ \not\in E \) for all \( B_+ \).

Proof. (1) If \( A \) defeats \( B \) based on a preference independent attack on a sub-argument \( B' \) of \( B \), then \( A_- \) preference independent attacks and defeats \( B_+ \) on \( B' \). If \( A \) preference dependent attacks and defeats \( B \) on \( B' \), then \( A \not\preceq B' \). By assumption of \( A \preceq A_- \), \( A_- \not\preceq B' \).
(2a) Let $A \in E$ and $B$ defeats any $A_\rightarrow$. Then $B$ also defeats $A$ by 1. Then there exists a $C$ that defeats $B$, so $A_\rightarrow$ is acceptable with respect to $A$, so (since $E$ is a complete extension) $A_\rightarrow \in E$.

(2b) Let $B \notin E$. Then there exists an $A$ that defeats $B$, and no $C \in E$ s.t. $C \rightarrow A$. But then $A$ defeats $B_\rightarrow$ by 1, so $B_\rightarrow$ is not acceptable with respect to $E$. QED

**Proposition 27** Let $\Delta$ be the (c)-SAF $(A, C, \preceq)$ defined on the basis of an AT for which $\preceq$ is defined such that for any $A \in A$, $A \preceq A_\rightarrow$. Let $\Delta_\rightarrow$ be the (c)-SAF $(A_\rightarrow, \preceq_\rightarrow, \preceq_\rightarrow)$ where:

- $A_\rightarrow$ is the set of premise minimal arguments in $A$.
- $C^{-}_\rightarrow = \{(X, \neg Y)(X, Y) \in C, X, Y \in A_\rightarrow\}$.
- $\preceq_\rightarrow = \{(X, \neg Y)(X, Y) \in \preceq, X, Y \in A_\rightarrow\}$.

Then for $T \in \{\text{complete, grounded, preferred, stable}\}$, $E$ is a $T$-extension of $\Delta_\rightarrow$ if $E'$ is a $T$-extension of $\Delta$, where $E' \subseteq E$ and $\bigcup_{X \in E} \text{Conc}(X) = \bigcup_{Y \in E'} \text{Conc}(Y)$.

**Proof.**

$T = \text{complete}$:

1) Suppose $E$ is a complete extension of $\Delta$. We show that $E_\rightarrow$ is a complete extension of $\Delta_\rightarrow$, where $E_\rightarrow = E \cap A_\rightarrow$.

Note first that $E_\rightarrow$ is (attack) conflict-free by construction.

(i) Let $A \in E_\rightarrow$ and consider any $B \in A_\rightarrow$ that defeats $A$. Since $A \in E$, there exists a $C \in E$ that defeats $B$. But then for all $C_\rightarrow$ we also have that $C_\rightarrow$ defeats $B$. Since all such $C_\rightarrow \in E$ (by Lemma 42(2)), all such $C_\rightarrow$ are in $E_\rightarrow$, and so $A$ is acceptable with respect to $E_\rightarrow$.

(ii) Let $A \in A_\rightarrow, A \notin E_\rightarrow$. Then $A \notin E$, some $B$ defeats $A$, $\neg\exists C \in E, C$ defeats $B$. There exists a $B_\rightarrow$ that (by Lemma 42(1)) defeats $A$. Suppose for contradiction that $A$ is an acceptable w.r.t. $E_\rightarrow$. Then, $\exists C' \in E_\rightarrow, C'$ defeats $B_\rightarrow$. By Lemma 42(1), $C'$ defeats $B$. Since $E_\rightarrow \subseteq E$, $C' \in E$, contradicting $\neg\exists C \in E, C$ defeats $B$. So $A$ is not acceptable with respect to $E_\rightarrow$. Given i) and ii), $E_\rightarrow$ is a complete extension of $\Delta_\rightarrow$.

2) Suppose $E$ is a complete extension of $\Delta_\rightarrow$. We show that $E_+ \rightarrow$ is a $T_\rightarrow$-extension of $\Delta_\rightarrow$, where $E_+ \rightarrow = E \cup \{A_+ \in A | A \in E_\rightarrow \text{ and } \text{Prem}(A_+) \subseteq E\}$. Suppose $E$ is a complete extension of $\Delta_\rightarrow$.

Note first that $E_+ \rightarrow$ is conflict-free by construction.

(i) For any $A \in E$ and any $B \in A$ that defeats $A$, we have that some $B_\rightarrow \in A_\rightarrow$ defeats $A$ by Lemma 42(1), so some $C \in E$ defeats $B_\rightarrow$. But then $C$ also defeats $B$ by Lemma 42(1). Since $C \in E_\rightarrow$, we have that $A$ is acceptable with respect to $E_+ \rightarrow$.

(ii) Consider any $A_+ \in E_+ \rightarrow, A_+ \notin E$. Suppose $B \in A$ defeats $A_+$, then $B$ defeats $A_+$ on some $A'$ that is a premise in $\text{Prem}(A_+)$. By definition of $E_+ \rightarrow$ and subargument closure of $E$, $A' \in E$. By Lemma 42(1), $B_\rightarrow \in A_\rightarrow$ defeats $A'$, and $B_\rightarrow$ is defeated by some $C \in E$. Since $C$ defeats $B$ (by Lemma 42(1)) and $C \in E_\rightarrow$, $A_+$ is acceptable with respect to $E_+ \rightarrow$.

(iii) Consider finally any $A \in A$, $A \notin E_+ \rightarrow$. Then no $A_\rightarrow$ is in $E$, so for all $A_\rightarrow$ there exists a $B$ defeats $A_\rightarrow, \neg\exists C \in E, C$ defeats $B$. By Lemma 42(1) $B$ also defeats $A$. Suppose for contradiction that $A$ is acceptable w.r.t. $E_+ \rightarrow$, and so $\exists C' \in E_+ \rightarrow, C'$ defeats $B$. Hence $C'$ must be some $C_\rightarrow$, where by Lemma 42-2 and construction of $E_+ \rightarrow$, $C_\rightarrow \in E_\rightarrow$ and $E_\rightarrow$. Since by assumption $C' \preceq C$, it remains the case that $C$ defeats $B$, contradicting $\neg\exists C \in E, C$ defeats $B$.

By i), ii) and iii), $E_+ \rightarrow$ is a complete extension of $\Delta_\rightarrow$.

$T = \text{preferred}$:

1) Suppose $E$ is a preferred extension of $\Delta$. Suppose for contradiction that $E_\rightarrow$ is not a preferred extension of $\Delta_\rightarrow$. We have shown that $E_\rightarrow$ is a complete extension of $\Delta_\rightarrow$. Hence
there must be some $E' \supset E$—that is a complete extension of $\Delta$. We have shown that $E' +$ is a complete extension of $\Delta$. It is easy to see by construction of $E$ and $E' +$ that $E \subset E' +$, contradicting $E$ is a preferred extension of $\Delta$.

2) Suppose $E$ is a preferred extension of $\Delta$. Suppose for contradiction that $E +$ is not a preferred extension of $\Delta$. We have shown that $E +$ is a complete extension of $\Delta$. Hence there must be some $E' \supset E+$ that is a complete extension of $\Delta$. We have shown that $E'$ is a complete extension of $\Delta$. It is easy to see by construction of $E'$ and $E+$, that $E \subset E'$, contradicting $E$ is a preferred extension of $\Delta$.

$T$ = grounded:

1) Suppose $E$ is the grounded extension of $\Delta$. Suppose for contradiction that $E$ is not the grounded extension of $\Delta$. We have shown that $E$ is a complete extension of $\Delta$. Hence there must be some $E' \subset E$ that is a complete extension of $\Delta$. We have shown that $E'$ is a complete extension of $\Delta$. It is easy to see by construction of $E$ and $E'$ that $E' + \subset E$, contradicting $E$ is the grounded extension of $\Delta$.

2) Suppose $E$ is the grounded extension of $\Delta$. Suppose for contradiction that $E +$ is not the grounded extension of $\Delta$. We have shown that $E +$ is a complete extension of $\Delta$. Hence there must be some $E' \subset E+$ that is a complete extension of $\Delta$. We have shown that $E'$ is a complete extension of $\Delta$. It is easy to see by construction of $E$ and $E'$, that $E' \subset E$, contradicting $E$ is the grounded extension of $\Delta$.

$T$ = stable:

1) Let $E$ be a stable (and so preferred) extension of $\Delta$. Then $E$ is a preferred extension of $\Delta$. Suppose for contradiction that $E$ is not a stable extension. Then $\exists B \in A$, $B \notin E$, and $B$ is not defeated by an argument in $E$. Note that $B \in A$. It cannot be that $B \in E$ since the fact that $B \in A$ would imply by construction of $E$ that $B \in E$. Since $E$ is stable, some $C \in E$ defeats $B$. By Lemma 42(1) all $C$ also defeat $B$, and by Lemma 42(2a) all such $C$ are in $E$. By construction of $E$ all such $C$ are in $E$, and so $B$ is defeated by an argument in $E$. Contradiction.

2) Let $E$ be a stable (and so preferred) extension of $\Delta$. Then $E$ is a preferred extension of $\Delta$. Suppose for contradiction that $E +$ is not a stable extension. Then $\exists B \in A$, $B \notin E+$, and $B$ is not defeated by an argument in $E$. Since $E$ is preferred, $B$ is defeated by a $C \in A \setminus E+$, where $C$ is not defeated by an argument in $E$. $C$ defeats $B$ on some $\varphi \in \text{Prem}(B)$. Note that $\varphi \in A$. If $\varphi \in E$ then, since $\varphi$ acceptable w.r.t $E$, $C$ is defeated by some argument in $E$. But since $E \nsubseteq E+$ this contradicts that $C$ is not defeated by an argument in $E$. If $\varphi \notin E$ then, since $E$ is a stable extension of $\Delta$, we have that $\varphi$ is defeated by a $D \in E$, but then $D$ also defeats $B$. Since $D \in E$ this contradicts that $B$ is not defeated by an argument in $E+$.

Finally, we clearly have for any $E$ and $E$ that $\text{Conc}(E) = \text{Conc}(E)$, and likewise for any $E$ and $E$. Then the proposition follows from (1) and (2) for each of the above semantics.

QED

Corollary 28 Given $\Delta$ and $\Delta^-$ as defined in Proposition 27:

1. $\varphi$ is a $T$ credulously (sceptically) justified conclusion of $\Delta$ iff $\varphi$ is a $T$ credulously (sceptically) justified conclusion of $\Delta^-$.  

2. For $\Delta^-$ satisfies all of closure under strict rules, direct consistency, indirect consistency and sub-argument closure under any of the semantics subsumed by the complete semantics.

Proof. 1) and closure under strict rules, direct consistency, and indirect consistency immediately follow from Proposition 27. For sub-argument closure expressed in Theorem 12, note that the proof of this theorem appeals to Lemma 33 which can straightforwardly be seen to apply to $\Delta^-$. The proof also depends on any sub-argument of $A \in E$ not being in
conflict with any argument in \( E \). This immediately follows for \( E- \) in the proof of Theorem 27, given that \( E- \subseteq E \).

**Proposition 29** Let \((A, C, \leq)\) be defined by an AL argumentation theory, where \( \leq \) is defined under the weakest or last link principles, based on the set orderings \( \prec_{E11} \) and \( \prec_{E12} \). Then \( \forall A, B \in A, \text{if } \text{Prem}(A) \subset \text{Prem}(B), \text{then } A \not\prec B \).

**Proof.** Since all arguments are strict extensions of ordinary premises, the last and weakest link principles are evaluated in the same way. For \( \prec_{E11}, \text{if } A \prec B, \text{then } \exists l \in \text{Prem}(A), l \not\prec l. \text{Contradiction. For } \prec_{E12}, \text{since } \text{Prem}(A) \setminus \text{Prem}(B) = \emptyset, \text{then it cannot be that } \text{Prem}(A) \prec_{E12} \text{Prem}(B). \)**

**Proposition 31** Let \( \Delta \) be the c-SAF defined by an AL argumentation theory. Then for any complete extension \( E \) of \( \Delta \): \( S = \{ \phi \in \text{Prem}(A), A \in E \} \) is AL-inconsistent iff \( S' = \text{Cl}(\{ \text{Conc}(A) \mid A \in E \}) \) is inconsistent.

**Proof.** Left to right: if \( S \) is AL inconsistent then for any \( \phi, \varphi, \neg \varphi \in Cn(S) \). By definition of \( R_s \), for some \( T, T' \subseteq S \) there exist rules \( T \rightarrow \varphi \) and \( T' \rightarrow \neg \varphi \) in \( R_s \). Since \( E \) is closed under sub-arguments and premises are sub-arguments, \( \{ \text{Conc}(A) \mid A \in E \} \) includes \( T \) and \( T' \). Hence \( \varphi, \neg \varphi \in S' \). That is, \( S' \) is inconsistent.

Right to left: If \( S' \) is inconsistent then \( \varphi, \neg \varphi \in S' \) for some \( \varphi \). Since \( E \) is closed under sub-arguments, \( S \subseteq S' \), and so \( S \vdash \varphi \) and \( S \vdash \neg \varphi \). By Def. 23(3b), \( \{ \varphi, \neg \varphi \} \) is LA-inconsistent, so \( Cn(\{ \varphi, \neg \varphi \}) = \emptyset \). But since \( \{ \varphi, \neg \varphi \} \subseteq Cn(S) \) and \( Cn(Cn(S)) = Cn(S) \), we have by monotonicity of \( Cn \) that \( Cn(S) = \emptyset \) so \( S \) is LA-inconsistent.

**8.6 Proofs for Section 5.3**

**Theorem 32** Let \((A, C, \leq)\) be a c-SAF corresponding to a default theory \( \Gamma \), and for any \( \Sigma \subseteq \Gamma \), let \( \text{Args}(\Sigma) \subseteq A \) be the set of all arguments with premises taken from \( \Sigma \). Then:

1) If \( \Sigma \) is a preferred subtheory of \( \Gamma \), then \( \text{Args}(\Sigma) \) is a stable extension of \((A, C, \leq)\).

2) If \( E \) is a stable extension of \((A, C, \leq)\), then \( \bigcup_{A \in E} \text{Prem}(A) \) is a preferred subtheory of \( \Gamma \).

**Proof.**

1) Firstly, we show that \( \text{Args}(\Sigma) \) is conflict free. Since \( \Sigma \) is consistent, \( \Sigma \not\vdash \alpha, \neg \alpha \) for any \( \alpha \). Suppose for contradiction that \( \text{Args}(\Sigma) \) is not conflict free, in which case \( \exists X, Y \in \text{Args}(\Sigma) \) s.t. \( \text{Conc}(X) \equiv \alpha, \neg \alpha \in \text{Prem}(Y) \). But then since every such argument is obtained by applying the strict rules encoding all classical inferences to \( \Sigma \), this implies \( \Sigma \vdash \alpha, \neg \alpha \). Contradiction.

We now show that for any \( Y \in A \setminus \text{Args}(\Sigma) \), \( \exists X \in \text{Args}(\Sigma) \) s.t. \( X \) defeats \( Y \). Consider any such \( Y \). Then \( \exists \gamma \in \text{Prem}(Y), \gamma \notin \Sigma \). By construction, \( \Sigma = \Sigma_1 \cup \ldots \cup \Sigma_n \) such that for \( i = 1, \ldots n, \Sigma_i \cup \ldots \cup \Sigma_n \) is a maximal consistent subset of \( \Gamma_1, \ldots, \Gamma_j \). Hence, suppose \( \gamma \in \Gamma_j \) for some \( j = 1, \ldots n \). Then \( \Sigma_1 \cup \ldots \cup \Sigma_j \cup \{ \gamma \} \vdash \perp \). Hence \( \Sigma_1 \cup \ldots \cup \Sigma_j \vdash \neg \gamma \). Hence, \( \exists X \in \text{Args}(\Sigma_1 \cup \ldots \cup \Sigma_j) \) s.t. \( \text{Conc}(X) \equiv \gamma \), and so \( X \rightarrow Y \). Since \( \gamma \in \Gamma_j \), and all premises in \( X \) are in \( \Gamma_i \), \( i \leq j \), then by the weakest or last link principle, under any of the definitions of \( \prec_\alpha \) in Def. 17, \( X \not\vdash Y \). Hence \( X \rightarrow Y \).

2) Firstly, we show that \( \bigcup_{A \in E} \text{Prem}(A) \) must be consistent. Suppose for contradiction that \( \exists X, Y \in E \) s.t. \( \text{Prem}(X) \cup \text{Prem}(Y) \vdash \perp \). Let \( \{ \alpha_1, \ldots, \alpha_m \} \) be a minimal (under set inclusion) subset of \( \text{Prem}(X) \cup \text{Prem}(Y) \) s.t. \( \alpha_1, \ldots, \alpha_m \vdash \perp \). Hence, \( \alpha_1, \ldots, \alpha_{m-1} \vdash \neg \alpha_m \). Since \( E \) is stable and so complete, then by sub-argument closure (Theorem 12), \( \{ A_1, \ldots, A_m \} \subseteq E \), where for \( i = 1, \ldots m, \text{Prem}(A_i) = \{ \alpha_i \} \). By Lemma 35, if \( \{ A_1, \ldots, A_m \} \subseteq E \), where \( E \) is a complete extension, then any strict extension of \( \{ A_1, \ldots, A_m \} \) is acceptable w.r.t. \( E \), and so in \( E \). Hence \( A \in E \) where \( A \) concludes \( \neg \alpha_m \).
Hence, \( A \rightarrow A_m \), contradicting \( E \) is conflict free.

Next, let \( E_1, \ldots, E_n \) be the partition of \( \text{Form}(E) \) s.t. for \( i = 1 \ldots n \), \( E_i \) is a (possibly) empty subset of \( \Gamma_i \) in the stratification \( \Gamma_1, \ldots, \Gamma_n \) of \( \Gamma \). Suppose for contradiction that \( \text{Form}(E) \) is not a preferred subtheory. Then, for some \( i \), for \( k = 1 \ldots i - 1, E_1, \ldots, E_k \) is a maximal consistent subset of \( \Gamma_1, \ldots, \Gamma_{i-1} \), and \( \exists \alpha \in \Gamma_i \) s.t. \( \alpha \notin E_i \), and \( E_1 \cup \ldots \cup E_{i-1} \cup E_i \cup \{\alpha\} \not\models \bot \). Hence, \( \exists Y \in A, \text{Prem}(Y) = \{\alpha\}, Y \notin E \). By assumption of \( E \) being a stable extension, \( \exists X \in E, X \rightarrow Y \). Since \( E_1 \cup \ldots \cup E_{i-1} \cup E_i \cup \{\alpha\} \not\models \bot \), then \( E_1 \cup \ldots \cup E_{i-1} \cup E_i \not\models \neg \alpha \), and so it must be that some \( \beta \in \text{Prem}(X) \) is in \( E_j, j > i \). But then \( \text{Prem}(X) \not\models \text{Prem}(Y) \), and \( X \prec Y \) under the weakest or last link principle, contradicting \( X \rightarrow Y \).

\[ \text{QED} \]

9 Proofs for Resolutions

**Proposition 0** Let \((A, C, \preceq)\) be a SAFE that \textit{preference-extends} \((A, C, \preceq')\), and let \(D'\) and \(D\) be defeat relations respectively defined by \(\preceq'\) and \(\preceq\). Then \(D' \subseteq D\).

\[ \text{Proof.} \text{ Proof is by contradiction. Suppose } A \rightarrow_{D'} B ((A, B) \in D') \text{ and } A \rightarrow_D B ((A, B) \notin D). \text{ Then } A \rightarrow B \text{ on } B', A \prec B'. \text{ Since } \preceq' \text{ extends } \preceq, A \prec' B', \text{ and so } A \prec_{D'} B. \text{ Contradiction.} \]

\[ \text{QED} \]

**Proposition 1** Let \(\Delta = (A, C, \preceq)\) be a (possibly c-consistent) SAFE defined by an argumentation theory consisting of a knowledge base \((K, \preceq_1)\) in an argumentation system \((\mathcal{L}, \neg, \mathcal{R}, \preceq_2)\), and where \(\preceq\) is defined according to the weakest or last principles. For any \(\preceq_1\) that extends \(\preceq_1\), and any \(\preceq_2\) that extends \(\preceq_2\), the SAFE \(\Delta' = (A, C, \preceq')\) defined by \((K, \preceq_1)\) in \((\mathcal{L}, \neg, \mathcal{R}, \preceq_2')\), \textit{preference-extends} \(\Delta\).

\[ \text{Proof.} \text{ In Section 5.1 the weakest and last link orderings over arguments are defined based on set orderings over the defeasible rules and non-axiom premises. The latter, set orderings } \prec_{s}, \text{ are in turn defined (Definition 17) based on the pre-orderings over the defeasible rules and non-axiom premises. Given } \Gamma = \{i_1, \ldots, i_n\}, \Gamma' \subseteq \Gamma, \Gamma'' \subseteq \Gamma, \text{ a pre-ordering on } \Gamma, \text{ then:} \]

- \(\Gamma' \prec_{s} \Gamma'' \) iff \( \exists \gamma' \in \Gamma' \) \( (\Gamma' \setminus \Gamma'') \) s.t. \( \forall \gamma'' \in \Gamma'' \) \( (\Gamma'' \setminus \Gamma') \), \( \gamma' \prec \gamma'' \).
- \(\Gamma' \prec_{s} \Gamma'' \) iff \( \forall \gamma' \in \Gamma' \) \( (\Gamma' \setminus \Gamma'') \) s.t. \( \exists \gamma'' \in \Gamma'' \) \( (\Gamma'' \setminus \Gamma') \), \( \gamma' \prec \gamma'' \).

\[ \text{It sufficient to show that these set orderings are preserved; that is, if } \Gamma' \prec_{s} \Gamma'' \text{ based on } \preceq, \text{ then } \Gamma' \prec_{s} \Gamma'' \text{ based on any } \preceq' \text{ that extends } \preceq. \text{ This immediate to see, given that } x \prec y \text{ implies } x \prec' y. \]

\[ \text{QED} \]

In what follows, if \(\Sigma\) is a preferred subtheory of \((\Gamma, \preceq) = (\Gamma_1, \ldots, \Gamma_m)\), we may write \(\Sigma\) as the tuple \((\Sigma_1, \ldots, \Sigma_m)\), where for \(i = 1 \ldots m\), \(\Sigma_i = \Sigma \cap \Gamma_i\). We say that \((\Sigma_1, \ldots, \Sigma_m)\) is the \(\Gamma\) partition of \(\Sigma\).

**Proposition 9** Let \((\Gamma, \preceq')\) be any default theory extending \((\Gamma, \preceq)\). Then \(\Sigma\) is a preferred subtheory of \(\Gamma'\) implies \(\Sigma\) is a preferred subtheory of \(\Gamma\).

\[ \text{Proof.} \text{ It follows from the definition of } \text{extends} \text{ (Definition 6 in the resolutions paper), that if } \Gamma' \text{ extends } \Gamma, \text{ then:} \]

- \(\Gamma = (\Gamma_1, \ldots, \Gamma_m)\) and \(\Gamma' = (\Gamma_1, \ldots, \Gamma_m)\) where:
  - \(\Gamma_1 \equiv \Gamma_1 \cup \ldots \cup \Gamma_m = \Gamma_{m_1} \cup \ldots \cup \Gamma_{m_k}\).
  - \(\Sigma_1 = \Sigma_1 \cup \ldots \cup \Sigma_{m_1} = \Sigma_1 \cup \ldots \cup \Sigma_{m_k}\).

\[ \text{Let } (\Sigma_1, \ldots, \Sigma_1, \ldots, \Sigma_{m_k}) \text{ be the } \Gamma' \text{ partition of } \Sigma. \text{ Then } (\Sigma_1, \ldots, \Sigma_m) \text{ is the } \Gamma \text{ partition of } \Sigma \text{ where:} \]

- \(\Sigma_1 = \Sigma_1 \cup \ldots \cup \Sigma_{m_1} \cup \ldots \cup \Sigma_{m_k}\).

\[ \text{We show that } \Sigma = (\Sigma_1, \ldots, \Sigma_m) \text{ is a preferred subtheory of } \Gamma, \text{ by showing that:} \]
Proof is by induction:

**Base case:** By definition \( \Sigma_1 = \Sigma_{i+1} \cup \ldots \cup \Sigma_{j+1} \) maximally consistently extends \( \Sigma_1 \cup \ldots \cup \Sigma_i, \ldots, \Sigma_{j+1} \), and \( \Sigma_{j+1} \) is consistent. Suppose \( \Sigma_1 \) is not a maximal consistent subset of \( \Gamma_1 \), and so \( \exists \beta \in \Gamma_1 \setminus \Sigma_1 \) s.t. \( \Sigma_1 \cup \{ \beta \} \not\models \top \). But then irrespective of which \( \Gamma_1 \), contains \( \beta, \beta \) can be consistently included in \( \Sigma_1 \), contradicting \( \Sigma_1 = \Sigma_{i+1} \cup \ldots \cup \Sigma_j \).

**General case:** Let \( \Sigma_{j+1}, \ldots, \Sigma_{j+k} \) maximally consistently extend \( \Sigma_{i+1}, \ldots, \Sigma_{i+k}, \ldots, \Sigma_{j+1} \), \( \Sigma_{j+1} \cup \ldots \cup \Sigma_{j+k} \). By the same reasoning in the base case, \( \Sigma_{j+1} \) maximally consistently extends \( \Sigma_1, \ldots, \Sigma_j \).

\[ \text{QED} \]

**Proposition 10** Let \( (\Gamma, \leq^1), \ldots, (\Gamma, \leq^m) \) be all the default theories extending \( (\Gamma, \leq) \). Suppose that for \( i = 1 \ldots n, \alpha \in \cap \mathcal{PS}(\Gamma_i) \). Then \( \alpha \in \cap \mathcal{PS}(\Gamma) \).

**Proof:** We prove the contraposition:

\[ \exists \Sigma \in \mathcal{PS}(\Gamma), \alpha \notin \Sigma \Rightarrow \exists \Sigma' \in \mathcal{PS}(\Gamma'), \alpha \notin \Sigma' \quad (1) \]

Let \( \Gamma \) be the \( \Gamma_1, \ldots, \Gamma_m \) and the associated complete preordering. (1) trivially holds if the strict counterpart \( \prec \) of \( \leq \) is a total ordering (i.e., each \( \Gamma_i \) is a singleton set) since \( \Gamma \) is then the only extension of \( \Gamma \). Hence for some \( i = 1 \ldots m, |\Gamma_i| > 1 \).

Now, suppose \( \Sigma \in \mathcal{PS}(\Gamma), \alpha \notin \Sigma \), and let \( (\Sigma_1, \ldots, \Sigma_m) \) be the \( \Gamma \) partition of \( \Sigma \). It suffices to show that we can always define an extension \( \Gamma' \) of \( \Gamma \) s.t. \( \Sigma \in \mathcal{PS}(\Gamma') \). Suppose:

i) \( \Gamma_i \setminus \Sigma_i = \emptyset \) (i.e., \( \Sigma_i = \Gamma_i \)), in which case it is straightforward to see that \( \Sigma \) is a preferred subtheory of any extension with partition \( \Gamma_1, \ldots, \Gamma_{i-1}, \Gamma_{i+1}, \ldots, \Gamma_{i+k-1}, \Gamma_{i+k+1}, \ldots, \Gamma_m \), where by virtue of \( |\Gamma_i| > 1 \), one can always guarantee \( k > 1 \) for such an extension.

ii) \( \Gamma_i \setminus \Sigma_i \neq \emptyset \). Then let \( \Gamma_1, \ldots, \Gamma_{i-1}, \Gamma_{i+1}, \ldots, \Gamma_{i+k}, \Gamma_{i+1}, \ldots, \Gamma_m \) be any extension s.t. \( \Sigma_i = \Gamma_1 \cup \ldots \cup \Gamma_{i+k} \cup \ldots \cup \Gamma_m \), where \( \Gamma_{i+k} = \Gamma_{i+k} \setminus \Sigma_i \), and where one can guarantee \( k > 1 \) by virtue of \( |\Gamma_i| > 1 \) (note that in the case that \( \Sigma_i = \emptyset, \Gamma_i, \ldots, \Gamma_{i+k} \) is just the partitioning of \( \Gamma_i \)). It is then straightforward to see that \( \Sigma \) is a preferred subtheory of any such extension.

\[ \text{QED} \]

**References**

[1] L. Amgoud and P. Besnard. Bridging the gap between abstract argumentation systems and logic. In *Proc. 3rd International Conference on Scalable Uncertainty (SUM’09)*, pages 12–27, 2009.

[2] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems. In *Proc. 4th International Conference on Scalable Uncertainty (SUM’10)*, pages 42–55, 2010.

[3] L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, J. van Yeon H. Prakken, and G.A.W. Vreeswijk. Final review and report on formal argumentation system. deliverable d2.6, aspic-isfp-002307. Technical report, 2006.

[4] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Automated Reasoning*, 29 (2):125–169, 2002.
[5] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[6] L. Amgoud and G. Vesic. Generalizing stable semantics by preferences. In *Proc. Computational Models of Argument 2010*, pages 39–50, 2010.

[7] L. Amgoud and S. Vesic. Repairing preference-based argumentation frameworks. In *Proc. 21st International Joint Conferences on Artificial Intelligence*, pages 665–670, 2009.

[8] L. Amgoud and S. Vesic. Handling inconsistency with preference-based argumentation. In *Proc. 4th International Conference on Scalable Uncertainty Management, (SUM’10)*, pages 56–69, 2010.

[9] L. Amgoud and S. Vesic. Two roles of preferences in argumentation frameworks. In *Proc. 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 86–97, 2011.

[10] L. Amgoud and S. Vesic. A new approach for preference-based argumentation frameworks. *Annals of Mathematics and Artificial Intelligence*, to appear.

[11] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[12] T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171:10–15, 2007.

[13] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001.

[14] P. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.

[15] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[16] G. Brewka. Preferred subtheories: an extended logical framework for default reasoning. In *Proc. 11th International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.

[17] G. Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge University Press, 1991.

[18] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[19] M. Caminada, W. Carnielli, and P.E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, to appear.

[20] C. Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In *Proc. 14th International Joint Conference on Artificial Intelligence*, pages 1443–1448, 1995.
[21] C. Cayrol, S. Doutre, and J. Mengin. On Decision Problems related to the preferred semantics for argumentation frameworks. *Journal of Logic and Computation*, 13(3):377–403, 2003.

[22] P. M. Dung. An argumentation semantics for logic programming with explicit negation. In *Proc. 10th Logic Programming Conference*, pages 616–630, 1993.

[23] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[24] P. E. Dunne and T. J. M. Bench-Capon. Two party immediate response dispute: Properties and efficiency. *Artificial Intelligence*, 149:221–250, 2003.

[25] M.L. Ginsberg. AI and nonmonotonic reasoning. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 1–33. Clarendon Press, Oxford, 1994.

[26] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.

[27] G. Governatori and M. J. Maher. An argumentation-theoretic characterization of defeasible logic. In *Proc. 14th European Conference on Artificial Intelligence*, pages 469–473, 2000.

[28] S. Kaci. Refined preference-based argumentation frameworks. In *Proc. Computational Models of Argument (COMMA) 2010*, pages 299–310, 2010.

[29] F. Lin and Y. Shoham. Argument systems: A uniform basis for nonmonotonic reasoning. In *Proc. 1st International Conference on Principles of Knowledge Representation and Reasoning (KR-89)*, pages 245–255, 1989.

[30] R.P. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 2:100–106, 1987.

[31] H. Mercier and D. Sperber. Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–747, 2011.

[32] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[33] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in AI*, pages 105–129. Springer-Verlag, 2009.

[34] S. Modgil and H. Prakken. Reasoning about preferences in structured extended argumentation frameworks. In *Proc. Computational Models of Argument*, pages 347–358, 2010.

[35] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[36] J. L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[37] J. L. Pollock. Justification and defeat. *Artificial Intelligence*, 67:377–408, 1994.
[38] J. L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, 1995.

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

[40] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.

[41] I. Rahwan and G. Simari, editors. *Argumentation in AI*. Springer-Verlag, 2009.

[42] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence*, 53:125–157, 1992.

[43] G. A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[44] G. A. W. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proc. 7th European Workshop on Logic for Artificial Intelligence*, pages 239–253, 2000.

[45] G.A.W. Vreeswijk. *Studies in Defeasible Argumentation*. Doctoral dissertation Free University Amsterdam, 1993.