



# Classical logic, argument and dialectic

M. D'Agostino<sup>a</sup>, S. Modgil<sup>b,\*</sup>

<sup>a</sup> Department of Philosophy, University of Milan, Italy

<sup>b</sup> Department of Informatics, King's College London, United Kingdom



## ARTICLE INFO

### Article history:

Received 10 July 2017

Received in revised form 10 May 2018

Accepted 24 May 2018

Available online 20 June 2018

### Keywords:

Classical logic

Argumentation

Dialectic

Rationality postulates

Natural deduction

Preferred subtheories

Bounded reasoning

## ABSTRACT

A well studied instantiation of Dung's abstract theory of argumentation yields argumentation-based characterisations of non-monotonic inference over possibly inconsistent sets of classical formulae. This provides for single-agent reasoning in terms of argument and counter-argument, and distributed non-monotonic reasoning in the form of dialogues between computational and/or human agents. However, features of existing formalisations of classical logic argumentation (*Cl-Arg*) that ensure satisfaction of rationality postulates, preclude applications of *Cl-Arg* that account for real-world dialectical uses of arguments by resource-bounded agents. This paper formalises *dialectical classical logic argumentation* that both satisfies these practical desiderata and is provably rational. In contrast to standard approaches to *Cl-Arg* we: 1) draw an epistemic distinction between an argument's premises accepted as true, and those assumed true for the sake of argument, so formalising the dialectical move whereby arguments' premises are shown to be inconsistent, and avoiding the *foreign commitment* problem that arises in dialogical applications; 2) provide an account of *Cl-Arg* suitable for real-world use by eschewing the need to check that an argument's premises are subset minimal and consistent, and identifying a minimal set of assumptions as to the arguments that must be constructed from a set of formulae in order to ensure that the outcome of evaluation is rational. We then illustrate our approach with a natural deduction proof theory for propositional classical logic that allows measurement of the 'depth' of an argument, such that the construction of *depth-bounded* arguments is a tractable problem, and each increase in depth naturally equates with an increase in the inferential capabilities of real-world agents. We also provide a resource-bounded argumentative characterisation of non-monotonic inference as defined by Brewka's Preferred Subtheories.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Argumentation is a form of reasoning that makes explicit the reasons for the conclusions that are drawn and how conflicts between reasons are resolved. While informal studies of argumentation have a rich tradition, recent years have witnessed intensive study of logic-based models of argumentation and their use in formalising agent reasoning, decision making, and inter-agent dialogue [11,53]. Much of this work builds on Dung's seminal theory of abstract argumentation [26], and the theory's provision of argumentative characterisations of nonmonotonic inference. Given a possibly inconsistent set of logical formulae (*base*) one defines the arguments and a binary *attack* relation denoting that one argument is a

\* Corresponding author.

E-mail address: sanjay.modgil@kcl.ac.uk (S. Modgil).

counter-argument to another. Various developments of Dung's theory additionally accommodate a preference relation over arguments, which is used to determine which attacks succeed as defeats [1,9,43]. The resulting directed graph of arguments related by defeats, referred to as an *argumentation framework* (AF), is then said to be 'instantiated' by the base. Evaluation of the justified arguments is then based on the intuitive principle that an argument is justified if all its defeaters are themselves defeated by justified arguments. The conclusions of justified arguments identify the 'argumentation defined' non-monotonic inferences from the instantiating base.

The widespread impact of Dung's theory is in large part due to this characterisation of non-monotonic inference in terms of the dialectical use of arguments and counter-arguments familiar in everyday reasoning and debate. The theory thus provides foundations for reasoning by individual computational and human agents, and distributed non-monotonic reasoning involving agents resolving conflicts amongst beliefs or deciding amongst alternative actions, or negotiating allocations of resources (e.g., [2,3,8,38,40,44,45,48,58]). These 'monological' and 'dialogical' applications have motivated the study of rationality postulates for logical instantiations of AFs [15,16], as well as desiderata for practical applications [27,44].

This paper focuses on classical logic instantiations of AFs (*Cl-Arg*) [1,33,43]. Features of the current paradigm have been shown to provide sufficient conditions for satisfaction of the rationality postulates. However, these features preclude satisfaction of practical desiderata that account for modelling real-world uses of arguments by resource-bounded agents. This paper therefore aims at an account of *Cl-Arg* that satisfies both practical desiderata and the rationality postulates.

In Section 2 we review Dung's theory, *Cl-Arg*, and the rationality postulates. In Section 3 we argue that monological and dialogical applications of Dung's theory require formalisation of real-world uses of argument suitable for resource-bounded agents. However, current approaches to *Cl-Arg* tacitly assume that *all* arguments defined by a base can be constructed and included in an AF, and that prior to inclusion the legitimacy of each constructed argument is verified by checking that its premises are consistent and not redundant in the strong sense that their conclusion is not entailed by any proper subset of the premises. These assumptions are computationally unfeasible (even in the propositional case) for real-world uses of argument by resource-bounded agents. However, they are proposed as sufficient conditions for satisfaction of the *consistency* and *closure* postulates [15] for first order *Cl-Arg* with preferences [43],<sup>1</sup> and of the 'non-contamination' postulates [16] for propositional *Cl-Arg* without preferences. Moreover, checking the legitimacy of arguments prior to inclusion in an AF fails to account for real-world uses of argument. *Firstly*, in real-world uses the inconsistency of arguments' premises is typically demonstrated dialectically. *Secondly*, agents do not interrogate premises for subset minimality. Rather, it is the specific proof-theoretic means for constructing arguments that determines whether or not premises are redundantly used in deriving the conclusion; that is, redundant in the obvious sense that they are syntactically disjoint from the remaining premises and the conclusion.

Section 3 then presents a new account of first order *Cl-Arg* that satisfies practical desiderata. Our approach introduces a new notion of argument that distinguishes amongst the premises accepted as true and those supposed true 'for the sake of argument'. We can therefore model a ubiquitous feature of dialectical *practice*, whereby the inconsistency of a set of premises  $\Gamma$  is shown *dialectically*, by defeats from arguments that claim that a contradiction is implied if one supposes (for the sake of argument) the truth of  $\Gamma$ . The distinction also solves the so called *foreign commitment* problem that arises in dialogical applications when agents are forced to commit to the premises of their interlocutors in order to challenge their arguments [17]. We also drop the computationally demanding checks on the legitimacy of arguments, and define 'partially instantiated' AFs that include subsets of the arguments defined by a base. We thus accommodate real-world uses of argument in which agents do not (or may not have sufficient resources to) construct all arguments from a base when determining whether arguments are justified. We show that our account satisfies standard properties of Dung's theory. We also show that despite dropping the legitimacy checks on arguments and making minimal assumptions as to the arguments defined by a base for inclusion in an AF, the consistency and closure postulates are satisfied (where the latter are adapted to account for the fact that not all defined arguments may be included in the AF). Moreover, in contrast with [43], these postulates are satisfied assuming *any* preference relation over arguments.

Finally, in Section 3 we identify the notion of an argument whose use of obviously redundant (in the sense described above) premises can be excluded proof-theoretically, in contrast with the use of impractical subset-minimality checks. We generalise the 'non-contamination' postulates defined for propositional instantiations of AFs in [16], to first order instantiations. We then show that despite dropping consistency and subset minimality checks on arguments' premises, our account of first order *Cl-Arg* satisfies these postulates under the assumption that preference relations are 'coherent'.

Standard accounts of *Cl-Arg* typically leave implicit the specific proof theoretic means by which one entails a conclusion from a set of premises. In Section 4 we illustrate use of our dialectical account of argumentation by formalising arguments as *intelim* trees: a new natural deduction formalism for propositional classical logic [20,21] that allows measurement of the 'depth' of an argument such that the construction of *depth-bounded* arguments is a tractable problem, and each increase in depth naturally equates with an increase in the inferential capabilities of real-world agents. We then show that AFs instantiated by arguments up to any given depth satisfy the rationality postulates and practical desiderata. Furthermore, *intelim* natural deduction allows for a notion of proof that excludes arguments that use obviously redundant premises. Section 4 also develops a resource-bounded argumentative characterisation of non-monotonic reasoning in Brewka's Preferred Subtheories [13].

<sup>1</sup> Consistency postulates closely related to [15] are also studied for propositional *Cl-Arg* without preferences in [33].

Section 5 reviews related work; in particular the practical desiderata described in [27], and works [34,61] that exclude arguments that make use of inconsistent formulae so as to ensure satisfaction of [16]'s non-contamination postulates. We then conclude in Section 6 and outline directions for future work.

## 2. Background

### 2.1. Classical logic instantiations of Dung's argumentation theory

We first recapitulate Dung's argumentation theory [26].

**Definition 1** (*Dung framework, conflict free and acceptability*). A Dung argumentation framework (AF) is a tuple  $(\mathcal{A}, \mathcal{C})$ , where  $\mathcal{A}$  is a set of arguments and  $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$  is a binary attack relation defined over  $\mathcal{A}$ . Let  $S \subseteq \mathcal{A}$ . Then:

- $S$  is *conflict free* iff  $\forall X, Y \in S: (X, Y) \notin \mathcal{C}$ .
- $X \in \mathcal{A}$  is *acceptable* with respect to  $S$  iff  $\forall Y \in \mathcal{A}$  such that  $(Y, X) \in \mathcal{C}: \exists Z \in S$  such that  $(Z, Y) \in \mathcal{C}$ .

**Definition 2** (*Dung semantics*). Let  $(\mathcal{A}, \mathcal{C})$  be an AF, and  $S \subseteq \mathcal{A}$  conflict free. Then  $S$  is an *admissible* extension iff  $X \in S$  implies  $X$  is acceptable w.r.t.  $S$ .

- An admissible extension  $S$  is a *complete* extension iff  $\forall X \in \mathcal{A}: X$  is acceptable w.r.t.  $S$  implies  $X \in S$ ;
- $S$  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  $\forall Y \in \mathcal{A}$ , if  $Y \notin S$  then  $\exists X \in S$  s.t.  $(X, Y) \in \mathcal{C}$ .

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

Extensions can also be defined in terms of an AF's characteristic function  $\mathcal{F}_{(\mathcal{A}, \mathcal{C})}$ :

$$\mathcal{F}_{(\mathcal{A}, \mathcal{C})}(S) = \{X | X \text{ is acceptable w.r.t. } S\}, \text{ where } S \subseteq \mathcal{A}.$$

Then for any conflict free  $S \subseteq \mathcal{A}$ ,  $S$  is: admissible iff  $S \subseteq \mathcal{F}_{(\mathcal{A}, \mathcal{C})}(S)$ ; complete iff  $S = \mathcal{F}_{(\mathcal{A}, \mathcal{C})}(S)$  (i.e.,  $S$  is a fixed point of  $\mathcal{F}_{(\mathcal{A}, \mathcal{C})}$ ); grounded iff  $S$  is the least fixed point of  $\mathcal{F}_{(\mathcal{A}, \mathcal{C})}$ ; preferred iff  $S$  is a maximal fixed point of  $\mathcal{F}_{(\mathcal{A}, \mathcal{C})}$ .

Dung explicitly considered 'instantiations' of AFs by sets of logical formulae  $\Delta$ , and showed that his theory provides a general framework for various species of non-monotonic reasoning. Given  $\Delta$  in some logic  $L$ , the conclusions of the justified arguments (i.e., the argumentation defined inferences from the instantiating  $\Delta$ ) are exactly those obtained from  $\Delta$  by the inference relation of the logic. Dung thus provides argumentation based characterisations of logic programming, Reiter's Default Logic [54] and Pollock's Inductive Defeasible logic [47].

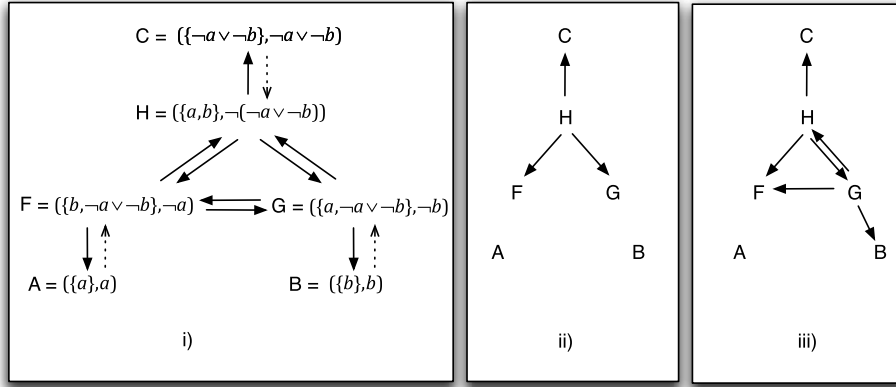
We now review classical logic instantiations of AFs (*Cl-Arg*). We assume a full first order language  $\mathcal{L}$  consisting of the usual logical operators  $\wedge, \vee, \rightarrow, \neg, \forall, \exists$ , a countable set of individual variables, a (possibly empty) set of function symbols of various arities, where as usual function symbols of arity 0 are interpreted as constant symbols, and a non-empty set of predicate symbols of various arities that includes the 0-ary symbol  $\perp$  interpreted as a patently false atomic formula (i.e. one that is always false under any interpretation of the language). A *term* is an individual variable or constant, or a functional expression  $f(t_1, \dots, t_n)$  where each  $t_i$  is a term. The wffs of  $\mathcal{L}$  are defined in the usual way. We shall use lower case Greek letters to refer to arbitrary wffs, and upper case Greek letters, possibly with subscripts, to refer to sets of classical wffs. The language of propositional classical logic is then defined as a sub-language of  $\mathcal{L}$  that consists only of the usual logical connectives and predicate symbols of arity 0 (i.e., 'propositional atomic formulae')  $\perp, a, b, c, \dots, z$ . We also assume the first order classical consequence relation  $\vdash_c$ , and write  $Cn(\Delta)$  to denote  $\{\alpha | \Delta \vdash_c \alpha\}$ . If  $Cn(\Delta) = \mathcal{L}$  we say that  $\Delta$  is inconsistent; else  $\Delta$  is consistent. We will make use of the following notions of a complement function and 'symbols' in a set of wff.

**Definition 3** (*Complement function*). Let  $\phi, \psi$  be classical wff. Then  $\overline{\phi} = \psi$  if  $\phi = \neg\psi$ ; else  $\overline{\phi} = \neg\phi$ .

**Definition 4** (*Symbols*). Let  $\Delta$  be a set of classical formulae. Then  $\text{symbols}(\Delta) = \{P | P \text{ is either a predicate or function symbol in } \Delta\}$ .

Note that in case  $\Delta$  is a set of propositional formulae,  $\text{symbols}(\Delta)$  consists of all the propositional atomic formulae in  $\Delta$ , whereas if  $\Delta$  is an arbitrary set of first order formulae,  $\text{symbols}(\Delta)$  consists of the predicate, function and constant symbols in  $\Delta$ . By defining symbols in arbitrary sets of first order formulae in this way, we are able to generalise the non-contamination postulates defined for propositional argumentation in [16], to the first order case (see later in Section 2.2).

Arguments are defined with respect to a set or 'base'  $\mathcal{B}$  of classical formulae, which we assume does not contain  $\perp$ ; in the sequel  $\perp$  will be exclusively used to express that an inconsistency has been reached when constructing an argument.



**Fig. 1.** i) Solid arrows are undermine attacks. If in addition to undermine attacks that target only premises, one also allowed attacks targeting conclusions, then one would also have the attacks represented by dotted arrows. In ii) and iii), arrows are defeats.

**Definition 5 (Base).** A base  $\mathcal{B}$  is a finite set of classical wff such that  $\wedge \notin \text{symbols}(\mathcal{B})$ .

**Definition 6 (Classical logic arguments).** [1,12,33]  $(\Delta, \phi)$  is an argument defined by a base  $\mathcal{B}$ , if  $\Delta \subseteq \mathcal{B}$ , and:

1.  $\Delta \vdash_c \phi$
2.  $Cn(\Delta) \neq \mathcal{L}$  ( $\Delta$  is consistent)
3.  $\neg \exists \Delta' \subset \Delta$  such that  $\Delta' \vdash_c \phi$  ( $\Delta$  is said to be ‘subset minimal’).

$\Delta$  and  $\phi$  are respectively referred to as the *premises* and *conclusion* of  $(\Delta, \phi)$ .

[33] studies propositional classical logic instantiations of *AFs*, identifying attack relations which ensure satisfaction of postulates closely related to [15]’s consistency postulates. For well-behaved attack relations, the premises of arguments in stable extensions of an *AF* instantiated by  $\mathcal{B}$  are simply the maximal (under set inclusion) consistent subsets of  $\mathcal{B}$ . Hence, [1,43] employ preferences over arguments to enable argumentation based arbitration of conflicts in a classical base.  $X$  successfully attacks (*defeats*)  $Y$  iff  $X$  attacks  $Y$  and  $Y$  is not strictly preferred to  $X$ . In this way, a defeat relation  $\mathcal{D} \subseteq \mathcal{C}$  is defined. The notions of conflict freeness, acceptability of arguments, and extensions are then defined exactly as in Definitions 1 and 2 above (with  $\mathcal{D}$  replacing  $\mathcal{C}$ ).

In [43], a general framework – *ASPIC*<sup>+</sup> (building on *ASPIC* [15]) – for instantiating *AFs* is defined. *ASPIC*<sup>+</sup> identifies properties of preference relations and logical instantiations of *AFs* that use strict (deductive) and defeasible inference rules, such that these properties ensure satisfaction of [15]’s consistency and closure rationality postulates. *ASPIC*<sup>+</sup> studies first order classical logic instantiations in which arguments are built from premises in a given base  $\mathcal{B}$ , and strict inference rules  $\{\phi_1, \dots, \phi_n \rightarrow \phi \mid \phi \in Cn(\{\phi_1, \dots, \phi_n\})\}$ . Hence arguments can be represented as in Definition 6, where an argument’s premises entail an argument’s conclusion through application of a single strict inference rule encoding the entailment condition 1 in Definition 6. (Note that the consistency check on an argument’s premises is enforced in [43], whereas the subset minimality check (condition 3 in Definition 6) is not enforced; we comment on this later.) Then:

**Definition 7 (Undermine attacks).**  $(\Delta, \phi)$  *undermine attacks*  $(\Delta', \phi')$  on  $\psi$  if  $\phi = \overline{\psi}$  for some  $\psi \in \Delta'$ .

Henceforth we may say that  $X$  ‘undermines’ rather than ‘undermine attacks’  $Y$ .

**Definition 8 (Elementary arguments).** Let  $\Delta$  be a set of classical wff. Then  $\forall \alpha \in \Delta$ ,  $(\{\alpha\}, \alpha)$  is said to be the ‘elementary argument’ defined by  $\alpha$ .

Undermine attacks only target premises, and as illustrated in the example below, can equivalently be said to target the ‘elementary arguments’ defined by such premises. Henceforth, we will for the sake of simplicity restrict ourselves to propositional examples.

**Example 1.** Let  $\mathcal{B}$  be a base  $\{a, b, \neg a \vee \neg b\}$ . A subset  $\mathcal{A}'$  of the arguments defined by  $\mathcal{B}$  are shown in Fig. 1i). Undermine attacks are indicated with solid arrows. For example,  $F$  attacks  $A$  on the premise  $a$  (i.e.,  $F$  attacks  $A$  on the elementary argument  $A$ ), and  $F$  attacks  $H$  on  $a$  (i.e.,  $F$  attacks  $H$  on  $A$ ).  $H$  attacks  $F$  on  $\neg a \vee \neg b$  (i.e., on  $C$ ), and  $G$  attacks  $H$  on  $b$  (i.e., on  $B$ ). There are three stable extensions – call them  $E_1, E_2$  and  $E_3$  – such that  $\{G, A, C\} \subseteq E_1$ ,  $\{F, B, C\} \subseteq E_2$

and  $\{H, A, B\} \subseteq E3$ . The premises (and conclusions) of arguments in each extension correspond to the maximal consistent subsets of  $\mathcal{B}$  (and their consequences). Hence, for example,  $(\Delta, \alpha) \in E3$  iff  $\alpha \in Cn(\{a, b\})$ .

ASPIC<sup>+</sup> [43] studies extensions of AFs consisting of arguments, and defeats obtained from attacks and a preference relation. It is then shown in [43] that [15]'s consistency and closure postulates are satisfied only if preferences are used to compare the attacker and the targeted elementary argument (premise).

**Definition 9 (Defeats).** Let  $(\mathcal{A}, \mathcal{C})$  be an AF where  $\mathcal{A}$  and  $\mathcal{C}$  are the arguments and undermine attacks defined by  $\mathcal{B}$ , and  $< \subseteq \mathcal{A} \times \mathcal{A}$  a strict partial ordering.

Let  $(Y, X) \in \mathcal{C}$ , where  $Y$  undermine attacks  $X$  on  $\psi$  (i.e., on  $X' = (\{\psi\}, \psi)$ ). Then  $(Y, X) \in \mathcal{D}$  ( $Y$  defeats  $X$ ) if  $Y \not< X'$ .

Modgil and Prakken [43] study specific preference relations. For example, given a partial ordering  $\leq$  over  $\mathcal{B}$  (with  $<$  and  $\approx$  defined in the usual way), then the 'Elitist' preference is defined as:

$$(\Gamma, \phi) <_E (\Delta, \theta) \text{ iff } \exists \alpha \in \Gamma \text{ such that } \forall \beta \in \Delta, \alpha < \beta \quad (\text{Elitist Preference})$$

**Example 2.** Referring to Fig. 1, suppose  $\neg a \vee \neg b < a \approx b$ . Then  $F <_E A$ ,  $G <_E B$ , and so  $F$  does not defeat  $A$ , or  $H$  or  $G$  (on  $A$ ), and  $G$  does not defeat  $B$ , or  $H$  or  $F$  (on  $B$ ). We obtain the AF with defeats in Fig. 1ii). This AF's single stable and preferred extension contains  $A, B, H$ . Suppose instead  $\neg a \vee \neg b < a$ . Then only  $F <_E A$ . Now  $G$ 's attacks on  $B, H$  and  $F$  succeed as defeats (Fig. 1iii)), and we have two stable and preferred extensions; one containing  $A, B$  and  $H$ , the other  $A, C$  and  $G$ .

Refs. [4,43,57] show that one can obtain an argumentation-based characterisation of non-monotonic inference defined by Brewka's Preferred Subtheories [13]. The latter starts with a *totally* ordered  $\mathcal{B}$  partitioned into equivalence classes  $(\mathcal{B}_1, \dots, \mathcal{B}_n)$  such that  $\forall \alpha \in \mathcal{B}_i, \forall \beta \in \mathcal{B}_j, i < j$  iff  $\beta < \alpha$ . A 'preferred subtheory' (*ps*) is obtained by taking a maximal (under set inclusion) consistent subset of  $\mathcal{B}_1$ , maximally extended with a subset of  $\mathcal{B}_2$ , and so on. Multiple *ps* may be constructed, and [43] shows that with undermine attacks defined over the arguments  $\mathcal{A}$  defined by  $\mathcal{B}$ , and the defeats  $\mathcal{D}$  obtained using the Elitist preference, then each *ps* corresponds to the premises of arguments in a stable extension of  $(\mathcal{A}, \mathcal{D})$ . Hence,  $\alpha$  is classically entailed from a *ps* iff  $\alpha$  is the conclusion of an argument in a stable extension. Then  $\alpha$  is a sceptical (credulous) *ps*-inference iff  $\alpha$  is entailed by all (respectively at least one) *ps*, iff  $\alpha$  is a sceptical (credulous) argumentation defined inference (i.e.,  $\alpha$  is the conclusion of an argument in all, respectively at least one, stable extension(s)). Preferences over arguments can of course be defined by criteria other than the Elitist comparison, and neither are preferences necessarily based on an ordering over  $\mathcal{B}$ . For example, preferences may be based on the relative importance of values promoted [9], or the goals realised [51], by actions whose justifying rationales are encoded in the arguments. Hence inference relations other than those corresponding to Preferred Subtheories can be defined by classical logic instantiations of AFs with preferences.

## 2.2. Rationality postulates for argumentation

The consistency and closure postulates [15] identify rational properties that one would want to hold of the complete (and hence grounded, preferred and stable) extensions of an instantiated AF. ASPIC [15] also identified sufficient conditions on logical instantiations which ensure satisfaction of these postulates. ASPIC<sup>+</sup> [43] then identified conditions for a broader range of instantiations while additionally accounting for the use of preferences. We now review these postulates as they apply to classical logic argumentation (sufficient conditions for satisfying these postulates will be reviewed in Section 3.2). Note that we will use the notation  $\text{conc}(E)$  to denote the set of conclusions of arguments in an extension  $E$ .

Recall that ASPIC and ASPIC<sup>+</sup> are general frameworks for defining instantiations of AFs. They assume an arbitrary language, and arguments are built from premises and (strict/defeasible) inference rules. An argument  $X$  is a tree whose root node is the argument's conclusion, the leaves are its premises, and any non-leaf node is the consequent  $\phi$  of an inference rule  $r$  where  $\phi'$  is a child node of  $\phi$  iff  $\phi'$  is a formula in the antecedent of  $r$ . Sub-arguments of  $X$  are identified as sub-trees of  $X$ , and the *sub-argument closure postulate* states that if  $X$  is in a complete extension  $E$  then all sub-arguments of  $X$  are in  $E$ . The standard representation of classical logic arguments consists only of premises paired with the entailed conclusion, and so a classical logic argument in ASPIC<sup>+</sup> applies a single strict inference rule to a set of premises (as described in Section 2.1), yielding a tree whose root node – the argument's conclusion – has child leaf nodes that are the premises. Thus, the sub-trees are simply the premises in the leaf nodes, and the sub-arguments are the elementary arguments associated with premises. Hence for *Cl-Arg*<sup>2</sup>:

– *Closure under sub-arguments* amounts to stating that (in the remainder of this section we assume complete extensions  $E, E' \dots$ ):

<sup>2</sup> This rendition of the postulate for classical logic argumentation is due to the 'schematic' representation of classical logic arguments as premises paired with a conclusion. As we suggest later, arguments 'proper' should include the proof theoretic means whereby one entails the conclusion from the premises, in which case the postulate would then require that all subproofs of a proof be included in a complete extension.

Let  $X = (\Delta, \phi) \in E$ . If  $\alpha \in \Delta$  then  $(\{\alpha\}, \alpha) \in E$

– *Closure under strict rules* states that if a strict rule with consequent  $\phi$  can be applied to the conclusions of arguments in a complete extension, then there is an argument in that extension that concludes  $\phi$ . For *Cl-Arg*:

If  $\text{conc}(E) \vdash_c \phi$ , then  $\phi \in \text{conc}(E)$

– *Direct* and *indirect consistency* respectively state that the conclusions of arguments in a complete extension, and their closure under strict rules, do not conflict:

$\forall \alpha, \beta \in \text{conc}(E) : \alpha \neq \bar{\beta}$  and  $\forall \alpha, \beta \in \text{Cn}(\text{conc}(E)) : \alpha \neq \bar{\beta}$

Clearly, *closure under strict rules* and *direct consistency* imply *indirect consistency*.

More recently, two further rationality postulates were proposed for instantiations of AFs based on propositional languages [16]. We introduce notation that (together with Definition 4) allows us to generalise these postulates to first order instantiations. This generalization can be seen as a modest contribution to the specification of rationality postulates for logic-based argumentation.

**Notation 3.** Let  $Sy$  denote a set of symbols (recall Definition 4).<sup>3</sup> Then  $\mathcal{B}_{|Sy} = \{\alpha \in \mathcal{B} \mid \text{symbols}(\{\alpha\}) \subseteq Sy\}$ , e.g.,  $\{\neg a \vee \neg b, c \wedge a\}_{|a,b} = \{\neg a \vee \neg b\}$  and  $\{\forall x P(x) \rightarrow Q(x), \exists z P(z) \wedge R(z)\}_{|P,Q} = \{\forall x P(x) \rightarrow Q(x)\}$ . Also,  $\mathcal{B} \parallel \mathcal{B}'$  denotes that  $\text{symbols}(\mathcal{B}) \cap \text{symbols}(\mathcal{B}') = \emptyset$ , and  $\mathcal{B}$  and  $\mathcal{B}'$  are said to be *syntactically disjoint*. Hence two sets of first order formulae are syntactically disjoint if they share no predicate or function (including constant) symbols (in the propositional case this equates to no shared propositional atomic formulae).

– *Non-interference* states that for any  $\mathcal{B}$  composed of two independent (i.e., syntactically disjoint) bases  $\mathcal{B}_1$  and  $\mathcal{B}_2$ , neither should ‘influence’ each other’s argumentation defined inferences. That is to say, given frameworks  $AF$  and  $AF'$  instantiated respectively by  $\mathcal{B}_1$  and  $\mathcal{B}_1 \cup \mathcal{B}_2$ , where  $\mathcal{B}_1 \parallel \mathcal{B}_2$ , then:

$E$  is an extension of  $AF$  iff there is an extension  $E'$  of  $AF'$ , where  $\text{conc}(E)_{| \text{symbols}(\mathcal{B}_1)} = \text{conc}(E')_{| \text{symbols}(\mathcal{B}_1)}$

– *Crash resistance* states that there is no ‘contaminating’  $\mathcal{B}_1$  that when merged with an unrelated  $\mathcal{B}_2$ , makes the latter irrelevant. That is to say, given frameworks  $AF$  and  $AF'$  instantiated respectively by  $\mathcal{B}_1$  and  $\mathcal{B}_1 \cup \mathcal{B}_2$ , where  $\mathcal{B}_1 \parallel \mathcal{B}_2$ , then:

$\{\text{conc}(E) \mid E \text{ is an extension of } AF\} \neq \{\text{conc}(E) \mid E \text{ is an extension of } AF'\}$

### 3. Dialectical classical logic argumentation

In this section we propose practical desiderata for applications of logical instantiations of Dung’s theory. We also discuss why features of standard approaches to *Cl-Arg* preclude satisfaction of these desiderata, while at the same time being proposed as sufficient conditions for satisfaction of the rationality postulates. We then formalise a provably rational account of *Cl-Arg* that satisfies practical desiderata.

#### 3.1. Applications and practical desiderata

Section 2.1 reviewed the definition of argumentation-based inference relations over logical (in particular classical logic) bases instantiating AFs. Formalisations of ‘instantiated argumentation’ have thus been widely advocated for individual agent reasoning over uncertain and conflicting beliefs (e.g., [15,39]) and decision making in which practical arguments for alternative actions attack each other. In the latter case preferences may be based on the prioritisation of agent goals or the values promoted by actions, and are used to arbitrate amongst these practical arguments (e.g., [3,8,39,51]).

Argumentation’s dialectical characterisation of non-monotonic inference provides for generalisation of the above ‘monological’ applications to *dialogues* involving both computational and human agents. Argumentation-based dialogues enable distributed reasoning amongst agents communicating to persuade one another of the truth of a proposition, decide amongst alternative action options, or negotiate allocations of resources (e.g., [2,38,45,48,58]). Agents submit locutions consisting of arguments and counter-arguments, as well as locutions that implicitly define arguments (e.g., when an agent asserts a claim, and subsequently asserts supporting reasons in response to another agent challenging the claim). The contents of locutions are incrementally included in a public *commitment store*  $\mathcal{B}_p$  [60], and agents can submit arguments constructed from their private distinct bases and  $\mathcal{B}_p$ . At any point in the dialogue, an agent can be said to ‘win the dialogue’ – i.e., successfully establish the dialogue ‘topic’  $\alpha$  which may be a belief or decision option – by countering arguments moved by

<sup>3</sup> In [16], the term ‘atoms’ is used instead of symbols. We use the term symbols as we generalise to the first order case.



his interlocutor, effectively showing (under the assumption of 'logically perfect' play, as discussed in the example below) that an argument  $X$  claiming  $\alpha$  is in an admissible extension  $E$  of the  $AF$  instantiated by  $\mathcal{B}_p$  [31,48]. These dialogues can be seen as generalising argument game proof theories (e.g. [42]) for monological applications in which an agent shows that a given argument  $X$  is justified by successfully countering defeating arguments that the agent can construct from his base (in this case the base, and hence instantiated  $AF$ , is assumed fixed rather than being incrementally updated). As in these argument games, dialogue rules regulating legal moves can then be varied to establish that the admissible  $E$  is a subset of an extension under different semantics.<sup>4</sup>

**Example 4.** Consider Example 1. Suppose  $Ag_1$  submits  $A$  constructed from its own base  $\mathcal{B}_1$ .  $A$  is countered by  $Ag_2$  submitting  $F$ , constructed from her base  $\mathcal{B}_2$ . At this point  $Ag_2$  is winning the dialogue. However note that  $A$  is in an admissible extension of the  $AF$  instantiated by the committed premises  $a, b, \neg a \vee \neg b$ . The agents have not played in a 'logically perfect' way [48]:  $Ag_1$  can construct and submit  $G$ .<sup>5</sup>

Another key reason for the impact of Dung's theory is that evaluation of justified arguments is essentially based on the intuitive and familiar 'reinstatement principle' whereby an argument is defended if all its defeaters are themselves defeated. The theory's compatibility with human modes of reasoning and debate has prompted empirical studies of human uses of the reinstatement principle [52], and development of theories and applications supporting both computational and human reasoning and dialogue [44]. For example, computational models of dialogue for normatively guiding human–human [58] and human–computer [37] dialogue. Hence, real-world modes of dialectical reasoning have been formalised,<sup>6</sup> often drawing on insights from philosophical and informal logic. For example, the notion of an 'audience' [46] to whom an argument is addressed, has been modelled in extensions to Dung's theory [10] and in making classical logic arguments more believable [35]. Dung's theory has also been extended to accommodate other real-world uses of argument, such as arguments supporting arguments (see [40] for a review) and the weighing up of arguments for and against claims (accrual) [41,49]. The above monological and dialogical applications involving computational and/or human agents, warrant consideration of *two key practical desiderata* for formalisations of argumentation:

1. One needs to account for the fact that real-world computational and human agents are resource-bounded.
2. Formalisations should be compatible with real-world modes of dialectical reasoning.

We now review the extent to which standard features of  $CI\text{-}Arg$  preclude satisfaction of these practical desiderata.

### 3.2. Practicality versus rationality

Both [1], [33], and  $ASPIC^+$ 's formalisation of  $CI\text{-}Arg$  with preferences [43], assume instantiation of an  $AF$  by *all arguments* defined by a base  $\mathcal{B}$  (which will in general be infinite).<sup>7</sup> This is clearly unfeasible for agents with limited resources, given the undecidability of first order classical logic, and that even in the propositional case, deciding whether  $\Delta \vdash_c \alpha$  is in general NP-hard, and therefore most likely intractable. Moreover, for each argument constructed, one must not only check consistency of the premises but also that they are subset-minimal; a problem in the second level of the polynomial hierarchy [30], so that in the worst case, for every constructed argument  $(\Delta, \alpha)$  one needs to verify that  $\forall \Delta' \subset \Delta, \Delta' \not\vdash_c \alpha$ . Hence, to accommodate uses by resource-bounded agents, one needs to study the outcomes of argument evaluation assuming  $AF$ s that may include only a subset of the arguments defined by a base (which henceforth we refer to as *partially instantiated*  $AF$ s) and avoid enforcing checks on the legitimacy of arguments before inclusion in an  $AF$ .

Standard formalisations of  $CI\text{-}Arg$  are also incompatible with real-world modes of dialectical reasoning. Firstly, in both monological and dialogical applications in which an agent attempts to establish whether an argument  $X$  is justified, agents may simply not construct all arguments defined by a base (be it an individual agent's base or one incrementally constructed by the dialogue). Agents pragmatically operate on the basis of beliefs or chosen actions that are provisionally acceptable, until later seeking further counter-arguments from a given base, or from a base that is subsequently augmented with newly acquired information. Moreover, it may not be necessary to construct all arguments, given that in argument game proof theories and dialogues, it may suffice to show membership of  $X$  in an admissible extension (and hence an extension

<sup>4</sup> In [42], variations in rules yield games for membership under grounded, preferred and stable semantics, and a game is proposed for showing sceptical justification under preferred semantics.

<sup>5</sup> Whether  $Ag_2$  is declared the winner depends on the context in which the dialogue model is deployed. One might argue that the onus is on  $Ag_1$  to use  $Ag_2$ 's premise  $\neg a \vee \neg b$  to construct  $G$  and counter  $F$ . Alternatively, a 'dialogue manager' [58] might be deployed to prompt the agents to submit arguments that can be constructed, so ensuring (and possibly enforcing) logically perfect play.

<sup>6</sup> See the series of workshops [www.cmna.info/CMNA14](http://www.cmna.info/CMNA14).

<sup>7</sup> In fact, although not considered in these works, it would suffice (under the assumption that undermine attacks are used) to consider a possibly finite sub-framework  $AF'$  of an infinite  $AF$ , such that for each  $\alpha$  in a base  $\mathcal{B}$  one includes the corresponding elementary argument and constructs and includes in  $AF'$  all arguments that claim  $\alpha$ . Then the justified elementary arguments in  $AF'$  will fully determine the status of any argument  $X = (\Delta, \gamma)$  defined by  $\mathcal{B}$  in the infinite  $AF$ ; if  $\forall \alpha \in \Delta, ((\alpha), \alpha)$  is justified, then  $(\Delta, \gamma)$  is justified. However, this would still be unfeasible for bases of any significant size, and more so given the required checks on the legitimacy of arguments.

under other semantics depending on the rules regulating legal moves). In other words, in practice agents are not logically perfect for pragmatic reasons other than that they are resource limited, further testifying to the need for studying partially instantiated *AFs*.

Secondly, establishing whether an individual argument's premises are consistent, is a special case of establishing the mutual consistency of a set of arguments' premises, and this is addressed dialectically in real-world dialogues. For example, an often used dialectical move illustrated in the *Socratic Elenchus* [59] involves showing that an opponent's arguments *collectively* make use of inconsistent premises, by constructing an argument from these premises to show that the opponent contradicts himself.

Thirdly, accounts of *Cl-Arg* typically make no reference to specific proof theories for constructing arguments, and so the subset minimality check is imposed as a somewhat blunt instrument for ensuring that no premises are redundant with respect to deriving the argument's conclusion.<sup>8</sup> However, in practice agents do not interrogate premises for subset minimality. Rather, the exclusion of trivially redundant premises is determined by the proof theoretic means used to derive the conclusion from the premises.

Fourthly, the exclusive targeting of premises by defeating arguments leads to the so called 'foreign commitment' problem in dialogues. Agents may be forced to commit to the truth of interlocutors' premises for which they have insufficient reasons to believe (see [17] for a concrete example of this problem), and may be subsequently obligated to defend [60]. Consider Example 4. After *A* is attacked by *F*, *Ag*<sub>1</sub> cannot counter with *A*, as attacks cannot target conclusions of arguments. *Ag*<sub>1</sub> has to counter *F* with either *G* or *H*, and so commit to *Ag*<sub>2</sub>'s premise  $\neg a \vee \neg b$ , respectively *b*. Suppose then that another agent *Ag*<sub>3</sub> were to attack *Ag*<sub>1</sub>'s argument *H* on *B*, requiring that *Ag*<sub>1</sub> now defend *Ag*<sub>2</sub>'s argument *B*.

We have argued that standard features of *Cl-Arg* preclude its suitability for use by real-world resource-bounded agents. However, dispensing with these features may result in violation of the rationality postulates. Dropping the consistency check on premises may lead to violation of *non-interference* and *crash-resistance*. Consider the syntactically disjoint  $\mathcal{B}_1 = \{s\}$  and  $\mathcal{B}_2 = \{p, \neg p\}$ . Assuming all attacks succeed as defeats, then  $(\{s\}, s)$  is in the grounded extension of the *AF* based on  $\mathcal{B}_1$ . However  $\emptyset$  is the grounded extension of the *AF'* based on  $\mathcal{B}_1 \cup \mathcal{B}_2$ , since in *AF'* every argument is defeated by arguments  $(\{p, \neg p\}, \phi)$  ( $\phi \in \mathcal{L}$ ),<sup>9</sup> and it is well known that the grounded extension is empty if there are no un-defeated arguments. Hence *non-interference* is violated. Moreover,  $\mathcal{B}_2$  is contaminating as the grounded extension of the framework based on  $\mathcal{B}_2$  is  $\emptyset$ , and for any  $\mathcal{B}_1$  (e.g.,  $\mathcal{B}_1 = \{s\}$ )  $\emptyset$  is the grounded extension of the *AF'* based on  $\mathcal{B}_1 \cup \mathcal{B}_2$ . Hence crash resistance is violated. Thus, arguments with inconsistent premises are said to be *contaminating* as they may attack arguments with syntactically unrelated premises.

As for the subset minimality check, we show in Section 3.7 that allowing non-subset minimal arguments may also result in violation of *non-interference*. That this is the case is suggested by a result shown for *ASPIC*<sup>+</sup>'s formulation of *Cl-Arg* [43]. The authors show that if one were to allow non-subset minimal arguments, then the argumentation defined inferences remain unchanged only if one assumes preference relations that do not strengthen subset minimal arguments by adding premises. That is to say, if *A* is subset minimal and a non-subset minimal *A'* is obtained by adding premises to the support of *A*, then  $A < B$  implies  $A' < B$ . (Intuitively, one would not want the strength of an argument to depend on premises that are redundantly used in deriving the conclusion.) Indeed, by definition, the *Elitist* preference (see Section 2) satisfies this property. Thus one can dispense with subset minimality checks. But for preference relations that do not satisfy this property (e.g., the *Democratic* preference relation studied in [43]), one needs to identify subset minimal arguments and impose this constraint on the preference relation. However this would not be necessary if one can identify a notion of non-redundancy enforceable by proof theories for constructing arguments, and such that the proof theoretic exclusion of arguments that use redundant premises suffices to ensure satisfaction of *non-interference* and *crash resistance* independently of the preference relation used. (We revisit this issue in Section 4.)

The rationality postulates are shown to hold [15,43] under the assumption that *all* arguments defined by a base are included in an *AF*. Consider the following example uses of argumentation. Suppose that in attempting to defend an argument, an agent constructs arguments  $P = (\{p, p \rightarrow \neg q\}, \neg q)$  and  $Q = (\{p, p \rightarrow q\}, q)$ . To ensure that no complete extension *E* can include *P* and *Q* and so violate direct consistency (i.e., that no such irrational defense is possible), one needs to assume that the *AF*, by virtue of including all defined arguments, includes the arguments *X*, *Y* and *Z*, where:

- $X = (\{p, p \rightarrow q\}, \neg(p \rightarrow \neg q))$  attacks  $A = (\{p \rightarrow \neg q\}, p \rightarrow \neg q)$ ;
- $Y = (\{p, p \rightarrow \neg q\}, \neg(p \rightarrow q))$  attacks  $B = (\{p \rightarrow q\}, p \rightarrow q)$ ;
- $Z = (\{p \rightarrow q, p \rightarrow \neg q\}, \neg p)$  attacks  $C = (\{p\}, p)$ .

Then [43] shows that *direct consistency* is satisfied under the assumption that the preference relation used is 'reasonable' ([43] shows that the *Elitist* preference relation is reasonable), which in this example means that either  $X \not\prec A$  or  $Y \not\prec B$  or  $Z \not\prec C$ . Now, suppose  $X < A$  and  $Y < B$  and  $<$  is reasonable. Then it must be that  $Z \not\prec C$ . This means that the agent cannot

<sup>8</sup> This can be contrasted with the weaker condition that all premises are 'used' (as opposed to necessary) to obtain the conclusion. There may well be inferences from non-subset minimal sets of premises in which all premises are used, e.g., three applications of modus ponens deriving *q* from *p*,  $p \rightarrow r$ ,  $p \rightarrow q$ ,  $r \rightarrow ((p \rightarrow q) \rightarrow q)$ .

<sup>9</sup> In particular  $(\{s\}, s)$  is defeated by an argument  $(\{p, \neg p\}, \neg s)$  that concludes a conclusion syntactically disjoint from the premises via the explosivity of the inconsistent premises.



use  $P$  and  $Q$  in its attempted defense, since  $Z$  will defeat both  $P$  and  $Q$  on  $C$ , and so to defend against  $Z$ , will mean including in its attempted defense (i.e.,  $E$ ), some  $V \in E$  that defeats  $Z$ . But any such  $V$  must then defeat  $P \in E$  or  $Q \in E$ , contradicting the admissibility of  $E$ .

However, if the agent does not construct all arguments then direct consistency may be violated. For example, resources may not suffice to construct  $Z$ ; indeed, in the resource bounded proof theory presented in Section 4, we show that construction of  $Z$  assumes a higher bound on inferential capabilities than construction of  $X$  or  $Y$ .

Consider also a variation of the above example in which an agent  $Ag_1$ , in the course of a persuasion dialogue with  $Ag_2$ , submits arguments that make use of premises  $p$ ,  $p \rightarrow q$  and  $p \rightarrow \neg q$  (without as yet having constructed and submitted arguments  $P$  and  $Q$ ). Again, one must assume a reasonable preference relation and that  $Ag_2$  can construct from these premises the arguments  $X$ ,  $Y$  and  $Z$ , to prevent  $Ag_1$  winning the dialogue by forcing  $Ag_1$  to move some  $V$  that defeats one of  $Ag_1$ 's own arguments (i.e.,  $Ag_2$  effectively shows that  $Ag_1$  is indirectly inconsistent).

We therefore want that the consistency postulates are satisfied by *partially instantiated AFs*. Moreover, one would intuitively expect that having explicitly recognised the inconsistency of premises (e.g.,  $p, p \rightarrow q, p \rightarrow \neg q$ ) in an extension  $E$ , through construction of arguments with conflicting conclusions (e.g.,  $P, Q$ ) then this should suffice to ensure that  $E$  cannot be rationally accepted as a defensible set of arguments, without needing to assume construction of the additional arguments  $X$ ,  $Y$  and  $Z$ , or *any* properties of preference relations.<sup>10</sup> Also note that the *closure under strict rules* postulate also assumes construction of all arguments defined by a base, and so assumes an idealised notion of an agent that is clearly unrealistic in its assumption of *logical omniscience* [56]. We therefore later modify this closure postulate to account for partially instantiated AFs.

Finally, suppose one were to avoid the foreign commitment problem by allowing attacks on conclusions, as indicated by the dotted arrows in Fig. 1. Then  $E' = \{A, B, C\}$  is now admissible and so a subset of a complete extension  $E$ . However  $E$  cannot be closed (neither  $F$ ,  $G$  or  $H$  can be in  $E$  since  $E$  would not then be conflict free). Also  $E$  would be indirectly inconsistent.

We now formalise an account of *Cl-Arg* that drops the consistency and subset minimality checks on arguments, formalises the dialectical demonstration that the premises of an argument, or set of arguments, is inconsistent, and solves the foreign commitment problem. We then show that despite dropping checks on the legitimacy of arguments, and making only minimal assumptions as to the arguments included in a partially instantiated AF, the closure (we will modify closure under strict rules to account for partially instantiated frameworks) and consistency postulates are satisfied assuming *any* preference relation over arguments. We also show that the non-interference and contamination postulates, generalised here so as to apply to first order instantiations of AFs, are satisfied under the assumption that preference relations are 'coherent'.

### 3.3. Dialectical frameworks and acceptability

We begin by observing that interlocutors in dialogues typically distinguish their own premises, namely those that they accept as true, from the premises that their opponent commits to and that they want to criticise: "on the basis of the premises I regard to be true, and supposing for the sake of argument what you regard to be true, then I can show some conclusion that contradicts one of your premises" or "I can show that your premises are inconsistent". This pattern is pervasive in real argumentation practice,<sup>11</sup> and motivates a dialectical 'ontology' for classical logic arguments that explicitly differentiates between an argument's assumptions that are premises regarded as true, and those supposed true for the sake of argument (the argument's *suppositions*). We also drop the subset minimality and consistency checks. Hence, arguments with inconsistent assumptions can now conclude  $\perp$ .

**Definition 10** (*Dialectical arguments*).  $X = (\Delta, \Gamma, \alpha)$  is a *dialectical argument* defined by a base  $\mathcal{B}$ , if  $(\Delta \cup \Gamma) \subseteq \mathcal{B}$ ,  $\Delta \cap \Gamma = \emptyset$ , and  $\Delta \cup \Gamma \vdash_c \alpha$ .

If  $\alpha = \perp$  then  $X$  is said to be a *falsum* argument. If  $\Gamma = \emptyset$  then  $X$  is said to be *unconditional*; else  $X$  is *conditional*. Finally, if  $\Delta = \emptyset$  then  $X$  is said to be *unassailable* (since  $X$  cannot then be attacked).

**Notation 5.** Let  $X = (\Delta, \Gamma, \alpha)$  be a dialectical argument,  $S$  a set of dialectical arguments.

- $\Delta$ ,  $\Gamma$  and  $\alpha$  are, respectively,  $X$ 's premises, suppositions and conclusion, denoted  $\text{prem}(X)$ ,  $\text{suppositions}(X)$  and  $\text{conclusion}(X)$ . Also,  $\text{assumptions}(X)$  denotes  $\text{prem}(X) \cup \text{suppositions}(X)$ .
- Generalising the above notation to sets of arguments in the obvious way:  $\text{prem}(S) = \bigcup_{X \in S} \text{prem}(X)$ ;  $\text{suppositions}(S) = \bigcup_{X \in S} \text{suppositions}(X)$ ;  $\text{assumptions}(S) = \text{prem}(S) \cup \text{suppositions}(S)$ ;  $\text{conclusion}(S) = \bigcup_{X \in S} \text{conclusion}(X)$ .

<sup>10</sup> It should be obvious that in real-world argumentation, an agent recognises that he is committing to inconsistent premises the moment he constructs arguments with conflicting conclusions.

<sup>11</sup> For example, in his *Dialogues Concerning Two New Sciences* [32], Galileo presents a famous refutation of Aristotle's theory of falling bodies, in the form of a dialogue between their respective alter-egos Salviati and Simplicio. Salviati demonstrates that the premises of Simplicio's arguments justifying that heavier bodies fall faster than lighter bodies, lead to a contradiction (i.e., are inconsistent).

**Example 6.** The dialectical arguments  $\mathcal{A}$  defined by  $\mathcal{B} = \{a, b, \neg a \vee \neg b\}$  include:

$A_1 = (\{a\}, \emptyset, a)$	$B_1 = (\{b\}, \emptyset, b)$	$C_1 = (\{\neg a \vee \neg b\}, \emptyset, \neg a \vee \neg b)$
$F_1 = (\{b, \neg a \vee \neg b\}, \emptyset, \neg a)$	$G_1 = (\{a, \neg a \vee \neg b\}, \emptyset, \neg b)$	$H_1 = (\{a, b\}, \emptyset, \neg(\neg a \vee \neg b))$
$F_2 = (\{b\}, \{\neg a \vee \neg b\}, \neg a)$	$G_2 = (\{a\}, \{\neg a \vee \neg b\}, \neg b)$	$H_2 = (\{a\}, \{b\}, \neg(\neg a \vee \neg b))$
$X_1 = (\emptyset, \{a, b, \neg a \vee \neg b\}, \wedge)$	$X_2 = (\{a, b, \neg a \vee \neg b\}, \emptyset, c)$	$X_3 = (\{b, \neg a \vee \neg b\}, \{a\}, \wedge)$
$X_4 = (\{a\}, \{b, \neg a \vee \neg b\}, \wedge)$	$Z_1 = (\{a, b, \neg a \vee \neg b\}, \emptyset, \neg a)$	$Z_2 = (\{b, \neg a \vee \neg b\}, \{a\}, \neg a)$

As in [43], we interpret an attack between two arguments as *declaratively* denoting that the conclusion of the attacking argument conflicts with some element in the attacked argument. In particular, we here formalise undermine attacks, denoting that the conclusion of the attacking argument conflicts with an assumption in the attacked argument that is a premise regarded as true. Preferences are then applied to determine the defeat relation, which is interpreted *dialectically*, in that it denotes the use of the defeating argument as a counter-argument to the defeated argument. We will then utilise the distinction between premises and suppositions when accounting for the dialectical use of defeats in determining the acceptability of arguments. Intuitively, when an agent constructs a candidate set of admissible arguments  $S$ , the agent commits to the premises of arguments in  $S$ . Any of these premises can then be supposed true when challenging the acceptability of some  $Y$  w.r.t.  $S$ . That is to say, when moving some  $X$  as a defeat on  $Y$ ,  $X$ 's suppositions can include premises of *any* arguments in  $S$ . In turn, when defending  $Y$  by some  $Z \in S$  that defeats  $X$ ,  $Z$ 's suppositions can include any of  $X$ 's premises. In the latter case,  $Z$  supposes for the sake of argument only  $X$ 's premises, since in order to invalidate the admissibility of  $S$ , it suffices that only one such defeat by some  $X$  on  $Y \in S$  be undefended (rather than having to construct some candidate set of arguments  $S'$  that includes  $X$ ). Note that a defeat by a falsum argument  $X = (\Delta, \Gamma, \wedge)$  on  $Y$  indicates that under the assumption that the premises  $\Delta$  is true, the 'supposed premises' (i.e., suppositions)  $\Gamma$  in  $S$  are inconsistent, and  $Y$  in particular is targeted since at least one of the suppositions in  $\Gamma$  is also a premise of  $Y$ .

**Definition 11 (Attacks).** Let  $\mathcal{A}$  be a set of dialectical arguments defined by  $\mathcal{B}$ . The attack relation  $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$  is defined as follows. For any  $X = (\Delta, \Gamma, \alpha)$ ,  $Y = (\Pi, \Sigma, \beta) \in \mathcal{A}$ :  $(X, Y) \in \mathcal{C}$  (which we may denote by writing  $X \rightarrow Y$ ) iff exactly one of the following holds:

- $\alpha \neq \wedge$  and  $\bar{\alpha} \in \Pi$  (we say that  $X$  attacks  $Y$  on  $\bar{\alpha}$ , equivalently on  $Y' = (\{\bar{\alpha}\}, \emptyset, \bar{\alpha})$ );
- $\alpha = \wedge$  and  $\Gamma \cap \Pi \neq \emptyset$  (we say that  $X$  attacks  $Y$  on any  $\phi \in \Gamma \cap \Pi$ , equivalently on any  $Y' = (\{\phi\}, \emptyset, \phi)$ ).

Notice that Definition 11 adapts the notion of an undermine attack (Definition 7) in the sense that the premises of an attacked argument are targeted by the attacking argument. We thus also adapt the term 'elementary argument' (Definition 8) to refer to  $Y' = (\{\bar{\alpha}\}, \emptyset, \bar{\alpha})$  in the above definition, and say that  $X$  is equivalently said to attack  $Y$  on  $Y'$  (recall the comment after Definition 8).

Notice that by definition, no argument in any set  $S$  can be attacked with *unconditional* falsum arguments  $(\Delta, \emptyset, \wedge)$ . Indeed, it would be insensible to allow such attacks, since an agent would be irrationally attacking with an argument that she herself recognises as based on inconsistent premises.<sup>12</sup>

We now define the defeat relation. Firstly, recall that the distinction between  $\Delta$  and  $\Gamma$  in  $X = (\Delta, \Gamma, \phi)$  is an *epistemic* distinction between assumptions one commits to as being true and supposed true for the sake of argument. Logically,  $X$  is equivalent to any  $X' = (\Delta', \Gamma', \phi)$  where  $\Delta' \cup \Gamma' = \Delta \cup \Gamma$ . Hence  $G_1$  and  $G_2$  in Example 6 are logically equivalent. We consider the strength (and hence relative preference) of an argument to be independent of whether assumptions are regarded as true or supposed true. To illustrate, suppose the *Elitist* preference relation (Section 2.1) defined on the basis of the well known temporal ordering over rules in the legal domain, such that later rules override earlier rules [50]. Suppose such an ordering on conflicting legal rules:  $a \rightarrow b <_E a \rightarrow \neg b$ , so that  $X = (\{a, a \rightarrow b\}, \emptyset, \neg(a \rightarrow \neg b))$  does not defeat  $Y = (\{a \rightarrow \neg b\}, \emptyset, a \rightarrow \neg b)$  since  $X <_E Y$ . In dialogical applications, the only distinction between  $X$  and  $X' = (\{a\}, \{a \rightarrow b\}, \neg(a \rightarrow \neg b))$  is that in  $X'$   $a \rightarrow b$  is a premise that an interlocutor commits to. (Clearly, no such distinction arises in monological applications.) Thus, one still uses the temporal ordering to conclude that  $X' <_E Y$ ; the fact that  $a \rightarrow b$  is a premise used by an interlocutor is irrelevant to defining the temporal based preference. Henceforth, we therefore assume preferences over arguments are *invariant modulo logical equivalence*:

**Definition 12 (Logically equivalent arguments).** Let  $\mathcal{A}$  be the dialectical arguments defined by  $\mathcal{B}$  and  $<$  a strict partial ordering over  $\mathcal{A}$ . Let  $X = (\Delta, \Gamma, \alpha) \in \mathcal{A}$ . Then:

- $[X] = \{X' = (\Delta', \Gamma', \alpha) \mid \Delta' \cup \Gamma' = \Delta \cup \Gamma\}$ .  
 $\forall Y, Z \in [X]$  we say that  $Y$  and  $Z$  are logically equivalent.
- $<$  is said to be *invariant modulo logical equivalence* (imle) if  $\forall Y' \in [Y], \forall X' \in [X] : Y < X$  iff  $Y' < X'$ .

<sup>12</sup> Note that allowing such attacks would not pose theoretical problems, since  $(\Delta, \emptyset, \wedge)$  would be counter-attacked by  $(\emptyset, \Delta, \wedge)$  which by virtue of having empty premises (and so being un-attackable) would be acceptable w.r.t. any set  $E$ .

In the case that  $X = (\emptyset, \Gamma, \lambda)$  attacks  $Y$ , the attack dialectically demonstrates that the proponent of  $Y$  commits to the inconsistent premises  $\Gamma$ , either because  $\Gamma \subseteq \text{prem}(Y)$ , or  $\Gamma \subseteq \text{prem}(Y) \cup \text{prem}(E)$  when challenging the acceptability of  $Y$  w.r.t.  $E$ . It would clearly be incoherent to use preferences to reject such an attack.

**Definition 13** (*Defeats*). Let  $\mathcal{A}$  be a set of dialectical arguments,  $\mathcal{C}$  the attack relation, and  $<$  a strict partial ordering, over  $\mathcal{A}$ . Then  $\forall (X, Y) \in \mathcal{C}$ ,  $(X, Y) \in \mathcal{D}$  ( $X$  defeats  $Y$ ) iff exactly one of the following holds:

- $X$  is an argument of the form  $(\emptyset, \Gamma, \lambda)$ <sup>13</sup>;
- $\exists \alpha \in \text{prem}(X)$  s.t.  $Y \rightarrow X$  on  $\alpha$ , and  $Y \not\prec (\{\alpha\}, \emptyset, \alpha)$ .

We may write  $Y \Rightarrow X$  to denote  $(Y, X) \in \mathcal{D}$ , and  $Y \not\Rightarrow X$  to denote  $(Y, X) \notin \mathcal{D}$ .

Recall that we are interested in *AFs* that do not necessarily include all arguments defined by a base. Hence, we define the notion of a *dialectical AF* that includes any subset of the arguments defined by a base. We will then identify (in this section and in Section 3.7) sufficient conditions on the arguments that must be included in order to ensure satisfaction of the rationality postulates.

**Definition 14** (*Dialectical AF instantiated by a base*). Let  $\mathcal{A}$  be any subset of the set of dialectical arguments defined by a base  $\mathcal{B}$ , and  $\mathcal{D}$  the defeat relation defined by the attack relation  $\mathcal{C}$ , and a strict partial ordering  $<$  over  $\mathcal{A}$ . Then  $(\mathcal{A}, \mathcal{D})$  is a dialectical *AF* defined by  $\mathcal{B}$  and  $<$ .

Recall that the suppositions of an argument  $X$  can suppose for the sake of argument the premises of arguments in  $S$  when challenging the acceptability of some  $Y$  w.r.t.  $S$ . Then  $Z \in S$  can suppose the premises of  $X$  when defeating  $X$  in order to defend  $Y$ .

**Definition 15** (*Dung semantics for dialectical AFs*). Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical *AF*, and  $S \subseteq \mathcal{A}$ ,  $X, Y \in \mathcal{A}$ . Then:

- $X$  defeats  $Y$  with respect to  $S$ , denoted  $X \Rightarrow_S Y$ , if  $(X, Y) \in \mathcal{D}$  and  $\text{suppositions}(X) \subseteq \text{prem}(S \cup \{Y\})$ .
- $S$  is conflict free if  $\forall Z, Y \in S$ ,  $Z \not\Rightarrow_S Y$ .
- $Y$  is acceptable w.r.t.  $S$  if  $\forall X$  s.t.  $X \Rightarrow_S Y$ ,<sup>14</sup>  $\exists Z \in S$  s.t.  $Z \Rightarrow_{\{X\}} X$ .
- Let  $S$  be conflict free. Then  $S$  is: an *admissible* extension iff  $X \in S$  implies  $X$  is acceptable w.r.t.  $S$ ; a *complete* extension iff  $S$  is admissible and if  $X$  is acceptable w.r.t.  $S$  then  $X \in S$ ; a *preferred* extension iff it is a set inclusion maximal complete extension; the grounded extension iff it is the set inclusion minimal complete extension; a stable extension iff  $\forall Y \notin S$ ,  $\exists X \in S$  s.t.  $X \Rightarrow_{\{Y\}} Y$ .
- For  $T \in \{\text{complete, preferred, grounded, stable}\}$ ,  $X$  is *sceptically*, respectively *credulously* justified under the  $T$  semantics if  $X$  belongs to all, respectively at least one,  $T$  extension.
- $\mathcal{F}_{(\mathcal{A}, \mathcal{D})}(S) = \{X \mid X \text{ is acceptable w.r.t. } S\}$ .

Notice that every stable extension of a dialectical *AF* is a preferred (and hence complete) extension. This result is shown in exactly the same way as for Dung *AFs*.<sup>15</sup>

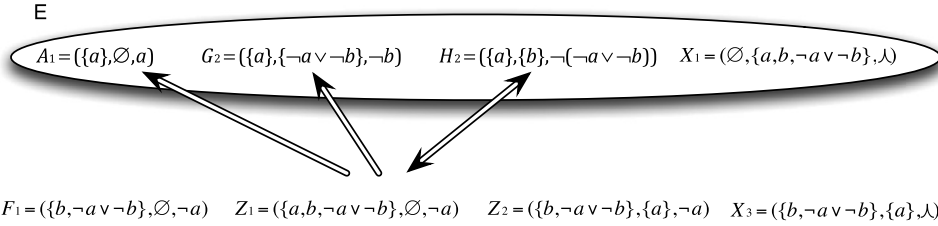
In Section 2.1, the argumentation defined inferences are obtained by detaching the conclusions of arguments in extensions. However, the extensions of dialectical *AFs* may contain conditional arguments that suppose the truth of assumptions when used dialectically. Once the extensions are defined, we detach only the conclusions (which we call ‘claims’) of *unconditional* arguments all of whose assumptions are premises presumed true.

**Definition 16** (*Conclusions (‘claims’) of an extension*). Let  $E$  be an extension of a dialectical *AF*. Then  $\text{claims}(E) = \{\phi \mid (\Delta, \emptyset, \phi) \in E\}$ .

<sup>13</sup> We later consider ‘coherent’ preference relations s.t.  $\forall \alpha \in \Gamma$  it cannot be that  $(\emptyset, \Gamma, \lambda) < (\{\alpha\}, \emptyset, \alpha)$ . Then, even if we don’t impose that attacks from such arguments always succeed as defeats, coherent preference relations guarantee that such attacks always succeed as a defeat on at least one premise in  $\Gamma$ .

<sup>14</sup> Of course, in practice, when submitting a defeating argument  $X$ , one need only reference the set  $S' \subseteq S$  such that  $S'$  are the arguments in  $S$  whose premises are referenced by  $\text{suppositions}(X)$ .

<sup>15</sup> Suppose  $E$  is stable,  $X \in E$  and  $Y \Rightarrow_E X$ . If  $Y \in E$  then this contradicts  $E$  is conflict free. If  $Y \notin E$ , then  $Z \in E$ ,  $Z \Rightarrow_{\{Y\}} Y$ , and so  $X$  is acceptable w.r.t.  $E$ . Moreover, if  $X \notin E$ , then  $X$  cannot be acceptable w.r.t.  $E$  since  $\exists Y \in E$  s.t.  $Y \Rightarrow_{\{X\}} X$ . Hence if  $X$  were acceptable w.r.t.  $E$  then  $Z \in E$ ,  $Z \Rightarrow_{\{Y\}} Y$  and so  $Z \Rightarrow_E Y$ , contradicting  $E$  is conflict free. Hence  $E$  is complete and must be maximal complete (i.e., preferred) since any  $E' \supset E$  cannot be conflict free.



**Fig. 2.** All arguments outside of  $E$  defeat  $A_1$ ,  $G_2$  and  $H_2$ .  $H_2$  defeats all arguments outside of  $E$ . Hence  $E$  is admissible.

The dialectical ontology for arguments solves the foreign commitment problem described in Section 3.2. Recall that  $Ag_1$  submits  $A$  and  $Ag_2$  counters with  $F$ . Then to defend  $A$ ,  $Ag_1$  must defeat  $F$  with either  $G$  or  $H$ , respectively committing to  $Ag_2$ 's premise  $b$  or  $\neg a \vee \neg b$  in  $Ag_2$ 's argument  $F$ . Now suppose the following continuation of Example 6.

**Example 7.** Consider  $E$  in Fig. 2. Suppose  $\forall V \in \{F_1, Z_1, Z_2, X_3\}$ ,  $V \not\Rightarrow_E A_1$ . Then for each  $V$  we have that  $V \Rightarrow_E A_1$ , and  $V \Rightarrow_E G_2$  on  $A_1$ ,  $V \Rightarrow_E H_2$  on  $A_1$ . Suppose also that  $G_1 < B_1$ ,  $H_1 \not< C_1$ , and since  $<$  is *imle*,  $G_2 < B_1$  and  $H_2 \not< C_1$ . Hence  $\forall V \in \{F_1, X_3, Z_1, Z_2\}$ ,  $G_2 \not\Rightarrow_{\{V\}} V$  and  $H_2 \Rightarrow_{\{V\}} V$  (on  $C_1$ ).

$E$  is admissible, despite  $Ag_1$  not having to commit to either  $\neg a \vee \neg b$  or  $b$  as premises, since  $G_2$  and  $H_2$  respectively include  $\neg a \vee \neg b$  and  $b$  as suppositions.  $H_2$  defends against each of  $V$ 's defeats on members of  $E$  (where  $H_2$ 's assumption  $b$  is only supposed true for the sake of argument).

**Remark 8.** Suppose a dialectical  $AF$  containing exactly the arguments in Example 6. No admissible extension  $E$  can contain the arguments  $X_2$  and  $Z_1$ , both of whose premises are inconsistent, since  $X_1 \Rightarrow_E X_2$  ( $X_1 \Rightarrow_E Z_1$ ), and the unassailable  $X_1$  cannot be defeated by virtue of its premises being empty. Hence  $X_2$  ( $Z_1$ ) cannot be defended and so cannot be acceptable w.r.t.  $E$ . Moreover, no *admissible* (and not just complete) extension  $E$  can contain arguments with conflicting conclusions. Suppose  $E$  contains  $A_1$  and  $F_1$ . But then  $X_1 \Rightarrow_E A_1$  on  $A_1$ , and  $X_1 \Rightarrow_E F_1$  on both  $B_1$  and  $C_1$ , indicating that the proponent of arguments in  $E$  has committed to inconsistent premises. No argument in  $E$  can defend against  $X_1$ , and so  $E$  cannot be admissible.

In contrast with Example 1, which assumes the standard ontology for arguments, Remark 8 highlights that one does not require that all arguments (i.e.,  $G_1$  and  $H_1$ ) can be constructed in order to guarantee satisfaction of the direct consistency postulate. It suffices that the assumptions of arguments with conflicting conclusions (i.e.,  $A_1$  and  $F_1$ ) be combined to yield a falsum argument (i.e.,  $X_1$ ), which we argue is a less demanding, and indeed more intuitive assumption as to the arguments that an agent can construct.

In general, if the inconsistency of a set of premises is recognised by constructing at least two arguments from these premises with conflicting conclusions, then this guarantees consistency of extensions under the assumption that one combines their assumptions to yield an argument concluding  $\lambda$ . Of course, if either of the two arguments is constructed from inconsistent premises, and this is recognisable through construction of an argument from those premises that concludes  $\lambda$ , then it is not necessary to assume that one combines their assumptions. This points to properties of any set  $\mathcal{A}$  of arguments in a dialectical  $AF$  that suffice to ensure satisfaction of the consistency and closure postulates:

**Definition 17** (Properties of sets of arguments in dialectical  $AF$ s). Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical  $AF$ . We define properties that  $\mathcal{A}$  may satisfy:

- (P1)  $\forall X \in \mathcal{A}$ :  $\alpha \in \text{prem}(X)$  implies  $(\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$ .
- (P2)  $\forall X \in \mathcal{A}$ :  $X' \in [X]$  implies  $X' \in \mathcal{A}$ .
- (P3) If  $(\Delta, \emptyset, \alpha) \in \mathcal{A}$  and  $(\Gamma, \emptyset, \bar{\alpha}) \in \mathcal{A}$ , then either  $(\Delta, \emptyset, \lambda) \in \mathcal{A}$  or  $(\Gamma, \emptyset, \lambda) \in \mathcal{A}$  or  $(\Delta \cup \Gamma, \emptyset, \lambda) \in \mathcal{A}$ .

We believe that requiring that  $\mathcal{A}$  satisfies P1, P2 and P3 would impose minimally restrictive assumptions as to the arguments that agents should be able to construct. P3 is discussed above. P1 is a kind of reflexivity property, and states that for any premise  $\alpha$  of an argument in  $\mathcal{A}$ , the corresponding 'elementary' argument  $(\{\alpha\}, \emptyset, \alpha)$  is also in  $\mathcal{A}$ . P2 says that if  $X \in \mathcal{A}$ , then  $\mathcal{A}$  contains all logically equivalent arguments that are distinct only in the epistemic distinction between premises and suppositions. Note that P2 does not imply use of additional resources in *constructing* arguments. Rather, given a constructed argument  $(\Delta, \Gamma, \alpha)$ , the practical meaning of P2 is that any differential labelling of the assumptions in  $\Delta \cup \Gamma$  is accessible to an agent when using the argument dialectically. Of course there is a distinct issue with respect to the computational effort required to list all logically equivalent arguments (which for each given argument are exponential in the number of assumptions). However, we suggest that in practice this will not be necessary. An interesting topic for future work would be to identify (given an  $AF$  instantiated by dialectical arguments moved in monological and dialogical applications) the minimal set of logically equivalent arguments that need to be listed to ensure satisfaction of the rationality postulates (e.g., given arguments  $(\Delta, \Gamma, \alpha)$  and  $(\Delta', \Gamma', \neg\alpha)$ , it may suffice to list only  $(\Gamma \cup \Gamma', \Delta \cup \Delta', \lambda)$ ).

### 3.4. Closure and consistency postulates

We prove closure and consistency postulates for dialectical AFs that include any subset of the arguments defined by a base, subject to the above assumptions P1–P3. We first show the following lemmas:

**Lemma 9.** *Let  $E$  be a complete extension of a dialectical AF  $(\mathcal{A}, \mathcal{D})$  where  $\mathcal{A}$  satisfies P1. Then  $\forall X \in \mathcal{A}$  such that  $\text{prem}(X) \subseteq \text{prem}(E)$ ,  $X \in E$ .*

**Proof.** By definition, a complete extension  $E$  is an admissible (and hence conflict free) set such that all arguments acceptable w.r.t.  $E$  are in  $E$ . Hence we need to show that  $X$  is acceptable w.r.t.  $E$  and  $E \cup \{X\}$  is conflict free.

– Suppose  $Y \Rightarrow_E X$  on some  $X' = (\{\alpha\}, \emptyset, \alpha)$ ,  $Y = (\Delta, \Gamma, \phi)$  ( $\phi = \perp$  or  $\phi = \bar{\alpha}$ ).<sup>16</sup> Since  $\text{prem}(X) \subseteq \text{prem}(E)$ , then  $\Gamma \subseteq \text{prem}(E)$ . Since  $\alpha \in \text{prem}(X) \subseteq \text{prem}(E)$ ,  $\exists X'' \in E$  s.t.  $Y \Rightarrow_E X''$  on  $X'$ . Since  $E$  is complete  $\exists Z \in E$  s.t.  $Z \Rightarrow_{\{Y\}} Y$ . Hence  $X$  is acceptable w.r.t.  $E$ .

– Suppose for contradiction that  $E \cup \{X\}$  is not conflict free. Then either:

1.  $Y \in E$ ,  $Z \in E$ ,  $Y \neq X$ ,  $Z \neq X$ , and  $Y \Rightarrow_{E \cup \{X\}} Z$ . Hence  $Y \Rightarrow_E Z$ , contradicting  $E$  is complete and so conflict free.
2.  $Y \in E$  and  $Y \Rightarrow_{E \cup \{X\}} X$ . Hence  $Y \Rightarrow_E X$ . Since  $X$  acceptable w.r.t.  $E$ ,  $\exists Z \in E$ ,  $Z \Rightarrow_{\{Y\}} Y$ . Hence  $Z \Rightarrow_E Y$ , contradicting  $E$  is complete and so conflict free.
3.  $Y \in E$  and  $X \Rightarrow_{E \cup \{X\}} Y$ . Hence  $X \Rightarrow_E Y$ . Since  $E$  is complete,  $\exists Z \in E$ ,  $Z \Rightarrow_{\{X\}} X$ . Hence  $Z \Rightarrow_E X$ . Since  $X$  is acceptable w.r.t.  $E$ ,  $\exists W \in E$ ,  $W \Rightarrow_{\{Z\}} Z$  and so  $W \Rightarrow_E Z$ , contradicting  $E$  is complete and so conflict free.
4.  $X \Rightarrow_{E \cup \{X\}} X$ . Hence  $X \Rightarrow_E X$ . Since  $X$  acceptable w.r.t.  $E$ ,  $\exists Y \in E$ ,  $Y \Rightarrow_{\{X\}} X$ . Hence  $Y \Rightarrow_E X$ , and so  $\exists Z \in E$ ,  $Z \Rightarrow_{\{Y\}} Y$  and  $Z \Rightarrow_E Y$ , contradicting  $E$  is complete and so conflict free.  $\square$

Recall that for *CI-Arg*, the sub-argument postulate states that if  $X$  is in a complete extension  $E$  then all the elementary arguments associated with its premises are in  $E$ .

**Theorem 10** (Sub-argument closure). *Let  $E$  be a complete extension of a dialectical AF  $(\mathcal{A}, \mathcal{D})$  such that  $\mathcal{A}$  satisfies P1. Let  $X \in E$ . Then if  $\alpha \in \text{prem}(X)$  then  $(\{\alpha\}, \emptyset, \alpha) \in E$ .*

**Proof.** By P1,  $X' = (\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$ . Since  $X \in E$ ,  $\text{prem}(X') \subseteq \text{prem}(E)$ . By Lemma 9,  $X' \in E$ .  $\square$

Direct consistency (see Section 2.2) states that no arguments in a complete extension have conflicting conclusions. Since we allow arguments concluding  $\perp$ , we show that  $\text{claims}(E)$  (recall that these are the conclusions of unconditional arguments that make no suppositions for the sake of argument) neither contains  $\perp$  nor conflicting conclusions, *for the more general case of admissible extensions*.

**Theorem 11** (Direct consistency). *Let  $E$  be an admissible extension of a dialectical AF  $(\mathcal{A}, \mathcal{D})$ . If  $\mathcal{A}$  satisfies P1, P2 and P3, then  $\forall \alpha, \beta \in \text{claims}(E)$ ,  $\alpha \neq \perp$  and  $\beta \neq \bar{\alpha}$ .*

**Proof.** *Case 1:* Suppose  $X = (\Delta, \emptyset, \perp) \in E$ . By P2,  $X' = (\emptyset, \Delta, \perp) \in \mathcal{A}$ .  $X'$  attacks  $X$  on every  $\alpha \in \Delta$ . Recall (Definition 13) that since  $X'$  is an unassailable falsum argument, every such attack succeeds as a defeat. Moreover,  $X'$  cannot be defeated, contradicting  $X$  is acceptable w.r.t.  $E$ .

*Case 2:*  $X, Y \in E$ ,  $X = (\Delta, \emptyset, \alpha)$ ,  $Y = (\Gamma, \emptyset, \bar{\alpha})$ . If either  $X' = (\Delta, \emptyset, \perp) \in \mathcal{A}$  or  $Y' = (\Gamma, \emptyset, \perp) \in \mathcal{A}$ , then (given P2) either  $(\emptyset, \Delta, \perp) \Rightarrow_E X$  or  $(\emptyset, \Gamma, \perp) \Rightarrow_E Y$ , contradicting  $X$  ( $Y$ ) is acceptable w.r.t.  $E$  as in case 1). Else, by P3,  $Z = (\Delta \cup \Gamma, \emptyset, \perp) \in \mathcal{A}$ . By P2,  $Z' = (\emptyset, \Delta \cup \Gamma, \perp) \in \mathcal{A}$ . Since  $\Delta \cup \Gamma \subseteq \text{prem}(E)$ ,  $Z' \Rightarrow_E X$  (on all  $\delta \in \Delta$ ) and  $Z' \Rightarrow_E Y$  (on all  $\gamma \in \Gamma$ ). Since  $Z'$  is unassailable, this contradicts  $X, Y$  acceptable w.r.t.  $E$ .  $\square$

Satisfaction of direct consistency for admissible (and not just complete) extensions is a desirable property for both monological and dialogical applications. As discussed in Sections 3.1 (see also footnote 4) and 3.2, in both application types it often suffices that an agent identify an admissible extension  $E$  containing the argument  $X$  whose status is to be established. Clearly, agents do not then proceed to include all arguments that can be defended by  $E$  so as to define a complete extension. Hence one would want that direct consistency holds for admissible extensions. Of course, in real-world applications an agent may simply refrain from submitting any arguments with conflicting conclusions, whereas an agent may unwittingly commit to arguments whose premises are mutually inconsistent (e.g., the example in Section 3.2 in which  $\text{Ag}_1$  commits to premises  $p$ ,  $p \rightarrow q$  and  $p \rightarrow \neg q$ , and the dialogue referred to in footnote 11). However, no admissible  $E$  can contain arguments with premises whose inconsistency is recognised by the construction (from those premises) of an argument concluding  $\perp$ , or a pair of arguments with directly conflicting conclusions.

<sup>16</sup> Henceforth we may simply write  $\phi$ , leaving it to the reader to interpret  $\phi$  as denoting  $\perp$  or  $\bar{\alpha}$ .

**Theorem 12** (Premise consistency). *Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF such that  $\mathcal{A}$  satisfies P2. If for some  $\Delta \subseteq \text{prem}(E) : (\Delta, \emptyset, \lambda) \in \mathcal{A}$ , then  $E$  cannot be an admissible extension of  $(\mathcal{A}, \mathcal{D})$ .*

**Proof.** Suppose for contradiction that  $\exists \Delta \subseteq \text{prem}(E)$  s.t.  $(\Delta, \emptyset, \lambda) \in \mathcal{A}$ , and  $E$  is admissible. By P2,  $Z = (\emptyset, \Delta, \lambda) \in \mathcal{A}$ , and so for all  $\beta \in \Delta$ ,  $\exists X \in E$  s.t.  $\beta \in \text{prem}(X)$  and  $Z \Rightarrow_E X$ , contradicting  $E$  is admissible.  $\square$

Note of course that under the assumption that P3 holds, a sufficient condition for satisfaction of premise consistency would be that for some  $\Delta, \Gamma \subseteq \text{prem}(E) : \exists (\Delta, \emptyset, \alpha) \in \mathcal{A}$  and  $\exists (\Gamma, \emptyset, \neg\alpha) \in \mathcal{A}$ .

We now adapt the closure under strict rules postulate to account for the fact that it may be that not all arguments defined by a base are included in a dialectical AF. In general, even though  $\text{claims}(E) \vdash_c \alpha$ , it may not be that there exists an  $X \in E$  such that  $X$  concludes  $\alpha$ , given that agents are not logically omniscient and do not construct all arguments from a base (e.g., because of insufficient resources). However, if  $\alpha$  is entailed by the claims of arguments  $E' \subseteq E$ , and there is an argument  $X$  that, from the premises in  $E'$ , concludes  $\alpha$ , then  $X$  is in  $E$ .

**Theorem 13** (Closure under strict rules for dialectical AFs). *Let  $E$  be a complete extension of  $(\mathcal{A}, \mathcal{D})$ , where  $\mathcal{A}$  satisfies P1. Let  $E' \subseteq E$  and  $\text{claims}(E') \vdash_c \alpha$ . If there exists an  $X = (\Delta, \emptyset, \alpha) \in \mathcal{A}$  such that  $\Delta = \text{prem}(E')$ , then  $X \in E$ .*

**Proof.** Follows immediately from Lemma 9.  $\square$

We have shown satisfaction of the consistency and (modified) closure postulates under the assumption that agents only construct possibly finite subsets of the set of all arguments defined by a base (provided the subsets satisfy P1, P2 and P3). Moreover, unlike the  $\text{ASPIC}^+$  framework [43], the postulates have been shown to hold without making any assumptions as to the properties of the preference relation used to determine whether attacks succeed as defeats (recall that  $\text{ASPIC}^+$  assumes ‘reasonable’ preference relations that satisfy a number of properties; see Section 3.2), and both consistency of conclusions and premises is shown for admissible extensions.

### 3.5. The fundamental lemma and monotonicity of the characteristic function

Two key properties of Dung AFs are the *Fundamental Lemma* (FL) [26, p. 7] – if  $X$  and  $X'$  are acceptable w.r.t. an admissible  $E$ , then  $E \cup \{X\}$  is admissible and  $X'$  is acceptable w.r.t.  $E \cup \{X\}$  – and monotonicity of an AF’s characteristic function [26, p. 9]:  $E \subseteq E'$  implies  $\mathcal{F}(E) \subseteq \mathcal{F}(E')$ . The FL implies that any admissible extension is a subset of a preferred extension. Monotonicity of  $\mathcal{F}$  allows one to show that  $\mathcal{F}$  has a unique least fixed point that is the grounded extension, and that iterating  $\mathcal{F}$  (starting with  $\emptyset$ ) yields the grounded extension.

Both these properties can straightforwardly be shown for Dung AFs, since when determining the acceptability of  $X$  w.r.t.  $E$ , the defeats on  $X$  are independent of the set  $E$  under consideration. However for dialectical AFs, the defeats on  $X$  w.r.t.  $E$  may be a subset of the defeats on  $X$  w.r.t.  $E' \supset E$  (since the additional premises committed to in  $E'$  can now be suppositions in arguments that did not defeat  $X$  w.r.t.  $E$ ). However, these properties can be shown by focusing on sets of arguments that are ‘epistemically maximal’. Intuitively, if  $X = (\Delta, \Gamma, \alpha) \in E$ , and some supposition  $\beta \in \Gamma$  is also a premise of an argument in  $E$ , then commitment to the truth of the premise  $\beta$  also implies a commitment to the logically equivalent  $X' = (\Delta \cup \{\beta\}, \Gamma \setminus \{\beta\}, \alpha)$ .

**Definition 18** (Epistemically maximal sets). *Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF. Then  $E \subseteq \mathcal{A}$  is epistemically maximal (em) iff:*

*If  $X = (\Delta, \Gamma, \alpha) \in E$ ,  $\Gamma' \subseteq (\Gamma \cap \text{prem}(E))$ , then  $X' = (\Delta \cup \Gamma', \Gamma \setminus \Gamma', \alpha) \in E$ .*

*The function  $Cl_{em} : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$  maps any  $E$  to its epistemically maximal set:*

*$Cl_{em}(E) = E \cup \{X' | X \in E, X' \in [X], \text{prem}(X) \subseteq \text{prem}(X'), \text{prem}(X') \subseteq \text{prem}(E)\}$*

We now show some straightforward results (note that when referring to the characteristic function  $\mathcal{F}_{(\mathcal{A}, \mathcal{D})}$  we will omit the subscript). Henceforth we assume that the arguments  $\mathcal{A}$  in a dialectical AF satisfy P1, P2 and P3 (Definition 17).

**Lemma 14.** *Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF. Then  $\forall S \subseteq \mathcal{A} : Cl_{em}(S) \subseteq \mathcal{A}$ .*

**Proof.** The result follows immediately from P2, which states that if  $X \in \mathcal{A}$  then all arguments logically equivalent to  $X$  are in  $\mathcal{A}$ .  $\square$

**Lemma 15.** *Let  $\mathcal{F}$  be the characteristic function of  $(\mathcal{A}, \mathcal{D})$ , and  $E \subseteq \mathcal{A}$  a fixed point of  $\mathcal{F}$  (i.e.,  $E = \mathcal{F}(E)$ ). Then  $E$  is epistemically maximal.*

**Proof.** Suppose for contradiction that  $E$  is not em, hence  $\exists X' \in Cl_{em}(E) \setminus E$ . Since  $\text{prem}(X') \subseteq \text{prem}(E)$  and  $E$  is complete,  $X' \in E$  by Lemma 9. Contradiction.  $\square$



**Lemma 16.** Let  $E' = Cl_{em}(E)$ , and  $\forall X \in E$ ,  $X$  is acceptable w.r.t.  $E'$ . Then  $\forall Y \in E' \setminus E$ ,  $Y$  is acceptable w.r.t.  $E'$ .

**Proof.** Suppose  $Y \in E' \setminus E$ ,  $Z \Rightarrow_{E'} Y$  on  $\alpha$ . Hence  $Z \not\vdash (\{\alpha\}, \emptyset, \alpha)$ . Since  $E' = Cl_{em}(E)$ , then  $\text{prem}(E') = \text{prem}(E)$ ,  $\alpha \in \text{prem}(E)$ , and so  $\exists X \in E$ ,  $\alpha \in \text{prem}(X)$ , and  $Z \Rightarrow_{E'} X$  on  $\alpha$ . By assumption of  $X$  acceptable w.r.t.  $E'$ ,  $\exists Q \in E'$ ,  $Q \Rightarrow_{\{Z\}} Z$ . Hence  $Y$  is acceptable w.r.t.  $E'$ .  $\square$

### 3.5.1. Proving a variant of the fundamental lemma and implied properties

We now present results for use in proving a variant of the fundamental lemma.

**Lemma 17.** Let  $E \subseteq \mathcal{A}$  such that every argument in  $E$  is acceptable w.r.t.  $E$ , and  $\mathcal{A}$  satisfies P1, P2 and P3. Then  $E$  is conflict free.

**Proof.** Suppose for contradiction that  $X, Y \in E$ ,  $X \Rightarrow_E Y$  on  $\alpha$ , where  $X = (\Delta, \Gamma, \phi)$ ,  $\phi = \lambda$  or  $\phi = \bar{\alpha}$ .

1) Suppose  $\phi = \lambda$ . Then (by Definition 11) it must be that  $\Gamma \neq \emptyset$ . By P2,  $Z = (\emptyset, \Delta \cup \Gamma, \lambda) \in \mathcal{A}$ . Since  $\Delta \cup \Gamma \subseteq \text{prem}(E)$ , then  $\forall \beta \in \Delta \cup \Gamma$ ,  $\exists W \in E$  s.t.  $Z \Rightarrow_E W$  on  $\beta$ , contradicting  $W$  is acceptable w.r.t.  $E$ .

2) Suppose  $\phi = \bar{\alpha}$ . By P2,  $X' = (\Delta \cup \Gamma, \emptyset, \bar{\alpha}) \in \mathcal{A}$ . By P1  $(\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$ . By P3, either:

- $(\Delta \cup \Gamma, \emptyset, \lambda) \in \mathcal{A}$  (in which case  $\Delta \cup \Gamma \neq \emptyset$ ) and  $\Delta \cup \Gamma \subseteq \text{prem}(E)$ , so that  $\forall \beta \in \Delta \cup \Gamma$ ,  $\exists W \in E$  s.t.  $(\emptyset, \Delta \cup \Gamma, \lambda) \Rightarrow_E W$  on  $\beta$ , contradicting  $W$  is acceptable w.r.t.  $E$ , or
- $(\{\alpha\}, \emptyset, \lambda) \in \mathcal{A}$ , and since  $\alpha \in \text{prem}(E)$ ,  $(\emptyset, \{\alpha\}, \lambda) \Rightarrow_E Y$  on  $\alpha$ , contradicting  $Y$  is acceptable w.r.t.  $E$ , or
- $(\Delta \cup \Gamma \cup \{\alpha\}, \lambda) \in \mathcal{A}$ . By P2,  $Z = (\emptyset, \Delta \cup \Gamma \cup \{\alpha\}, \lambda) \in \mathcal{A}$ . Since  $\Delta \cup \Gamma \cup \{\alpha\} \subseteq \text{prem}(E)$ ,  $Z \Rightarrow_E Y$  on  $\alpha$  and  $\forall \beta \in \Delta \cup \Gamma$ ,  $\exists W \in E$  s.t.  $Z \Rightarrow_E W$  on  $\beta$ , contradicting  $Y, W$  acceptable w.r.t.  $E$ .  $\square$

**Lemma 18.** Let  $E' \subseteq \mathcal{A}$ ,  $Z = (\Delta, \Gamma, \phi)$  and  $Z \Rightarrow_{E'} Y$  on  $\alpha$ . Let  $\Gamma' = \Gamma \cap \text{prem}(E \cup \{Y\})$  where  $E \subseteq E'$ . Then  $Z' = (\Delta \cup (\Gamma \setminus \Gamma'), \Gamma', \phi) \in \mathcal{A}$  and  $Z' \Rightarrow_E Y$ .

**Proof.** By P2,  $Z' \in \mathcal{A}$ . Since  $Z \Rightarrow_{E'} Y$  on  $\alpha$  and  $\prec$  is imle:  $Z' \not\vdash (\{\alpha\}, \emptyset, \alpha)$ .

Suppose  $\phi = \bar{\alpha}$ . By definition,  $\Gamma' \subseteq \text{prem}(E \cup \{Y\})$ . Hence  $Z' \Rightarrow_E Y$ .

Suppose  $\phi = \lambda$ . Then  $\alpha \in \Gamma$ , and by definition  $\alpha \in \Gamma'$ . Since  $\Gamma' \subseteq \text{prem}(E \cup \{Y\})$ ,  $Z' \Rightarrow_E X$ .  $\square$

We now prove a variant of the fundamental lemma:

**Lemma 19.** Let  $X, X'$  be acceptable w.r.t. an admissible extension  $E$  of  $(\mathcal{A}, \mathcal{D})$ . Then: 1)  $Cl_{em}(E \cup \{X\})$  is admissible, and 2)  $X'$  is acceptable w.r.t.  $Cl_{em}(E \cup \{X\})$ .

**Proof.** Let  $E' = E \cup \{X\}$  and  $E'' = Cl_{em}(E')$ . Note that  $\text{prem}(E') = \text{prem}(E'')$ .

*Proof of 1).* We show that every argument in  $E''$  is acceptable w.r.t.  $E''$ .

We first show that every argument in  $E'$  is acceptable w.r.t.  $E''$ . Suppose for contradiction that  $Y \in E'$  is not acceptable w.r.t.  $E''$ . Then  $\exists Z = (\Delta, \Gamma, \phi)$  ( $\phi = \lambda$  or  $\bar{\alpha}$ ) s.t.  $Z \Rightarrow_{E'} Y$  on  $Y' = (\{\alpha\}, \emptyset, \alpha)$  (hence  $Z \not\vdash Y'$ ), and:

$$\neg \exists Q \in E'' \text{ s.t. } Q \Rightarrow_{\{Z\}} Z \quad (1)$$

Suppose  $Y = X$ . Since  $\text{prem}(E \cup \{X\}) = \text{prem}(E') = \text{prem}(E'')$ ,  $Z \Rightarrow_E X$ . By assumption of  $X$  acceptable w.r.t.  $E$ ,  $\exists Q \in E$  (and so  $Q \in E''$ )  $Q \Rightarrow_{\{Z\}} Z$ , contradicting Eq. (1).

Suppose  $Y \neq X$ . Let  $\Gamma' = \Gamma \cap \text{prem}(E \cup \{Y\})$  (and so  $\Gamma \setminus \Gamma'$  are premises of  $X$  that are not premises in  $E$  or  $Y$ ). By Lemma 18,  $Z' = (\Delta \cup (\Gamma \setminus \Gamma'), \Gamma', \phi)$ ,  $Z' \Rightarrow_E Y$ .

By assumption of  $E$  admissible and so every  $Y$  in  $E$  acceptable w.r.t.  $E$ :

$$\exists W = (\Pi, \Sigma, \psi) \in E \text{ s.t. } W \Rightarrow_{\{Z'\}} Z' (\psi = \lambda \text{ or } \psi = \bar{\beta}) \quad (2)$$

on  $\beta \in \Delta \cup (\Gamma \setminus \Gamma')$ , where  $W \not\vdash (\{\beta\}, \emptyset, \beta)$ ,  $\Sigma \subseteq \Delta \cup (\Gamma \setminus \Gamma')$ . Consider two cases:

1. Suppose  $\beta \in (\Gamma \setminus \Gamma')$  (in which case  $\beta \in \text{prem}(X)$  and if  $\psi = \lambda$  then  $\beta \in \Sigma \cap (\Gamma \setminus \Gamma')$ ).

Let  $W'$  be the logically equivalent  $(\Sigma \cap \Delta, \Pi \cup (\Sigma \cap (\Gamma \setminus \Gamma')), \psi)$ . Since  $\Pi \subseteq \text{prem}(E)$  and  $(\Sigma \cap (\Gamma \setminus \Gamma')) \subseteq \text{prem}(X)$ , and  $W' \not\vdash (\{\beta\}, \emptyset, \beta)$ , then  $W' \Rightarrow_E X$ . By assumption of the acceptability of  $X$  w.r.t.  $E$ ,  $\exists Q \in E$  s.t.  $Q \Rightarrow_{\{W'\}} W'$ . Since  $\text{prem}(W') \subseteq \text{prem}(Z)$ ,  $Q \Rightarrow_{\{Z\}} Z$ , contradicting Eq. (1).

2. Suppose  $\beta \in \Delta$ , (hence  $\beta \in \text{prem}(Z)$  and if  $\psi = \lambda$  then  $\beta \in \Sigma \cap \Delta$ ). Let  $W'$  be the logically equivalent  $(\Pi \cup (\Sigma \cap (\Gamma \setminus \Gamma')), \Sigma \cap \Delta, \psi)$ . Since  $\text{prem}(W') \subseteq \text{prem}(E')$ , then  $W'$  is in the epistemic closure  $E''$  of  $E'$ . Since  $W' \not\vdash (\{\beta\}, \emptyset, \beta)$ ,  $W' \Rightarrow_{\{Z\}} Z$ , contradicting Eq. (1).

We have shown  $\forall Y \in E', Y'$  is acceptable w.r.t.  $E''$ . By Lemma 16, all arguments in  $E''$  are acceptable w.r.t.  $E''$ . By Lemma 17,  $E''$  is conflict free. Hence  $E'' = Cl_{em}(E')$  is admissible.

*Proof of 2)* Suppose for contradiction that  $X'$  is not acceptable w.r.t.  $E''$ . Then  $Z \Rightarrow_{E''} X'$  on  $(\{\alpha\}, \emptyset, \alpha)$ ,  $Z \not\vdash (\{\alpha\}, \emptyset, \alpha)$  and  $\neg \exists Q \in E''$  s.t.  $Q \Rightarrow_{\{Z\}} Z$ .

Let  $Z = (\Delta, \Gamma, \phi)$  ( $\phi = \perp$  or  $\overline{\alpha}$ ),  $\Gamma' = \Gamma \cap \text{prem}(E \cup \{X'\})$  (and so  $\Gamma \setminus \Gamma'$  are premises in  $X$  that are not premises in  $E$  or  $X'$ ). By Lemma 18,  $Z' = (\Delta \cup (\Gamma \setminus \Gamma'), \Gamma', \phi)$ ,  $Z' \Rightarrow_E X'$ .

By assumption of the acceptability of  $X'$  w.r.t.  $E$ ,  $\exists W = (\Pi, \Sigma, \psi) \in E$ ,  $W \Rightarrow_{\{Z'\}} Z'$  on  $(\{\beta\}, \emptyset, \beta)$ , where  $W \not\vdash (\{\beta\}, \emptyset, \beta)$  and  $\Sigma \subseteq \Delta \cup (\Gamma \setminus \Gamma')$ . We can now reason as in 1) above, contradicting  $\neg \exists Q \in E''$  s.t.  $Q \Rightarrow_{\{Z\}} Z$ .  $\square$

Assuming the admissible set  $E$  in Lemma 19 is epistemically maximal, it immediately follows that:

**Corollary 20.** *The set of all epistemically maximal admissible extensions of a dialectical AF form a complete partial order w.r.t. set inclusion.*

**Lemma 21.** *Let  $E$  be an admissible extension of  $(\mathcal{A}, \mathcal{D})$ . Then  $E \subseteq \mathcal{F}(E)$  and  $Cl_{em}(\mathcal{F}(E))$  is admissible.*

**Proof.** By assumption of the admissibility of  $E$ ,  $E \subseteq \mathcal{F}(E)$ . To show  $Cl_{em}(\mathcal{F}(E))$  is admissible, we make use of the function:  $F^*(E, X) = Cl_{em}(E \cup \{X\})$ .

Let  $\mathcal{F}(E) \setminus E$  be the possibly countably infinite set  $\{X_1, \dots, X_n, \dots\}$ , and let:

$$F_1^* = F^*(E, X_1), F_2^* = F^*(F_1^*, X_2), \dots, F_i^* = F^*(F_{i-1}^*, X_i), \dots$$

That is to say, the sequence obtained by adding, one by one, the arguments  $X_1, \dots, X_n, \dots$ , whereupon after each argument is added, the set is closed to obtain an *em* set. One can straightforwardly show that:

$$\forall i : Cl_{em}(E \cup \{X_1, \dots, X_i\}) = F_i^* \quad (3)$$

That is to say, adding all arguments up to some  $i$ , and then closing, yields the same result as adding each argument one by one and closing prior to each subsequent addition.

Given Eq. (3), it suffices to show that  $\forall F_i^*$  in the above sequence,  $F_i^*$  is admissible:

*Base case:* By assumption of the acceptability of  $X_1$  w.r.t.  $E$ , and Lemma 19-1),  $F_1^* = F^*(E, X_1)$  is admissible. By Lemma 19-2),  $X_2, \dots, X_i$  are acceptable w.r.t.  $F_1^*$ .

*General case:* By inductive hypothesis,  $F_{i-1}^*$  is admissible, and  $X_i$  is acceptable w.r.t.  $F_{i-1}^*$ . By Lemma 19-1),  $F_i^*$  is admissible.  $\square$

We can now show the following key result:

**Proposition 22.** *Every admissible extension of a dialectical AF is a subset of a preferred extension.*

**Proof.** Let  $E$  be an admissible extension. We consider two cases:

1. Suppose  $E = \mathcal{F}(E)$ . By Lemma 15,  $E$  is epistemically maximal.
2. Suppose  $E \neq \mathcal{F}(E)$ . By Lemma 21,  $E \subseteq \mathcal{F}(E)$ , and  $E' = Cl_{em}(\mathcal{F}(E))$  is admissible, where trivially,  $E \subseteq E'$ .

In both cases,  $E \subseteq E'$  ( $E' = E$  in case 1), where  $E'$  is a *em* admissible set. By Corollary 20, there exists a maximal under set inclusion *em* admissible set  $E'' \supseteq E'$ . There cannot exist a fixed point of  $\mathcal{F}$  that is a strict superset of  $E''$ , since by Lemma 15, such a fixed point would be *em*. Hence  $E'' \supseteq E$  is a preferred extension.  $\square$

The importance of Proposition 22 for monological and dialogical applications, is that it suffices to show that an argument  $X$  is in an admissible extension, in order to show that  $X$  is credulously justified under the preferred semantics.

### 3.5.2. Monotonicity and a constructive definition of the grounded extension

We now show monotonicity of a dialectical AF's characteristic function, restricted to the domain of epistemically maximal admissible sets.

**Lemma 23.** *Let  $E, E'$  be two *em* admissible extensions such that  $E \subseteq E'$ . Then  $\mathcal{F}(E) \subseteq \mathcal{F}(E')$ .*

**Proof.** We show that  $X$  acceptable w.r.t.  $E$  ( $X \in \mathcal{F}(E)$ ) implies  $X$  acceptable w.r.t.  $E'$  ( $X \in \mathcal{F}(E')$ ). Suppose  $X \in \mathcal{F}(E)$ ,  $Z = (\Delta, \Gamma, \phi)$  ( $\phi = \perp$  or  $\overline{\alpha}$ ), and  $Z \Rightarrow_{E'} X$  on  $X' = (\{\alpha\}, \emptyset, \alpha)$ , and so  $Z \not\prec X'$ .

1) Suppose  $Z \Rightarrow_E X$ . Since  $X \in \mathcal{F}(E)$ ,  $\exists Y \in E$  s.t.  $Y \Rightarrow_{\{Z\}} Z$ . Since  $E \subseteq E'$ ,  $Y \in E'$ , and so  $X \in \mathcal{F}(E')$ .

2) Suppose  $Z \not\Rightarrow_E X$ . We show  $X \in \mathcal{F}(E')$ , by showing an argument in  $E'$  that defeats  $Z = (\Delta, \Gamma, \phi)$ .

Let  $\Gamma' = \Gamma \cap \text{prem}(E \cup \{X\})$  ( $\Gamma \setminus \Gamma'$  are premises in  $E'$  that are not in  $E \cup \{X\}$ ).

By Lemma 18,  $Z' = (\Delta \cup \Gamma \setminus \Gamma', \Gamma', \phi)$ ,  $Z' \Rightarrow_E X$ . Since  $X$  is acceptable w.r.t.  $E$ :

$\exists W = (\Pi, \Sigma, \psi) \in E$  ( $\psi = \perp$  or  $\overline{\beta}$ ),  $W \Rightarrow_{\{Z'\}} Z'$  on  $(\{\beta\}, \emptyset, \beta)$  and  $W \not\prec (\{\beta\}, \emptyset, \beta)$ .

- Suppose  $\beta \in \Gamma \setminus \Gamma'$  (where if  $\psi = \perp$  then  $\beta \in \Sigma \cap (\Gamma \setminus \Gamma')$ ). Hence  $\exists X' \in E' \setminus E \cup \{X\}$  s.t.  $\beta \in \text{prem}(X')$ .  
Given  $\Sigma \subseteq \Delta \cup (\Gamma \setminus \Gamma')$ , let  $W' = (\Sigma \cap \Delta, \Pi \cup (\Sigma \cap (\Gamma \setminus \Gamma')), \psi)$ . Since  $W \in E$ , then  $\Pi \subseteq \text{prem}(E)$ . Since  $\Sigma \cap (\Gamma \setminus \Gamma') \subseteq \text{prem}(E')$ ,  $W' \Rightarrow_{E'} X'$  on  $(\{\beta\}, \emptyset, \beta)$ . By assumption of admissibility of  $E'$ ,  $\exists Q \in E'$  s.t.  $Q \Rightarrow_{\{W'\}} W'$ . Since  $\text{prem}(W') \subseteq \text{prem}(Z)$ ,  $Q \Rightarrow_{\{Z\}} Z$ .
- Suppose  $\beta \in \Delta$  (if  $\psi = \perp$ ,  $\beta \in \Sigma \cap \Delta$ ). Let  $W' = (\Pi \cup (\Sigma \cap (\Gamma \setminus \Gamma')), \Sigma \cap \Delta, \psi)$ . By assumption of  $E'$  being *em*,  $W' \in E'$ . Since  $\Sigma \cap \Delta \subseteq \text{prem}(Z)$ ,  $W' \Rightarrow_{\{Z\}} Z$ .  $\square$

By definition of  $Cl_{em}$ , for any  $S \subseteq S'$ ,  $Cl_{em}(S) \subseteq Cl_{em}(S')$ . Hence Lemma 23 immediately implies that:

**Corollary 24.** Let  $E, E'$  be two *em* admissible extensions such that  $E \subseteq E'$ . Then  $Cl_{em}(\mathcal{F}(E)) \subseteq Cl_{em}(\mathcal{F}(E'))$ .

Consider now a characteristic function  $\mathcal{F}_p$  whose domain is sets  $E$  that are *em* admissible, and that returns  $Cl_{em}(\mathcal{F}(E))$ . By Lemma 21,  $\mathcal{F}_p$  returns a *em* admissible set. We then show that the fixed points of  $\mathcal{F}$  and  $\mathcal{F}_p$  coincide. Corollary 24 then guarantees existence of a least fixed point of  $\mathcal{F}_p$  and hence  $\mathcal{F}$ . Formally:

**Definition 19.** Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF and  $\mathcal{A}_p^2$  the set of all *em* admissible subsets of  $\mathcal{A}$ . Then  $\mathcal{F}_p : \mathcal{A}_p^2 \mapsto \mathcal{A}_p^2$ , where  $\mathcal{F}_p(E) = Cl_{em}(\mathcal{F}(E))$ .

**Proposition 25.** Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF. Then there exists a least fixed point of  $\mathcal{F}_{(\mathcal{A}, \mathcal{D})}$ .

**Proof.** Firstly, let  $E$  be a fixed point of  $\mathcal{F}_p$ . Since  $E$  is admissible,  $E \subseteq \mathcal{F}(E)$ . By definition of  $Cl_{em}$ ,  $\mathcal{F}(E) \subseteq Cl_{em}(\mathcal{F}(E))$ . Hence since  $E = \mathcal{F}_p(E) = Cl_{em}(\mathcal{F}(E))$ , then it must be that  $E = \mathcal{F}(E)$ .

Secondly, let  $E$  be a fixed point of  $\mathcal{F}$  ( $E = \mathcal{F}(E)$ ). By Lemma 15,  $E = \mathcal{F}(E)$  is *em*. By definition of  $Cl_{em}$ , if  $S$  is *em* then  $Cl_{em}(S) = S$ . Hence  $E = \mathcal{F}(E) = Cl_{em}(\mathcal{F}(E))$ . That is,  $E$  is a fixed point of  $\mathcal{F}_p$ . We have shown:

$$E \text{ is a fixed point of } \mathcal{F}_p \text{ iff } E \text{ is a fixed point of } \mathcal{F} \quad (4)$$

By Corollary 24,  $\mathcal{F}_p$  is monotonic, and so there exists a least fixed point of  $\mathcal{F}_p$ , which given Eq. (4) is a least fixed point of  $\mathcal{F}$ .  $\square$

We can therefore identify the grounded extension of a dialectical AF as the least fixed point of the framework's characteristic function. If we define a sequence, starting with the empty set, and iteratively applying  $\mathcal{F}_p$ , the monotonically increasing sequence approximates, and in the case of a *finitary* dialectical AF (see below) constructs, the least fixed point (*lfp*) of  $\mathcal{F}_p$ , i.e., the grounded extension.

**Proposition 26.** Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF, and  $F^0 = \emptyset$ ,  $F^{i+1} = \mathcal{F}_p(F^i)$ . Let  $E$  be the grounded extension of  $(\mathcal{A}, \mathcal{D})$ . Then:

1.  $E \subseteq \bigcup_{i=0}^{\infty} (F^i)$ .
2. If  $(\mathcal{A}, \mathcal{D})$  is finitary, i.e.,  $\forall X \in \mathcal{A}$ , the set  $\{Y \mid (Y, X) \in \mathcal{D}\}$  is finite, then  $E = \bigcup_{i=0}^{\infty} (F^i)$ .

**Proof.** Trivially,  $F^0 = \emptyset$  is *em* admissible. By Corollary 24,  $F^0, \dots, F^i$  is a monotonically increasing sequence. 1 is then shown, given that  $E$  is the *lfp* of  $\mathcal{F}_p$ . 2 is then shown by showing that  $\mathcal{F}_p$  is  $\omega$ -continuous. Let  $E' = F^0 \cup \dots \cup F^n \cup \dots$ , and let  $X \in \mathcal{F}_p(E')$ . Since there are finitely many arguments that defeat  $X$ , there exists a number  $m$  such that  $X \in \mathcal{F}_p(F^m)$ . Therefore  $\mathcal{F}_p(E') = F^0 \cup \dots \cup F^n \cup \dots$ .  $\square$

Finally, before formally showing satisfaction of these postulates, we discuss what it means for a preference relation to be *dialectically coherent* when determining the success of attacks as defeats. A preference  $Y \prec X$  invalidating an attack from  $Y = (\Delta, \emptyset, \overline{\alpha})$  to  $X = (\{\alpha\}, \emptyset, \alpha)$ , can be interpreted as:

from amongst the inconsistent  $\Delta \cup \{\alpha\}$ , one preferentially accepts (i.e., includes in a complete extension  $E$ ) arguments constructed from  $\alpha$  and rejects (excludes from  $E$ ) arguments constructed from  $\Delta \setminus \{\alpha\}$ .

Now, let  $Y' = (\Delta \cup \{\alpha\}, \emptyset, \phi)$  ( $\phi = \perp$  or  $\phi = \overline{\alpha}$ ). Then  $Y < X$  should imply that  $Y' < X$ . To see why, suppose  $Y' \not< X$ . But then this would contradict the interpretation ascribed above to  $Y < X$ , since  $Y' \not< X$  would mean that from amongst the inconsistent  $\Delta \cup \{\alpha\}$ , one *does not* preferentially accept arguments constructed from  $\alpha$  and reject arguments constructed from  $\Delta \cup \{\alpha\} \setminus \{\alpha\}$  (i.e., from  $\Delta \setminus \{\alpha\}$ ).

Moreover, suppose that for every  $\alpha_i$  in some inconsistent  $\Delta = \{\alpha_1, \dots, \alpha_n\}$ , one constructs the argument  $A_i = (\Delta \setminus \{\alpha_i\}, \emptyset, \overline{\alpha_i})$ , and that  $\forall i: A_i < (\{\alpha_i\}, \emptyset, \alpha_i)$ . Reasoning as above, we would then have  $\forall i: A'_i = (\Delta, \emptyset, \phi_i) < (\{\alpha_i\}, \emptyset, \alpha_i)$  ( $\phi_i = \perp$  or  $\phi_i = \overline{\alpha_i}$ ). In other words one preferentially accepts arguments built from  $\alpha_1$  to those built from  $\Delta$ , from  $\alpha_2$  to those built from  $\Delta$ , and so on. But then (by Lemma 9), if  $E$  includes arguments with premises  $\alpha_1, \dots, \alpha_n$  then  $E$  includes arguments built from premises  $\Delta$ . Contradiction. Hence, it would be incoherent to have that  $\forall \alpha \in \Delta: (\Delta, \emptyset, \perp) < (\{\alpha\}, \emptyset, \alpha)$  and (since  $<$  is *imle*)  $(\emptyset, \Delta, \perp) < (\{\alpha\}, \emptyset, \alpha)$ . Indeed, all the results shown thus far would hold if instead of enforcing that attacks from  $(\emptyset, \Delta, \perp)$  on each  $\alpha \in \Delta$  always succeed as a defeat (given that it would be incoherent to reject the dialectical demonstration of inconsistency), one instead assumed a coherent preference relation implying that for at least one  $\alpha \in \Delta$ ,  $(\emptyset, \Delta, \perp)$  defeats  $(\{\alpha\}, \emptyset, \alpha)$ .<sup>17</sup>

**Definition 20** (*Dialectically coherent preference relations*). Let  $(\mathcal{A}, \mathcal{D})$  be a *pdAF* defined by  $\mathcal{B}$  and  $<$ . Then  $<$  is *dialectically coherent* iff

$$\forall (\emptyset, \Delta, \perp) \in \mathcal{A}: \exists \alpha \in \Delta \text{ such that } (\emptyset, \Delta, \perp) \not< (\{\alpha\}, \emptyset, \alpha). \quad (\text{Pref1})$$

$$\forall X = (\{\alpha\}, \emptyset, \alpha), Y = (\Delta, \emptyset, \overline{\alpha}), Y' = (\Delta \cup \{\alpha\}, \emptyset, \phi) (\phi = \overline{\alpha} \text{ or } \phi = \perp): Y < X \text{ implies } Y' < X. \quad (\text{Pref2})$$

In Section 4 we show that the Elitist preference satisfies *Pref1* and *Pref2*. Notice that since  $<$  is *imle*, *Pref2* implies that if  $(\Delta, \Gamma, \overline{\alpha}) < (\{\alpha\}, \emptyset, \alpha)$  then  $(\Delta, \Gamma \cup \{\alpha\}, \phi) < (\{\alpha\}, \emptyset, \alpha)$ . To see why preference relations that do not satisfy *Pref2* may yield counter-intuitive results, suppose that in Example 28 we have that  $C < A$  but  $D \not< A$ , thus violating *Pref2*. We would then have an additional complete extension  $E_2$  containing  $C$  and  $D$ , since the defeats by  $A$  and  $B$  on  $C$  and  $D$  would now be defended by defeats from  $D$  on  $A$  and  $B$ . Counter-intuitively one would obtain a complete extension containing  $C$ , despite a strict preference for  $A$  over  $C$ .

Finally, in now proving the non-interference and crash resistance postulates, we will make use of the following result that holds for preference relations satisfying *Pref1*:

**Lemma 27.** Let  $(\mathcal{A}, \mathcal{D})$  be a *pdAF* defined by  $\mathcal{B}$  and  $<$ , where  $<$  satisfies *Pref1*. Then for any  $Y = (\Delta, \Gamma, \perp)$ ,  $Y' = (\Gamma, \Delta, \perp): Y \Rightarrow_{\{Y\}} Y'$  or  $Y' \Rightarrow_{\{Y\}} Y$ .

**Proof.** Suppose for contradiction that  $Y \not\Rightarrow_{\{Y\}} Y'$  and  $Y' \not\Rightarrow_{\{Y\}} Y$ . Since  $<$  is *imle*,  $\forall \alpha \in \Delta \cup \Gamma: (\emptyset, \Delta \cup \Gamma, \perp) < (\{\alpha\}, \emptyset, \alpha)$ , contradicting *Pref1*.  $\square$

### 3.6. Contaminated arguments

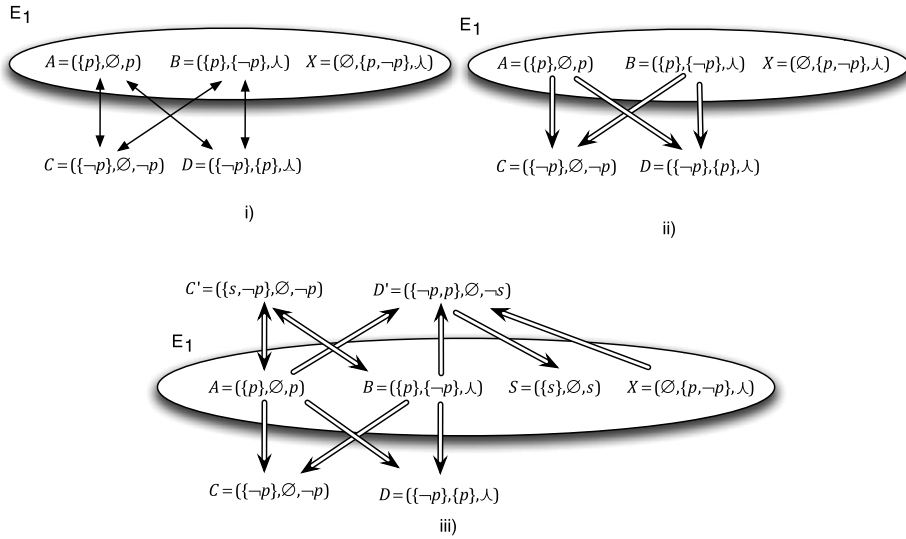
Dialectical *AFs* satisfy consistency and closure given relatively undemanding assumptions as to the arguments constructed from a base, while assuming arbitrary preference relations and eschewing the need for checking the legitimacy of arguments' assumptions. However, dropping consistency and subset minimality checks on arguments' assumptions may result in violation of the non-contamination postulates.

**Example 28.** Let  $\mathcal{B} = \{p, \neg p\}$ . Fig. 3-i) shows arguments and attacks defined by  $\mathcal{B}$ . Suppose  $C < A$ ,  $D < A$ , and so we have the defeats in Fig. 3-ii) and  $E_1$  is the single grounded, preferred and stable extension (notice that  $C \not\Rightarrow_{E_1} B$  since  $C$  attacks  $B$  on  $A$  and  $C < A$ ). Suppose we now add  $s$  to  $\mathcal{B}$ , and suppose the defined arguments in Fig. 3-iii). Again, assume only  $C < A$ ,  $D < A$ . Intuitively, since  $s$  is syntactically disjoint from  $p, \neg p$ , we should continue to obtain a single grounded, preferred and stable extension  $\{A, B, X, S\}$ .

Firstly, note that in contrast with the example in Section 3.2, we avoid the 'contaminating effect' (in the sense that the non-contamination postulates are violated) of arguments that are used to entail syntactically disjoint conclusions via the 'explosivity' (recall footnote 9) of classically inconsistent premises. Although  $D'$  defeats  $S$  (Fig. 3-iii), membership of  $S$  in any complete extension  $E$  can never be invalidated since any such  $E$  will contain the unassailable  $X$  that defeats  $D'$  and so defends  $S$ . Hence crash resistance is not violated ( $X$ 's defence of  $S$  also prevents violation of non-interference that may occur due to the defeat by  $D'$ ).

Secondly, with the addition of  $s$  to  $\mathcal{B}$ , we can also construct  $C' = (\{s, \neg p\}, \emptyset, \neg p)$ . Since  $C' \not< A$ , we have that  $C' \Rightarrow_{E_1} A$ . We then have an additional complete extension  $E_2 = \{C, C', D, S, X\}$ , since although  $A \Rightarrow_{E_2} V$ ,  $V \in \{C, C', D\}$ , each such defeat is defended given that  $C' \Rightarrow_{\{A\}} A$ . This is clearly counter-intuitive. We no longer obtain a single grounded, preferred

<sup>17</sup> In Remark 8, it would remain the case that no admissible extension can contain  $A_1$  and  $F_1$  since a coherent  $<$  would imply that the unassailable  $X_1$  defeats either  $A_1$  on  $A_1$  or  $F_1$  on either  $B_1$  or  $C_1$ .



**Fig. 3.** i) All attacks are bidirectional; ii) Defeats resulting in single complete extension  $E_1$ ; iii) Arguments and defeats given base  $\{p, \neg p, s\}$  and  $C < A$ ,  $C' \not\prec A$ .

and stable extension containing  $A$ ; indeed the grounded extension  $E$  will now only contain  $S$  and  $X$ . Thus, by adding the syntactically disjoint  $s$  to  $B$ ,  $A$  is no longer a sceptically justified argument under any semantics, and hence non-interference is violated.

The above example describes two types of contamination. *The first type of contamination* may occur when the assumptions  $\Gamma$  of an argument  $X$ , with conclusion  $\alpha \neq \lambda$ , include some inconsistent  $\Delta$  that is syntactically disjoint from  $\Gamma \setminus \Delta \cup \{\alpha\}$ , and  $\alpha$  may be derived by the explosivity of  $\Delta$  (i.e.,  $\Delta$  contaminates  $X$  by explosivity).

**Remark 29.** Notice that:

1. if  $\alpha = \lambda$  then  $X$  is of the form  $(\Sigma, \Lambda, \lambda)$  and cannot have a contaminating effect due to explosivity, since  $X$  can only attack arguments on assumptions in  $\Lambda \neq \emptyset$ , indicating that  $\Sigma \cup \Lambda$  is inconsistent (recall Definition 11).
2. We emphasise above that  $\alpha$  may be derived by explosivity since we are presently considering representations of arguments independently of the proof theory used for their construction. Hence, given an argument represented as  $(\{p, \neg p, r, r \rightarrow q\}, \emptyset, q)$  it may be that  $q$  is derived from  $r, r \rightarrow q$ , or explosively from  $p, \neg p$ ; there is no way of telling which is the case without reference to the proof theory. Note that the same reasoning applies to  $(\{p, \neg p\}, \emptyset, q \vee \neg q)$  in that the tautology may be derived from empty assumptions or explosively.
3. In addition to the arguments  $\mathcal{A}$  in a dialectical AF satisfying P1, P2 and P3 in Definition 17, we will later stipulate that if  $X \in \mathcal{A}$  and  $\Delta$  contaminates  $X$  by explosivity, then  $\mathcal{A}$  also contains a falsum argument with assumptions  $\Delta$ . (Intuitively, if an argument's inconsistent assumptions are used to conclude a syntactically disjoint conclusion via explosivity, then this can be explicitly recognised by concluding the syntactically disjoint  $\lambda$ .)

Violation of the non-contamination postulates due to arguments contaminated by explosivity is then avoided if the assumption in 3 above is satisfied. In Example 28,  $\{p, \neg p\}$  contaminates  $D'$  by explosivity. However, we also have  $X = (\emptyset, \{p, \neg p\}, \lambda)$  that defeats  $D'$ .

A *second type of contamination* may arise because we do not enforce that arguments' assumptions are subset minimal. Adding  $s$  to the premises of  $C$  yields the non-subset minimal  $C'$  that is stronger than its subset minimal counterpart  $C$  ( $C < A$  but  $C' \not\prec A$ , and so  $C' \Rightarrow_{E_1} A$ ). As discussed in Section 3.2, [43] shows that if subset minimal arguments are not strengthened when adding redundant assumptions, then the argumentation defined inferences of a framework consisting only of subset minimal arguments, is not changed when adding non-subset minimal arguments to the framework. Intuitively, this also means that non-interference would not be violated due to inclusion of non-subset minimal arguments. However, we have argued that ensuring that assumptions are not redundant with respect to deriving an argument's conclusion, should not be enforced by computationally expensive checks on subset minimality, but rather should in principle be enforceable by the proof system used for constructing arguments. Recall that the issue here is not whether assumptions are *actually used* in deriving the conclusion (footnote 8 on page 12 provides an example of a non-subset minimal argument in which all assumptions are used in the proof), since one can readily exclude assumptions not actually used in a given proof. Rather, the focus is on identifying whether arguments specified 'abstractly' as tuples consisting of assumptions paired with a conclusion (i.e., without reference to a specific proof system) include obviously redundant assumptions. One such obvious notion of

redundancy is implied by the following property of classical logic (recall – Notation 3 – that  $\parallel$  denotes ‘syntactically disjoint’; to spare on parentheses, we also assume that  $\cup$  binds more tightly than  $\parallel$ ):

**Proposition 30.** *Let  $\mathcal{B}$  be a base. For all  $\Gamma \subseteq \mathcal{B}$ , if  $\Gamma \vdash_c \alpha$  and  $\exists \Delta \subseteq \Gamma$  such that  $\Delta \parallel (\Gamma \setminus \Delta) \cup \{\alpha\}$ , then either  $\Delta \vdash_c \perp$  or  $\Gamma \setminus \Delta \vdash_c \alpha$ .*

**Proof.** Suppose for contradiction that  $\Delta \not\vdash_c \perp$  and  $\Gamma \not\vdash_c \alpha$ . By classical semantics, there is a model  $M_1$  such that the value of all the formulas in  $\Gamma$  is 1 and the value of  $\alpha$  is 0. The same holds true of every model  $M'_1$  that agrees with  $M_1$  on the interpretation of symbols( $\Gamma \cup \{\alpha\}$ ). Moreover, since  $\Delta$  is consistent, there is a model  $M_2$  such that the value of all the formulas in  $\Delta$  is 1. The same holds true of every model  $M'_2$  that agrees with  $M_2$  on the interpretation of symbols( $\Delta$ ). Now, since  $\Delta \parallel \Gamma \cup \{\alpha\}$ , there is a model  $M_3$  that agrees with  $M_1$  on the interpretation of symbols( $\Gamma \cup \{\alpha\}$ ) and with  $M_2$  on the interpretation of symbols( $\Delta$ ). For such a model the values of all the formulas in  $\Delta \cup \Gamma$  is 1 and the value of  $\alpha$  is 0, contradicting the hypothesis that  $\Delta \cup \Gamma \vdash_c \alpha$ .  $\square$

**Remark 31.** Suppose a dialectical AF  $(\mathcal{A}, \mathcal{D})$  such that  $Y = (\Gamma, \emptyset, \alpha) \in \mathcal{A}$ , and there is a non-empty  $\Delta \subseteq \Gamma$  such that  $\Delta \parallel (\Gamma \setminus \Delta) \cup \{\alpha\}$ . Then, given Proposition 30:

1. If  $\Gamma \setminus \Delta \neq \emptyset$  then either:
  - i)  $\Delta$  ‘redundantly weakens’<sup>18</sup>  $Y' = (\Gamma \setminus \Delta, \emptyset, \alpha)$  (i.e.,  $\Delta \subseteq \text{assumptions}(Y)$  are trivially unnecessary for obtaining  $\alpha$ ), or
  - ii)  $\Gamma \setminus \Delta$  redundantly weakens  $Y' = (\Delta, \emptyset, \alpha)$  (where  $\Delta$  explosively entails  $\alpha$  by virtue of the inconsistency of  $\Delta$ ).

For example, given  $Y = (\{s, \neg p\}, \emptyset, \neg p)$  then  $\Delta = \{s\}$  redundantly weakens  $Y' = (\{\neg p\}, \emptyset, \neg p)$  (case 1i). Given  $Y = (\{q, p, \neg p\}, \neg s)$ , then  $\Gamma \setminus \Delta = \{q\}$  redundantly weakens  $Y' = (\{p, \neg p\}, \emptyset, \neg s)$  (case 1ii).
2. If  $\Gamma \setminus \Delta = \emptyset$  then either:
  - i)  $\Delta \not\vdash_c \perp$  and so by classical semantics  $\alpha$  must be a tautology, and  $\Gamma$  redundantly weakens  $Y' = (\emptyset, \emptyset, \alpha)$ , or
  - ii)  $\Delta \vdash_c \perp$  and so either  $\alpha$  is a tautology and  $\Gamma$  redundantly weakens  $Y' = (\emptyset, \emptyset, \alpha)$ , or  $\Delta$  explosively entails  $\alpha$  by virtue of the inconsistency of  $\Delta$ .

For example, given  $Y = (\{s\}, \emptyset, p \vee \neg p)$ , then  $\Gamma = \{s\}$  redundantly weakens  $Y' = (\emptyset, \emptyset, p \vee \neg p)$  (case 2i). Given  $(\{s, \neg s\}, \emptyset, p \vee \neg p)$ , then either  $\Gamma = \{s, \neg s\}$  redundantly weakens  $Y' = (\emptyset, \emptyset, p \vee \neg p)$ , or  $p \vee \neg p$  is explosively entailed from the inconsistent  $\{s, \neg s\}$  (case 2ii).
3. In addition to P1, P2 and P3, we will later stipulate that if  $Y \in \mathcal{A}$ , then:
  - In case 1  $\mathcal{A}$  also includes the redundantly weakened  $Y' = (\Gamma \setminus \Delta, \emptyset, \alpha)$ , or, given the redundantly weakened explosive  $Y' = (\Delta, \emptyset, \alpha)$ ,  $\mathcal{A}$  also includes  $(\Delta, \emptyset, \perp)$  (e.g.,  $Y'' = (\{p, \neg p\}, \emptyset, \perp)$ ) as stipulated in Remark 29-3.
  - In case 2,  $\mathcal{A}$  also includes the redundantly weakened  $Y' = (\emptyset, \emptyset, \alpha)$ , or, given the explosive  $(\Delta, \emptyset, \alpha)$ ,  $\mathcal{A}$  also includes  $(\Delta, \emptyset, \perp)$  (e.g.,  $(\{s, \neg s\}, \emptyset, \perp)$ ).

Let us formally define the above notions. We first identify the notion of a contaminated argument, and then further distinguish when such an argument is redundantly and/or explosively contaminated.

**Definition 21** (*Contaminated arguments*). Let  $Y$  be an argument with  $\text{assumptions}(Y) = \Gamma$  and  $\text{conclusion}(Y) = \alpha$ . We say that  $Y$  is *contaminated* if there exists some non-empty  $\Delta \subseteq \Gamma$  such that the following conditions hold:

1.  $\Delta \parallel (\Gamma \setminus \Delta) \cup \{\alpha\}$ ;
2. either  $\Gamma \setminus \Delta \neq \emptyset$  or  $\alpha \neq \perp$ .

We also say that any such  $\Delta$  is a *contaminating set* for  $Y$ .

Note that if an argument  $Y$  satisfies condition 1 and is not contaminated, then it violates condition 2; that is to say,  $\text{assumptions}(Y)$  cannot be partitioned into non-empty syntactically disjoint subsets, so that *the only contaminating*  $\Delta$  for  $Y$  is such that  $\Gamma \setminus \Delta = \emptyset$ , and  $\text{conclusion}(Y) = \perp$  (recall that  $\perp$  does not appear in a base and so it must be that  $\Gamma = \Delta$  is syntactically disjoint from  $\perp$ ). Hence  $Y$  has not been obtained by some assumptions redundantly weakening an argument in any of the senses described in Remark 31. Moreover, recalling Remark 29-1, such an argument is not explosively contaminating. To illustrate,  $(\{p, \neg p\}, \emptyset, \perp)$  is an example of such a  $Y$  that is not contaminated (either redundantly or explosively). On the other hand  $(\{p, \neg p, r\}, \emptyset, \perp)$  is contaminated redundantly but not explosively, and  $(\{p, \neg p\}, \emptyset, \{q\})$  is contaminated explosively but not redundantly.

**Definition 22** (*R-contaminated arguments*). Let  $Y$  be a contaminated argument such that  $\text{assumptions}(Y) = \Gamma$  and  $\text{conclusion}(Y) = \alpha$ . We say that  $Y$  is *redundantly contaminated* or *R-contaminated* if:

<sup>18</sup> We use the term ‘weakens’ in keeping with the standard use of the term weakening in logic to denote that an entailment from  $\Delta$  continues to be valid when adding some  $\Gamma$  to  $\Delta$ .



- either (i) for some contaminating set  $\Delta \subseteq \Gamma$  for  $Y$ ,  $\Gamma \setminus \Delta \neq \emptyset$ , or  
(ii) the only contaminating set for  $Y$  is  $\Gamma$  itself (i.e.,  $\Gamma \setminus \Delta = \emptyset$ ) and  $\Delta \not\vdash_c \perp$ .

Conditions i) and ii) respectively state that  $Y$  is an R-contaminated argument of the type described in 1 and 2i) of Remark 31. Arguments of the type described in 2ii), such as  $Y = (\{p, \neg p\}, \emptyset, q \vee \neg q)$ , cannot be definitively recognised as being R-contaminated. This is because it may be that  $X = (\emptyset, \emptyset, q \vee \neg q) \notin \mathcal{A}$ . That is to say  $Y$  has not been obtained by  $\{p, \neg p\}$  redundantly weakening  $X$ . Rather,  $q \vee \neg q$  is *only* derived by explosivity and so we have (given Remark 29-3) that  $(\{p, \neg p\}, \emptyset, \perp) \in \mathcal{A}$  (and so the property described in Remark 31-3 is satisfied). On the other hand one can definitively identify  $Y = (\{p, \neg p, q, q \rightarrow r\}, \emptyset, r)$  as being R-contaminated, since satisfaction of the property in Remark 31-3 means that either  $(\{p, \neg p\}, \emptyset, r) \in \mathcal{A}$  or  $(\{q, q \rightarrow r\}, \emptyset, r) \in \mathcal{A}$ . In either case  $Y$  includes a redundantly weakening set of assumptions  $(\{q, q \rightarrow r\}$  respectively  $\{p, \neg p\})$ .

We now formally define explosively contaminated arguments:

**Definition 23** (*E-contaminated arguments*). Let  $Y$  be a contaminated argument such that  $\text{conclusion}(Y) = \alpha$ . We say that  $Y$  is *explosively contaminated* or *E-contaminated* if  $\alpha \neq \perp$  and  $\Delta \vdash_c \perp$  for some  $\Delta$  that is a contaminating set for  $Y$ .

**Example 32** (*Examples of (non) contaminated arguments*).

1.  $X = (\{p, \neg p, q, q \rightarrow r\}, \emptyset, r)$ ,  $Y = (\{p, \neg p, q\}, \emptyset, r)$ ,  $Z = (\{p, \neg p, q, \neg q\}, \emptyset, r)$  are all both R-contaminated and E-contaminated;
2. Both  $(\{p, \neg p\}, \emptyset, \{q\})$  and  $(\{p, \neg p\}, \emptyset, \{q \vee \neg q\})$  are E-contaminated, but not R-contaminated;
3. Both  $(\{p, \neg p\}, \emptyset, \{p \wedge q\})$  and  $(\{p, \neg p, r\}, \emptyset, r \wedge p)$  are neither R-contaminated nor E-contaminated (as neither is contaminated);
4.  $(\{p, \neg p, r\}, \emptyset, \perp)$  is R-contaminated but not E-contaminated.  $(\{p, \neg p\}, \emptyset, \perp)$  is neither R-contaminated nor E-contaminated.

For the examples in 1 above, if the property in Remark 31-3 is satisfied, then:

- if  $X \in \mathcal{A}$  then either  $X' = (\{q, q \rightarrow r\}, \emptyset, r) \in \mathcal{A}$  and  $X'$  has been redundantly weakened by  $\{p, \neg p\}$ , or the explosive  $X'' = (\{p, \neg p\}, \emptyset, r)$  has been redundantly weakened by  $\{q, q \rightarrow r\}$ , and so  $(\{p, \neg p\}, \emptyset, \perp) \in \mathcal{A}$ .
- If  $Y \in \mathcal{A}$  then  $(\{p, \neg p\}, \emptyset, \perp) \in \mathcal{A}$ .
- If  $Z \in \mathcal{A}$  then:
  - either  $(\{q, \neg q\}, \emptyset, r) \in \mathcal{A}$  or  $(\{p, \neg p\}, \emptyset, \perp) \in \mathcal{A}$ , and
  - either  $(\{p, \neg p\}, \emptyset, r) \in \mathcal{A}$  or  $(\{q, \neg q\}, \emptyset, \perp) \in \mathcal{A}$ .

### 3.7. Coherent preference relations and partially instantiated dialectical AFs

Section 4 describes a propositional natural deduction proof theory that does not license construction of R-contaminated arguments. However, satisfaction of the non-contamination postulates when using proof systems that do generate such arguments, requires that preference relations are ‘redundance-coherent’ in the sense that arguments are not strengthened when redundantly weakening with syntactically disjoint assumptions (the *Elitist* preference is shown to be redundance-coherent in Section 4):

**Definition 24** (*Redundance-coherent preference relations*). Let  $\mathcal{A}$  be a set of dialectical arguments. A strict partial ordering  $<$  over  $\mathcal{A}$  is *redundance-coherent* iff

$$\forall X, Y, Y' \text{ such that } Y = (\Gamma, \emptyset, \alpha), Y' = (\Delta \cup \Gamma, \emptyset, \alpha), \text{ and } \Delta \parallel \Gamma \cup \{\alpha\}: \text{ if } Y < X \text{ then } Y' < X.$$

We now define *partially instantiated dialectical AFs* (*pdAFs*) whose arguments  $\mathcal{A}$  satisfy P1, P2 and P3, and the property described in Remark 31-3 (which, as is easy to verify, subsumes the property in Remark 29-3). We also distinguish non-redundant *pdAFs* whose arguments are obtained by proof theories that do not generate R-contaminated arguments. In this case it suffices that  $\mathcal{A}$  satisfies the property relating to arguments contaminated by explosivity described in Remark 29-3.

**Definition 25** (*(Non-redundant) partially instantiated dialectical AF*). Let  $(\mathcal{A}, \mathcal{D})$  be a dialectical AF such that  $\mathcal{D}$  is the defeat relation over  $\mathcal{A}$  defined by the attacks  $\mathcal{C}$  and the strict partial ordering  $<$  over  $\mathcal{A}$ .

- $(\mathcal{A}, \mathcal{D})$  is a *partially instantiated dialectical AF* (*pdAF*) if  $\mathcal{A}$  is any subset of the dialectical arguments defined by a base  $\mathcal{B}$  such that  $\mathcal{A}$  satisfies P1, P2 and P3, and:

(P4) If  $(\Gamma, \emptyset, \alpha) \in \mathcal{A}$ ,  $\Delta \subseteq \Gamma$ ,  $\Delta \neq \emptyset$  and  $\Delta \parallel \Gamma \setminus \Delta \cup \{\alpha\}$ , then either  $(\Delta, \emptyset, \perp) \in \mathcal{A}$  or  $(\Gamma \setminus \Delta, \emptyset, \alpha) \in \mathcal{A}$ .

•  $(\mathcal{A}, \mathcal{D})$  is a *non-redundant pdAF* if  $\mathcal{A}$  is any subset of the dialectical arguments defined by a base  $\mathcal{B}$  such that  $\mathcal{A}$  satisfies P1, P2 and P3, and:

(P4') If  $(\Gamma, \emptyset, \alpha) \in \mathcal{A}$ ,  $\Delta \subseteq \Gamma$ ,  $\Delta \neq \emptyset$  and  $\Delta \parallel (\Gamma \setminus \Delta) \cup \{\alpha\}$ , then  $(\Delta, \emptyset, \lambda) \in \mathcal{A}$ .

(P5)  $\neg \exists Y \in \mathcal{A}$  such that  $Y$  is R-contaminated.

We say that  $(\mathcal{A}, \mathcal{D})$  is a (non-redundant) *pdAF* defined by  $\mathcal{B}$  and  $\prec$ .

By virtue of non-redundant *pdAFs* satisfying P5, any argument in such a *pdAF* that satisfies the precondition of P4 (and hence P4') can only be explosively contaminated, and the argument's assumptions cannot be partitioned into non-empty syntactically disjoint subsets. Hence it suffices that a non-redundant *pdAF* satisfy P4'. Indeed, we leave it to the reader to verify that:

**Remark 33.** Let  $(\mathcal{A}, \mathcal{D})$  be a non-redundant *pdAF*, and  $(\Gamma, \emptyset, \alpha) \in \mathcal{A}$ ,  $\Delta \subseteq \Gamma$ ,  $\Delta \neq \emptyset$  and  $\Delta \parallel (\Gamma \setminus \Delta) \cup \{\alpha\}$ . Then it must be the case that  $\Delta = \Gamma$ .

Henceforth, when referring to '*pdAFs*' we assume *pdAFs* in the general sense or non-redundant *pdAFs*. Finally note that since the arguments in (non-redundant) *pdAFs* satisfy P1, P2 and P3, the results in Sections 3.4 and 3.5 all hold for such frameworks.

Let us now review the significance of the concepts thus far introduced. Suppose the *pdAF*  $(\mathcal{A}, \mathcal{D})$  in Example 28 satisfies P4. Given the R-contaminated  $C' \in \mathcal{A}$ , P4 implies  $C \in \mathcal{A}$ . Moreover, if the *pdAF's* preference relation is redundancy-coherent, then  $C \prec A$  implies  $C' \prec A$ . Hence  $C' \not\Rightarrow_{\{A\}} A$  and non-interference is no longer violated. Furthermore, we have already noted that if P4 is satisfied, then an E-contaminated argument  $Y$  implies existence of an argument concluding  $\lambda$  from the inconsistent  $\Delta$  in assumptions( $Y$ ), and this ensures satisfaction of the non-contamination postulates. Now, as will be shown in Section 4, one can define proof theories that do not generate R-contaminated arguments (such as  $C'$ ). However, to the best of our knowledge, proof theories cannot be defined so as to exclude non-subset minimal arguments. We have therefore defined non-redundant *pdAFs* that only need satisfy the property P4' implied by P4, and as shown in the following section, need not assume redundancy-coherent preference relations (given the absence of R-contaminated arguments) in order that the non-contamination postulates are satisfied.

### 3.7.1. Proving the non-interference and crash resistance postulates

The non-interference and crash resistance postulates [16] (reviewed in Section 2.2) are stated with respect to *AFs* defined by bases  $\mathcal{B}_1$ ,  $\mathcal{B}_2$  and  $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ . For *pdAFs*, we also need to account for preferences over arguments defined by these bases.

**Definition 26** (*Composing pdAFs*). Let  $(\mathcal{A}_1, \mathcal{D}_1)$  defined by  $\mathcal{B}_1$  and  $\prec_1$ , and  $(\mathcal{A}_2, \mathcal{D}_2)$  defined by  $\mathcal{B}_2$  and  $\prec_2$ . Then  $(\mathcal{A}, \mathcal{D})$  defined by  $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$  and  $\prec$  is said to be *composed from*  $(\mathcal{A}_1, \mathcal{D}_1)$  and  $(\mathcal{A}_2, \mathcal{D}_2)$ , denoted  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ , iff:

1.  $\mathcal{A}_1 \cup \mathcal{A}_2 \subseteq \mathcal{A}$ .<sup>19</sup>
2.  $\forall X \in \mathcal{A}$  : if assumptions( $X$ )  $\subseteq \mathcal{B}_1$  then  $X \in \mathcal{A}_1$  and if assumptions( $X$ )  $\subseteq \mathcal{B}_2$  then  $X \in \mathcal{A}_2$ .
3.  $\prec$  is any preference ordering such that:
  - $\forall X_1, Y_1 \in \mathcal{A}_1$  :  $(X_1, Y_1) \in \prec_1$  iff  $(X_1, Y_1) \in \prec$
  - $\forall X_2, Y_2 \in \mathcal{A}_2$  :  $(X_2, Y_2) \in \prec_2$  iff  $(X_2, Y_2) \in \prec$

The above formalises the assumption that preferences defined over any set of arguments  $\mathcal{A}$ , remain unchanged when adding further arguments to  $\mathcal{A}$  (3), and that the arguments constructed from some base  $\mathcal{B}$  are exactly those constructed from the subset  $\mathcal{B}$  of some base  $\mathcal{B} \cup \mathcal{B}'$  (1 and 2). Intuitively, if the available resources are the limiting factor determining construction of arguments, then we assume the same resources available for constructing arguments from  $\mathcal{B}$ , independently of whether  $\mathcal{B}$  is included in a larger base  $\mathcal{B} \cup \mathcal{B}'$ .

**Remark 34** (*Assumptions on composed pdAFs*). In what follows, when stating that  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ , we assume the *pdAFs* defined by  $\mathcal{B}_1$ ,  $\prec_1$  and  $\mathcal{B}_2$ ,  $\prec_2$  and  $\mathcal{B}_1 \cup \mathcal{B}_2$ ,  $\prec$  respectively, and that  $\mathcal{B}_1 \parallel \mathcal{B}_2$  ( $\mathcal{B}_1$  is syntactically disjoint from  $\mathcal{B}_2$ ).

**Remark 35** (*Assumptions on preference relations in (non-redundant) pdAFs*). In the following results, when referring to *pdAFs* in general, we will assume *pdAFs* defined by preference relations that satisfy *Pref1* (and hence the results can make use

<sup>19</sup> It is obvious to then see that for the attack relations  $\mathcal{C}_1$ ,  $\mathcal{C}_2$  and  $\mathcal{C}$  that are used to obtain the defeats  $\mathcal{D}_1$ ,  $\mathcal{D}_2$  and  $\mathcal{D}$ :  $(\mathcal{C}_1 \cup \mathcal{C}_2) \subseteq \mathcal{C}$ .

of Lemma 27) and are redundance-coherent. When referring to non-redundant *pdAFs*, we assume only that the preference relations satisfy *Pref1*.

**Lemma 36.** Let  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ . Suppose  $Y \in \mathcal{A} \setminus \mathcal{A}_1$ ,  $X \in \mathcal{A}_1$ , and  $Y \Rightarrow_{S \subseteq \mathcal{A}} X$  on  $X' = (\bar{\alpha}, \emptyset, \bar{\alpha})$ . Let  $Y = (\Delta, \Gamma, \phi)$ ,  $\Delta_1 = \Delta \cap \mathcal{B}_1$ ,  $\Gamma_1 = \Gamma \cap \mathcal{B}_1$ ,  $\Delta_2 = \Delta \cap \mathcal{B}_2$ ,  $\Gamma_2 = \Gamma \cap \mathcal{B}_2$ . Then:

1.  $(\Delta_2, \Gamma_2, \lambda) \in \mathcal{A}_2$ , or
2.  $(\Delta_1, \Gamma_1, \lambda) \in \mathcal{A}_1$ , or
3.  $(\Delta_1, \Gamma_1, \alpha) \in \mathcal{A}_1$

**Proof.** Suppose  $\phi = \lambda$ . Since  $Y \in \mathcal{A} \setminus \mathcal{A}_1$ , then by Definition 26.2,  $\Delta_2 \cup \Gamma_2 \neq \emptyset$ . Since  $\lambda$  is not an atom in  $\mathcal{B}$ , then  $\Delta_2 \cup \Gamma_2 \parallel \Delta_1 \cup \Gamma_1 \cup \{\lambda\}$ . By P4 and P2, either  $Y_1 = (\Delta_1, \Gamma_1, \lambda) \in \mathcal{A}$  and (by Definition 26.2),  $Y_1 \in \mathcal{A}_1$ , or  $Y_2 = (\Delta_2, \Gamma_2, \lambda) \in \mathcal{A}$  and (by Definition 26.2),  $Y_2 \in \mathcal{A}_2$ . Suppose  $\phi = \alpha$ . Since  $X' \in \mathcal{A}_1$ ,  $\Delta_2 \cup \Gamma_2 \parallel \alpha$ . Hence  $\Delta_2 \cup \Gamma_2 \parallel \Delta_1 \cup \Gamma_1 \cup \{\alpha\}$ . By P4 and P2:  $Y_1 = (\Delta_1, \Gamma_1, \alpha) \in \mathcal{A}$  hence  $Y_1 \in \mathcal{A}_1$ , or  $Y_2 = (\Delta_2, \Gamma_2, \lambda) \in \mathcal{A}$  hence  $Y_2 \in \mathcal{A}_2$ .  $\square$

**Lemma 37.** Let  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ , and  $E' \subseteq \mathcal{A}$ ,  $E = E' \cap \mathcal{A}_1$ . Suppose  $Y = (\Delta, \Gamma, \phi) \in \mathcal{A}$ ,  $X \in \mathcal{A} \cap \mathcal{A}_1$ ,  $Y \Rightarrow_{E'} X$  on  $X' = (\{\bar{\alpha}\}, \emptyset, \bar{\alpha})$ , and:

- A1  $X$  is acceptable w.r.t.  $E$  in  $(\mathcal{A}_1, \mathcal{D}_1)$ ;
- A2  $\forall \Delta \subseteq \text{prem}(E' \setminus E): (\text{prem}(A) = \Delta) \rightarrow (A \in E' \text{ and } A \text{ is acceptable w.r.t. } E')$ .

Then  $\exists Z \in E'$  such that  $Z \Rightarrow_{\{Y\}} Y$ .

**Proof.** 1) Suppose  $Y \in \mathcal{A}_1$ . Since  $E = E' \cap \mathcal{A}_1$  and (given  $X \in \mathcal{A}_1$ )  $\Gamma \subseteq \text{prem}(E' \cup \{X\}) \cap \mathcal{B}_1$ , then (given  $Y \not\prec X'$  implies  $Y \not\prec_1 X'$ )  $Y \Rightarrow_{E \subseteq \mathcal{A}_1} X$ .

Given A1,  $\exists Z \in E$  s.t.  $Z \Rightarrow_{\{Y\}} Y$  on  $Y'$ ,  $Z \not\prec_1 Y'$  and so  $Z \not\prec Y'$ . Since  $E \subseteq E'$ , then  $Z \in E'$  and  $Z \Rightarrow_{\{Y\}} Y$  on  $Y'$ .

2) Suppose  $Y \in \mathcal{A} \setminus \mathcal{A}_1$ . By Lemma 36, either: i)  $Y_1 = (\Delta \cap \mathcal{B}_1, \Gamma \cap \mathcal{B}_1, \lambda) \in \mathcal{A}_1$ , or ii)  $Y_1 = (\Delta \cap \mathcal{B}_1, \Gamma \cap \mathcal{B}_1, \alpha) \in \mathcal{A}_1$ , or iii)  $Y_2 = (\Delta \cap \mathcal{B}_2, \Gamma \cap \mathcal{B}_2, \lambda) \in \mathcal{A}_2$ .

2.1) Suppose we are in cases i) or ii).  $\Gamma \cap \mathcal{B}_1 \subseteq \text{prem}(E)$  and by redundance-coherence of  $\prec$ ,  $Y \not\prec X'$  implies  $Y_1 \not\prec X'$ . Hence  $Y_1 \not\prec_1 X'$  and  $Y_1 \Rightarrow_{E \subseteq \mathcal{A}_1} X$ . Given A1,  $\exists Z \in E \subseteq E'$ ,  $Z \Rightarrow_{\{Y_1\}} Y_1$ . Since  $\text{prem}(Y_1) \subseteq \text{prem}(Y)$ ,  $Z \Rightarrow_{\{Y\}} Y$ .

2.2) Suppose we are in case iii). By P2,  $Y'_2 = (\Gamma \cap \mathcal{B}_2, \Delta \cap \mathcal{B}_2, \lambda) \in \mathcal{A}_2$ .  $\Gamma \cap \mathcal{B}_2 \subseteq \text{prem}(E' \setminus E)$  and so by A2,  $Y'_2 \in E'$ . By Lemma 27:

$Y'_2 \Rightarrow_{\{Y_2\}} Y_2$  and (since  $\text{prem}(Y_2) \subseteq \text{prem}(Y)$ ) the result is shown for  $Z = Y'_2$ , or

$Y_2 \Rightarrow_{\{Y'_2\}} Y'_2$ , in which case, since  $Y'_2$  is acceptable w.r.t.  $E'$  (by A2), then  $\exists Z \in E'$ ,  $Z \Rightarrow_{\{Y_2\}} Y_2$ , hence  $Z \Rightarrow_{\{Y\}} Y$ .  $\square$

**Corollary 38.** Lemma 37 holds for non-redundant *pdAFs*  $(\mathcal{A}, \mathcal{D})$ .

**Proof.** 1) is shown as in Lemma 37. In 2),  $Y \notin \mathcal{A}_1$  implies  $(\Delta \cap \mathcal{B}_2) \cup (\Gamma \cap \mathcal{B}_2) \neq \emptyset$ . Suppose  $(\Delta \cap \mathcal{B}_1) \cup (\Gamma \cap \mathcal{B}_1) \neq \emptyset$ . Then  $(\Delta \cap \mathcal{B}_2) \cup (\Gamma \cap \mathcal{B}_2) \parallel (\Delta \cap \mathcal{B}_1) \cup (\Gamma \cap \mathcal{B}_1) \cup \{\phi\}$ , contradicting  $(\mathcal{A}, \mathcal{D})$  is non-redundant (recall Remark 33). Hence  $\text{assumptions}(Y)$  is a non-empty subset of  $\mathcal{B}_2$  and  $Y \in \mathcal{A}_2$ . If  $\phi = \alpha$ ,<sup>20</sup> then since  $\text{assumptions}(Y) \parallel \{\alpha\}$ , by P4'  $(\Delta, \Gamma, \lambda) \in \mathcal{A}_2$ . We then reason as in 2.2) in Lemma 37, which does not assume redundance-coherent preference relations, to conclude  $\exists Z \in E'$ ,  $Z \Rightarrow_{\{Y\}} Y$ .  $\square$

**Lemma 39.** Let  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ . Then:

If  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$  then  $\exists E' \supseteq E$  s.t.  $E'$  is a complete extension of  $(\mathcal{A}, \mathcal{D})$  and  $E = E' \cap \mathcal{A}_1$

**Proof.** Suppose  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ . Let  $E' \supseteq E$ , such that  $E = E' \cap \mathcal{A}_1$ , and let us assume:

$E' \setminus E = \{Y \in \mathcal{A} \setminus \mathcal{A}_1 \mid Y \text{ is acceptable w.r.t. } E'\}$  (Ass1).

(That there exists such an  $E'$  with  $E' \setminus E \neq \emptyset$ , can be seen by the limiting case in which  $E' \setminus E$  is the set of unassailable arguments in  $\mathcal{A} \setminus \mathcal{A}_1$ ; these are acceptable w.r.t. any  $E' \subseteq \mathcal{A}$ .)

Firstly, suppose  $\Delta \subseteq \text{prem}(E' \setminus E)$ ,  $\text{prem}(V) = \Delta$ . Hence  $V \in \mathcal{A} \setminus \mathcal{A}_1$ . Suppose  $Z \Rightarrow_{E'} V$  on some  $\beta$ . Since  $\text{prem}(V) \subseteq \text{prem}(E' \setminus E)$ , then for some  $Y \in E' \setminus E$ ,  $Z \Rightarrow_{E'} Y$  on  $\beta$ . By Ass1,  $Y$  is acceptable w.r.t.  $E'$  and so  $\exists W \in E'$ ,  $W \Rightarrow_{\{Z\}} Z$ . Hence  $V$  is acceptable w.r.t.  $E'$ . Hence:

$$\forall \Delta \subseteq \text{prem}(E' \setminus E): (\text{prem}(A) = \Delta) \rightarrow A \in E' \text{ and } A \text{ acceptable w.r.t. } E' \quad (5)$$

<sup>20</sup> Note that in fact, it cannot be that  $\phi = \lambda$  since this would mean  $\text{prem}(X) \cap \Gamma \neq \emptyset$ , contradicting  $\text{assumptions}(Y) \subseteq \mathcal{B}_2$ .

We show  $X \in E$  implies  $X$  acceptable w.r.t.  $E'$ . Let  $Y \in \mathcal{A}$ ,  $Y \Rightarrow_{E'} X$ . Since  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ ,  $X$  is acceptable w.r.t.  $E$  in  $(\mathcal{A}_1, \mathcal{D}_1)$ . Hence, given Eq. (5), we can apply [Lemma 37](#) to conclude that  $\exists Z \in E'$  s.t.  $Z \Rightarrow_{\{Y\}} Y$ . Hence  $X$  acceptable w.r.t.  $E'$ .

Given Ass1 we now have that all arguments in  $E'$  are acceptable w.r.t.  $E'$ . Hence by [Lemma 17](#),  $E'$  is conflict free. To show  $E'$  is complete, it then remains to show (given Ass1) that  $X \in \mathcal{A}_1$ ,  $X \notin E'$  implies  $X$  is not acceptable w.r.t.  $E'$ . Suppose some such  $X$ . Since  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ ,  $X$  is not acceptable w.r.t.  $E$ . Hence:

$$\exists Y \in \mathcal{A}_1, Y \Rightarrow_E X, \neg \exists Z \in E \text{ s.t. } Z \Rightarrow_{\{Y\}} Y. \quad (6)$$

Since  $E \subseteq E'$ ,  $Y \Rightarrow_{E'} X$ . Suppose for contradiction that  $X$  is acceptable w.r.t.  $E'$  in  $(\mathcal{A}, \mathcal{D})$ . Hence:

$$\exists Z = (\Delta, \Gamma, \phi) \in E' \text{ s.t. } Z \Rightarrow_{\{Y\}} Y \text{ on } Y' = (\{\bar{\alpha}\}, \emptyset, \bar{\alpha}) \text{ (and so } Z \not\prec Y') \quad (7)$$

1) Suppose  $Z \in \mathcal{A}_1$ . Hence  $Z \in E$ .  $Z \not\prec Y'$  implies  $Z \not\prec_1 Y'$ , and so  $Z \Rightarrow_{\{Y\}} Y$  in  $(\mathcal{A}_1, \mathcal{D}_1)$ , contradicting Eq. (6).

2) Suppose  $Z \in \mathcal{A} \setminus \mathcal{A}_1$ . Since  $\Gamma \subseteq \text{prem}(Y)$ ,  $Y \in \mathcal{A}_1$ , it must be that  $\Gamma \subseteq \mathcal{B}_1$ . Hence by [Lemma 36](#) either: i)  $Z_1 = (\Delta \cap \mathcal{B}_1, \Gamma, \lambda) \in \mathcal{A}_1$ , or ii)  $Z_1 = (\Delta \cap \mathcal{B}_1, \Gamma, \alpha) \in \mathcal{A}_1$ , or iii)  $Z_2 = (\Delta \cap \mathcal{B}_2, \emptyset, \lambda) \in \mathcal{A}_2$ .

2.1) In case i) or ii), since  $(\Delta \cap \mathcal{B}_1) \subseteq \text{prem}(E)$ , and  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ , then by [Lemma 9](#),  $Z_1 \in E$ . Since  $Z \not\prec Y'$ , and  $<$  is redundancy-coherent,  $Z_1 \not\prec Y'$ . Hence  $Z_1 \not\prec_1 Y'$ , and  $Z_1 \Rightarrow_{\{Y\}} Y$ , contradicting Eq. (6).

2.2) In case iii). By Eq. (5),  $Z_2 \in E'$ . But then  $(\emptyset, \Delta \cap \mathcal{B}_2, \lambda) \Rightarrow_{E'} Z_2$ , contradicting  $Z_2$  acceptable w.r.t.  $E'$ .  $\square$

**Corollary 40.** *Lemma 39 holds for non-redundant pdAFs.*

**Proof.** Proof proceeds exactly as in [Lemma 39](#), except that:

- ‘Corollary 38’ replaces ‘Lemma 37’ (underlined in the proof of [Lemma 39](#));
- In 2) in the proof of [Lemma 39](#), we show as in [Corollary 38](#) that  $Z \in \mathcal{A}_2$  (substituting ‘ $Z \in \mathcal{A}_2$ ’ for ‘ $Y \in \mathcal{A}_2$ ’ in [Corollary 38](#)) and  $(\Delta, \emptyset, \lambda) \in \mathcal{A}_2$ . We then only need reason to a contradiction as in 2.2) of [Lemma 39](#), which does *not* assume redundancy-coherent preference relations.  $\square$

**Lemma 41.** *Let  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ . Then if  $E'$  is a complete extension of  $(\mathcal{A}, \mathcal{D})$  then  $E = E' \cap \mathcal{A}_1$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ .*

**Proof.** Suppose  $E'$  is a complete extension of  $(\mathcal{A}, \mathcal{D})$ .

**Part 1** Since  $E'$  is conflict free then  $E = E' \cap \mathcal{A}_1$  is conflict free.<sup>21</sup> Suppose for contradiction that  $E = E' \cap \mathcal{A}_1$  is *not* an admissible extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ . Hence  $\exists X \in E$ ,  $Y \in \mathcal{A}_1$ ,  $Y \Rightarrow_E X$  on  $X'$  (hence  $Y \not\prec_1 X'$ ), and:

$$\neg \exists Z \in E \text{ s.t. } Z \Rightarrow_{\{Y\}} Y. \quad (8)$$

Since  $Y \not\prec X'$ , and  $\text{prem}(E) \subseteq \text{prem}(E')$ , then  $Y \Rightarrow_{E'} X$ . By assumption of  $E'$  being a complete extension of  $(\mathcal{A}, \mathcal{D})$ :

$$Z = (\Delta, \Gamma, \phi) \in E' \text{ s.t. } Z \Rightarrow_{\{Y\}} Y \text{ on } Y' = (\{\bar{\alpha}\}, \emptyset, \bar{\alpha}), Z \not\prec Y'.$$

1) Suppose  $Z \in \mathcal{A}_1$ . Then  $Z \in E$ . Since  $Z \not\prec_1 Y'$ ,  $Z \Rightarrow_{\{Y\}} Y$ , contradicting Eq. (8).

2) Suppose  $Z \in \mathcal{A} \setminus \mathcal{A}_1$ . Since  $\Gamma \subseteq \text{prem}(Y)$ ,  $Y \in \mathcal{A}_1$ , it must be that  $\Gamma \subseteq \mathcal{B}_1$ . We then reason to a contradiction as in 2) in the proof of [Lemma 39](#), except that:

in 2.1), we have that since  $(\Delta \cap \mathcal{B}_1) \subseteq \text{prem}(E')$  and  $E'$  is a complete extension of  $(\mathcal{A}, \mathcal{D})$ , then by [Lemma 9](#)  $Z_1 \in E'$ , hence  $Z_1 \in E$ ;

in 2.2),  $Z_2 \in E'$  by virtue of  $E'$  being complete and [Lemma 9](#).

**Part 2** We have shown that  $E$  is admissible. Suppose for contradiction that  $E$  is not complete. Hence  $\exists X \notin E$  that is acceptable w.r.t.  $E$ . We show  $X$  is acceptable w.r.t.  $E'$ . Suppose:

$$Y = (\Delta, \Gamma, \phi), Y \Rightarrow_{E'} X \text{ on } X' = (\{\bar{\alpha}\}, \emptyset, \bar{\alpha}), \text{ hence } Y \not\prec X'.$$

Since  $E'$  is complete, then (by [Lemma 9](#)):

$$\forall \Delta \subseteq \text{prem}(E' \setminus E) : (\text{prem}(A) = \Delta) \rightarrow A \in E' \text{ and } A \text{ acceptable w.r.t. } E' \quad (9)$$

Given Eq. (9) and  $X$  is acceptable w.r.t.  $E$ , we can apply [Lemma 37](#) to conclude that  $\exists Z \in E'$  s.t.  $Z \Rightarrow_{\{Y\}} Y$ . Hence  $X$  is acceptable w.r.t.  $E'$ . Since  $E'$  is complete,  $X \in E'$ . Hence  $X \in E = E' \cap \mathcal{A}_1$ . Contradiction.  $\square$

<sup>21</sup> Trivially,  $\forall E, E' \text{ s.t. } E \subseteq E', X \Rightarrow_E Y \text{ implies } X \Rightarrow_{E'} Y$ . Hence suppose  $E'$  is conflict free and  $E$  is not conflict free. Then  $X, Y \in E \subseteq E', X \Rightarrow_E Y$  and so  $X \Rightarrow_{E'} Y$ , contradicting  $E'$  is conflict free.

**Corollary 42.** Lemma 41 holds for non-redundant *pdAFs*.

**Proof.** Proof as in Lemma 41, except:

- ‘Corollary 38’ replaces ‘Lemma 37’ (underlined);
- In 2) in Part 1 of the proof of Lemma 41, we show as in Corollary 38 that  $Z \in \mathcal{A}_2$  (substituting ‘ $Z \in \mathcal{A}_2$ ’ for ‘ $Y \in \mathcal{A}_2$ ’ in Corollary 38) and  $(\Delta, \emptyset, \lambda) \in \mathcal{A}_2$ . Since  $(\Delta, \emptyset, \lambda) \in E'$  by virtue of  $E'$  being complete and Lemma 9, this leads to a contradiction as in 2.2) in the proof of Lemma 39.  $\square$

The following is a corollary of Lemmas 39 and 41 and Corollaries 40 and 42:

**Corollary 43.** Let  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ , where  $(\mathcal{A}, \mathcal{D})$ ,  $(\mathcal{A}_1, \mathcal{D}_1)$  and  $(\mathcal{A}_2, \mathcal{D}_2)$  are (non-redundant) *pdAFs*. Then:

1. If  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$  then there exist complete extensions  $\{E_1, \dots, E_n\}$  of  $(\mathcal{A}, \mathcal{D})$  such that for  $i = 1 \dots n$ ,  $E = E_i \cap \mathcal{A}_1$ .
2. If  $\{E_1, \dots, E_n\}$  are complete extensions of  $(\mathcal{A}, \mathcal{D})$  such that for  $i = 1 \dots n$ ,  $E = E_i \cap \mathcal{A}_1$ , then  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$ .

That there is a one to (possibly) many mapping between complete extensions of  $(\mathcal{A}_1, \mathcal{D}_1)$  and  $(\mathcal{A}, \mathcal{D})$  is witnessed by an example in which  $\mathcal{B}_1 = \{a\}$ ,  $\mathcal{B}_2 = \{b, \neg b\}$ ,  $\mathcal{B} = \{a, b, \neg b\}$ , and  $\prec_1 = \prec_2 = \prec = \emptyset$ . Then  $E$  is a single complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$  containing  $A = (\{a\}, \emptyset, a)$  and  $E_1$  and  $E_2$  are the complete extensions of  $(\mathcal{A}, \mathcal{D})$ , where  $E_1$  contains  $A$  and  $(\{b\}, \emptyset, b)$ , and  $E_2$  contains  $A$  and  $(\{\neg b\}, \emptyset, \neg b)$ .

Note that Corollary 43 immediately implies that:

1.  $X$  is credulously justified under the complete semantics in  $(\mathcal{A}_1, \mathcal{D}_1)$  iff  $X$  is credulously justified under the complete semantics in  $(\mathcal{A}, \mathcal{D})$ .
2.  $X$  is sceptically justified under the complete semantics in  $(\mathcal{A}_1, \mathcal{D}_1)$  iff  $X$  is sceptically justified under the complete semantics in  $(\mathcal{A}, \mathcal{D})$ .

In [16,61], the non-interference and crash resistance postulates are formulated w.r.t. the ‘consequences’ of *propositional* instantiations of *AFs*. Our formalisation of symbols (Definition 4) and syntactic disjointedness for first order theories (Notation 3) allows us to generalise these postulates to *first order* instantiations of *AFs*, and show their satisfaction by (non-redundant) *pdAFs*.

**Definition 27** (Consequence relation for *pdAFs*). Let  $(\mathcal{A}, \mathcal{D})$  be a (non-redundant) *pdAF*. Then  $Cn_T((\mathcal{A}, \mathcal{D})) = \{\text{claims}(E_1), \dots, \text{claims}(E_n)\}$  where  $E_1, \dots, E_n$  are the  $T$  extensions of  $(\mathcal{A}, \mathcal{D})$ ,  $T \in \{\text{complete, grounded, preferred, stable}\}$ .

Referring to the above example, and writing  $Cn_{cp}$  as an abbreviation for  $Cn_{complete}$ ,  $Cn_{cp}((\mathcal{A}_1, \mathcal{D}_1)) = \{Cn(\{a\})\}$ <sup>22</sup> and  $Cn_{cp}((\mathcal{A}, \mathcal{D})) = \{Cn(\{a, b\}), Cn(\{a, \neg b\})\}$ . Recall that we define  $\text{claims}(E)$  as the conclusions of unconditional arguments that have empty suppositions. Also recall the notation  $\mathcal{B}_{|S_Y}$  in Notation 3.

**Definition 28** (Non interference). Let  $(\mathcal{A}_1, \mathcal{D}_1)$  be defined by  $\mathcal{B}_1$  and  $\prec_1$ ,  $(\mathcal{A}_2, \mathcal{D}_2)$  defined by  $\mathcal{B}_2$  and  $\prec_2$ , and  $(\mathcal{A}, \mathcal{D})$  defined by  $\mathcal{B}_1 \cup \mathcal{B}_2$  and  $\prec$ , where  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ , and  $\mathcal{B}_1 \parallel \mathcal{B}_2$ . Then non-interference is satisfied iff:

$$Cn_{cp}((\mathcal{A}_1, \mathcal{D}_1))_{| \text{symbols}(\mathcal{B}_1)} = Cn_{cp}((\mathcal{A}, \mathcal{D}))_{| \text{symbols}(\mathcal{B}_1)}$$

**Theorem 44** (Non interference). Non-interference is satisfied by (non-redundant) *pdAFs*.

**Proof.** By Corollary 43:  $E$  is a complete extension of  $(\mathcal{A}_1, \mathcal{D}_1)$  iff  $\{E_1, \dots, E_n\}$  are complete extensions of  $(\mathcal{A}, \mathcal{D})$ , where for  $i = 1 \dots n$ ,  $E = E_i \cap \mathcal{A}_1$ . Then assuming unconditional arguments  $X$  s.t.  $\text{symbols}(\text{conclusion}(X)) \subseteq \text{symbols}(\mathcal{B}_1)$ :

- 1) If  $X \in E$  then  $\forall E_i$ ,  $X \in E_i$ .
- 2) If  $X \in E_i$ ,  $X \in \mathcal{A} \cap \mathcal{A}_1$ , then for  $j = 1 \dots n$ ,  $X \in E_j$  and  $X \in E$ .
- 3) If  $X = (\Delta, \emptyset, \alpha) \in E_i$ ,  $X \in \mathcal{A} \setminus \mathcal{A}_1$ , then  $\Delta = \Pi \cup \Sigma$ , where  $\Pi \subseteq \mathcal{B}_1$ ,  $\Sigma \subseteq \mathcal{B}_2$ ,  $\Sigma \neq \emptyset$  (since if  $\Sigma = \emptyset$  then given Definition 26.2,  $X \in \mathcal{A}_1$ ) and so  $\Sigma \parallel \Pi \cup \{\alpha\}$ .

• Assuming arbitrary *pdAFs*, then by P4, either:

- 3.1)  $X' = (\Pi, \emptyset, \alpha) \in \mathcal{A}$  and so (by Definition 26.2),  $X' \in \mathcal{A}_1$ . By Lemma 9,  $X' \in E_i$ , and we are in case 2), or
- 3.2)  $X' = (\Sigma, \emptyset, \lambda) \in \mathcal{A}$ . By Lemma 9,  $X' \in E_i$ , contradicting Theorem 11.

• Assuming non-redundant *pdAFs*, then it must be that  $\Pi = \emptyset$  (recall Remark 33), and so by P4',  $X' = (\Sigma, \emptyset, \lambda) \in \mathcal{A}$ , and we are in case 3.2).

<sup>22</sup> Where ‘ $Cn$ ’ is the classical consequence relation.

Cases 2) & 3) show that if an unconditional  $X$  is in some  $E_i$  and  $\text{symbols}(\alpha) \subseteq \text{symbols}(\mathcal{B}_1)$ , then for  $j = 1 \dots n$ , there is an  $X' \in E_j$  where  $X'$  has conclusion  $\alpha$ , and  $X' \in E$ . This, together with 1), establishes that:  $\text{claims}(E)_{|\text{symbols}(\mathcal{B}_1)} = \text{claims}(E_i)_{|\text{symbols}(\mathcal{B}_1)}$ , where for  $i = 1 \dots n$ ,  $E = E_i \cap \mathcal{A}_1$ . Hence the theorem is shown.  $\square$

In our example,  $\text{Cn}_{cp}((\mathcal{A}_1, \mathcal{D}_1))_{|\text{symbols}(\mathcal{B}_1)} = \{\{a\}\}$  and  $\text{Cn}_{cp}((\mathcal{A}, \mathcal{D})) = \{\{a\}\}$ .

We now prove crash resistance, adapting the strategy employed in [16,61]. We show that (non-redundant) *pdAFs* are non-trivial, and then show that this property, together with non-interference, implies crash resistance.

**Lemma 45.** *Let  $Sy$  be any set of predicate, function and constant symbols. Then there exist (non-redundant) *pdAFs*  $(\mathcal{A}_1, \mathcal{D}_1)$  and  $(\mathcal{A}_2, \mathcal{D}_2)$ , respectively defined by  $\mathcal{B}_1, <_1$  and  $\mathcal{B}_2, <_2$ , s.t.  $\text{symbols}(\mathcal{B}_1) = \text{symbols}(\mathcal{B}_2) = Sy$ , and  $\text{Cn}_{cp}((\mathcal{A}_1, \mathcal{D}_1))_{|Sy} \neq \text{Cn}_{cp}((\mathcal{A}_2, \mathcal{D}_2))_{|Sy}$ .*

**Proof.** Let  $\{P_1, \dots, P_n\}$  be the predicate symbols in  $Sy$ , and for  $i = 1 \dots n$ , let  $\alpha_i$  denote the atomic formula  $P_i(\vec{t}_i)$  where  $\vec{t}_i$  is a vector of 0 or more terms, and such that  $\bigcup_{i=1}^n \text{symbols}(\vec{t}_i) = \{t | t \text{ is a function or constant symbol in } Sy\}$ . Let  $\mathcal{B}_1 = \{\alpha_1, \dots, \alpha_n\}$ ,  $\mathcal{B}_2 = \{\alpha_1 \rightarrow \alpha_1, \dots, \alpha_n \rightarrow \alpha_n\}$ ,  $<_1 = <_2 = \emptyset$ . Then each *pdAF* has a single complete extension whose unconditional arguments have conclusions that differ:  $\text{Cn}_{cp}((\mathcal{A}_1, \mathcal{D}_1)) = \{\text{Cn}(\{\alpha_1, \dots, \alpha_n\})\} \neq \text{Cn}_{cp}((\mathcal{A}_2, \mathcal{D}_2)) = \{\text{Cn}(\{\alpha_1 \rightarrow \alpha_1, \dots, \alpha_n \rightarrow \alpha_n\})\}$ .  $\square$

**Definition 29** (Contaminating base). Let  $\mathcal{B}_1$  be a base such that  $\text{symbols}(\mathcal{B}_1) \subset \text{symbols}(\mathcal{L})$ .  $\mathcal{B}_1$  is said to be *contaminating* for (non-redundant) *pdAFs* iff:

there exists a  $(\mathcal{A}_1, \mathcal{D}_1)$  defined by  $\mathcal{B}_1, <_1$ , such that for any  $\mathcal{B}_2$  such that  $(\mathcal{A}_2, \mathcal{D}_2)$  is defined by  $\mathcal{B}_2, <_2$ , and  $\mathcal{B}_1 \parallel \mathcal{B}_2$ :  $\text{Cn}_{cp}((\mathcal{A}_1, \mathcal{D}_1)) = \text{Cn}_{cp}((\mathcal{A}, \mathcal{D}))$ , where  $(\mathcal{A}, \mathcal{D}) = (\mathcal{A}_1, \mathcal{D}_1) \oplus (\mathcal{A}_2, \mathcal{D}_2)$ .

**Theorem 46** (Crash resistance). *There does not exist a contaminating base  $\mathcal{B}$  for *pdAFs* and non-redundant *pdAFs*.*

**Proof.** Suppose for contradiction that there exists a contaminating base  $\mathcal{B}_1$ . Hence  $\text{symbols}(\mathcal{B}_1) \subset \text{symbols}(\mathcal{L})$ . Let  $Sy = \text{symbols}(\mathcal{L}) \setminus \text{symbols}(\mathcal{B}_1)$ . By Lemma 45, there exists a *pdAF*<sub>3</sub> and *pdAF*<sub>4</sub> respectively defined by  $\mathcal{B}_3, <_3$  and  $\mathcal{B}_4, <_4$ , such that  $\text{symbols}(\mathcal{B}_3) = \text{symbols}(\mathcal{B}_4) = Sy$ , and:

$$\text{Cn}_{cp}(\text{pdAF}_3)_{|Sy} \neq \text{Cn}_{cp}(\text{pdAF}_4)_{|Sy}. \quad (1)$$

Since  $\mathcal{B}_1$  is contaminating, then there is a *pdAF*<sub>1</sub> =  $(\mathcal{A}_1, \mathcal{D}_1)$  such that:

$$\text{Cn}_{cp}(\text{pdAF}_1) = \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_3) = \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_4)$$

Hence  $\text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_3)_{|Sy} = \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_4)_{|Sy}$ . This, together with (1) implies that  $\text{Cn}_{cp}(\text{pdAF}_3)_{|Sy} \neq \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_3)_{|Sy}$  or  $\text{Cn}_{cp}(\text{pdAF}_4)_{|Sy} \neq \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_4)_{|Sy}$ .

Since  $\text{symbols}(\mathcal{B}_3) = \text{symbols}(\mathcal{B}_4) = Sy$ , we immediately have that either  $\text{Cn}_{cp}(\text{pdAF}_3)_{|\mathcal{B}_3} \neq \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_3)_{|\mathcal{B}_3}$  or  $\text{Cn}_{cp}(\text{pdAF}_4)_{|\mathcal{B}_4} \neq \text{Cn}_{cp}(\text{pdAF}_1 \oplus \text{pdAF}_4)_{|\mathcal{B}_4}$ . In either case, Theorem 44 (non-interference) is violated.  $\square$

## 4. Instantiating dialectical classical frameworks

### 4.1. C-inteligm argumentation

We now instantiate *pdAFs* in which, in contrast with standard approaches to *Cl-Arg* [1,33,43], the proof theoretic means for constructing arguments is given. Indeed, we suggest that arguments conceived of as assumptions in support of a conclusion should be understood as ‘argument schemata’ that are concretely realised as ‘arguments proper’ through provision of the proof theoretic means by which the conclusion is inferred from the assumption. We argue that the persuasive force of a non-trivial argument partly depends on whether valid reasoning steps have been employed in inferring the argument’s conclusion. For example, a mathematical argument claiming that a certain theorem follows from the Euclidean axioms would be incomplete and entirely unpersuasive without explicit representation of the proof steps involved. It is the means by which the conclusion is obtained that renders the argument understandable, and furthermore, one would not want to rely on the recipient expending resources to reconstruct the proof from assumptions to conclusion (we revisit this issue in Section 5 when discussing the ‘transparency’ postulate proposed for practical applications [27]).

In this section we therefore ‘instantiate’ *pdAFs* with *propositional* natural deduction proofs. We present a non-standard version of *classical* natural deduction which, given some straightforward restrictions on the application of the rules, does not allow construction of R-contaminated arguments and admits definition of a simple notion of argument ‘depth’. The latter may be taken to reflect the inferential capabilities of resource-bounded agents. We then show that arguments constructed under any fixed bound on their depth, define non-redundant *pdAFs* that satisfy the rationality postulates. In dialectical frameworks we use the word “assumption” as a general term that refers to both premises and suppositions.



**Table 1**  
The intelim rules.

INTRODUCTION RULES			
$\frac{\varphi \quad \psi}{\varphi \wedge \psi} \wedge \mathcal{I}$	$\frac{\neg \varphi}{\neg(\varphi \wedge \psi)} \neg \wedge \mathcal{I}1$	$\frac{\neg \psi}{\neg(\varphi \wedge \psi)} \neg \wedge \mathcal{I}2$	
$\frac{\neg \varphi \quad \neg \psi}{\neg(\varphi \vee \psi)} \neg \vee \mathcal{I}$	$\frac{\varphi}{\varphi \vee \psi} \vee \mathcal{I}1$	$\frac{\psi}{\varphi \vee \psi} \vee \mathcal{I}2$	
$\frac{\varphi \quad \neg \psi}{\neg(\varphi \rightarrow \psi)} \neg \rightarrow \mathcal{I}$	$\frac{\neg \varphi}{\varphi \rightarrow \psi} \rightarrow \mathcal{I}1$	$\frac{\psi}{\varphi \rightarrow \psi} \rightarrow \mathcal{I}2$	
$\frac{\varphi}{\neg \neg \varphi} \neg \neg \mathcal{I}$			
ELIMINATION RULES			
$\frac{\varphi \vee \psi \quad \neg \varphi}{\psi} \vee \mathcal{E}1$	$\frac{\varphi \vee \psi \quad \neg \psi}{\varphi} \vee \mathcal{E}2$	$\frac{\neg(\varphi \vee \psi)}{\neg \varphi} \neg \vee \mathcal{E}1$	$\frac{\neg(\varphi \vee \psi)}{\neg \psi} \neg \vee \mathcal{E}2$
$\frac{\neg(\varphi \wedge \psi) \quad \varphi}{\neg \psi} \neg \wedge \mathcal{E}1$	$\frac{\neg(\varphi \wedge \psi) \quad \psi}{\neg \varphi} \neg \wedge \mathcal{E}2$	$\frac{\varphi \wedge \psi}{\varphi} \wedge \mathcal{E}1$	$\frac{\varphi \wedge \psi}{\psi} \wedge \mathcal{E}2$
$\frac{\varphi \rightarrow \psi \quad \varphi}{\psi} \rightarrow \mathcal{E}1$	$\frac{\varphi \rightarrow \psi \quad \neg \psi}{\neg \varphi} \rightarrow \mathcal{E}2$	$\frac{\neg(\varphi \rightarrow \psi)}{\varphi} \neg \rightarrow \mathcal{E}1$	$\frac{\neg(\varphi \rightarrow \psi)}{\neg \psi} \neg \rightarrow \mathcal{E}2$
$\frac{\neg \neg \varphi}{\varphi} \neg \neg \mathcal{E}$			

**Table 2**  
Falsum rules and RB.

$\frac{\varphi \quad \neg \varphi}{\perp} \text{RNC}$	$\frac{\perp}{\varphi} \text{XFQ}$	$\frac{\Gamma, [\varphi] \quad \mathcal{D}_1}{\psi}$	$\frac{\Delta, [\neg \varphi] \quad \mathcal{D}_2}{\psi}$
		$\frac{\psi}{\psi} \text{RB}$	

#### 4.1.1. C-intelim natural deduction

Natural deduction proofs apply intuitive introduction and elimination rules (or “intelim rules” for short) that are akin to natural modes of human reasoning (in contrast to other proof theories e.g., axiomatic systems, Gentzen-style sequent calculi or resolution), and are thus particularly appropriate if one is to simulate human understanding and reasoning.<sup>23</sup>

We use a non-standard version of classical natural deduction [20,22,23], that we call “C-intelim” (for “classical intelim”). As discussed in [20,24], this version is more faithful to the intuitive classical meaning of the logical operators and naturally suggests a simple measure of the “depth” of an argument. This is used to define a hierarchy of tractable, albeit increasingly complex, approximations to classical propositional logic that converge to it in the limit (for ideal “unbounded” agents). Each level in the hierarchy equates with increments in the maximum depth of the arguments that can be constructed and the resources required to construct such arguments. Thus, C-intelim proofs provide a perspicuous account of argumentation for resource-bounded agents employing natural rules that are faithful to the classical interpretation of the logical operators.

The C-intelim rules are displayed in Table 1. Besides the intelim rules we have two extra rules for the “falsum” constant  $\perp$  (the first two displayed in Table 2), representing a sentence which is false in every possible world. RNC (“Rule of Non-Contradiction”) simply says that two contradictory sentences cannot be both true, and the XFQ (*Ex Falso Quodlibet*) rule says that every sentence follows from  $\perp$ . The two falsum rules taken together imply that any arbitrary conclusion follows from a contradiction.

<sup>23</sup> For a recent contribution that makes use of a restricted version of standard Gentzen-style natural deduction in the context of formal argumentation theory see [36].

The *intelim* and *falsum* rules contain *no discharge rules*, namely rules that involve the temporary introduction of assumptions that are subsequently “discharged”, to the effect that their conclusion no longer depends on them.<sup>24</sup> To obtain a complete set of rules for classical propositional logic we only need add a *single* discharge rule: if we have a deduction  $\mathcal{D}_1$  of  $\psi$  depending on assumptions  $\Gamma \cup \{\varphi\}$  and a deduction  $\mathcal{D}_2$  of  $\psi$  depending on assumptions  $\Delta \cup \{\neg\varphi\}$ , we thereby have a deduction of  $\psi$  depending on  $\Gamma \cup \Delta$ . This typical pattern of classical case reasoning relies on the Aristotelian principle of bivalence; a cornerstone of classical semantics whereby any sentence is either true or false, and there are no other possibilities. Hence we call this pattern *Rule of Bivalence* (RB), which is the third rule displayed in Table 2 where the conclusion  $\psi$  does not depend on the “discharged” assumptions  $\varphi$  and  $\neg\varphi$  that are enclosed in square brackets.

C-intelim proofs can be presented using the standard natural deduction tree format in which the conclusion is at the root and the assumptions at the leaves. However this format involves a good deal of redundancy,<sup>25</sup> and we therefore resort to a different more concise format that is better suited to algorithmic treatment.<sup>26</sup> In this new format proofs are still represented by trees, but these trees branch downwards (like Smullyan’s Tableaux). The premises of each application of the *intelim* or the *falsum* rules do not occur on adjacent branches, but on the same branch as the conclusion, and anywhere above it. So, the application of these rules is *sequential*, and we call an *intelim sequence* for  $\Gamma$  any sequence of formulae generated in this way. On the other hand, each application of RB splits an *intelim sequence* into two branches: one containing a formula  $\varphi$  and the other its negation  $\neg\varphi$ , as in the examples displayed in Fig. 4. In this case we say that RB has been *applied to*  $\varphi$ , and  $\varphi$  is the *RB-formula* of this RB application. Every application of RB introduces, on each branch, an extra assumption, that we call *virtual* to distinguish it from the *actual* assumptions that are usually listed at the top, starting from the root. For example, in the leftmost branch of the first tree in Fig. 4, the formula  $\neg q$  is obtained from the actual assumption  $p \rightarrow \neg q$  at the top of the tree and from the virtual assumption  $p$  by means of an application of the rule  $\rightarrow \mathcal{E}1$ ; and  $r$  is obtained from the derived formula  $\neg q$  and the actual assumption  $q \vee r$  by means of an application of  $\vee \mathcal{E}1$ . [22,23] argue that the minimum number of nested applications of RB required to develop a deductive argument, provides a natural and plausible measure of the ‘difficulty’ involved in constructing it.

In the sequel we shall use the expression “tree of formulae” as an abbreviation of “tree whose nodes (except possibly the root) are labelled with formulae”. The special case when the root is unlabelled will be used to represent proofs from the empty set of assumptions (starting with an application of the RB rule).

**Definition 30** (*C-intelim trees, proofs, refutations*). A *C-intelim tree* based on a set  $\Gamma$  is a tree of formulae such that  $\Gamma$  is the set of all its actual assumptions and every other node results from the application of one of the C-intelim rules. A *C-intelim proof* of  $\varphi$  depending on  $\Gamma$  is a C-intelim tree based on  $\Gamma$  such that (i)  $\varphi$  occurs as the final formula in every branch and (ii) all actual assumptions in  $\Gamma$  are used as premises of some rule application. A *C-intelim refutation* of  $\Gamma$  is a C-intelim proof of  $\perp$  depending on  $\Gamma$ .

C-intelim is sound and complete for classical propositional logic: if  $\varphi$  is a classical consequence of  $\Gamma$ , then there is a C-intelim proof of  $\varphi$  depending on some  $\Delta \subseteq \Gamma$ . Fig. 4 shows examples of C-intelim proof trees. The actual assumptions are marked with a \*.

A crucial role for practical applications in classical logic argumentation is played by *normal* C-Intelim proofs. These normal proofs can be generated by straightforward restrictions on the applications of the rules that (i) limit the choice of the RB-formula of an RB-application to a tractable space defined by the assumptions and the conclusion and (ii) avoid obvious redundancies in the construction of a proof. More specifically normal C-intelim proofs: (a) enjoy the (weak) *subformula property*,<sup>27</sup> which makes their construction amenable to algorithmic treatment, and (b) do not allow the construction of R-contaminated arguments (as described in Section 3.6, Definition 22).

It can be shown (see [24]) that restricting to normal C-intelim proofs involves no loss of deductive power in that for each C-intelim proof there exists a normal one with the same assumptions and the same conclusion. The simple restrictions on the construction of normal C-intelim proofs ensure that such proofs are not trivially redundant (see [24] for a discussion). This allows us to bar convoluted proofs such as those shown on the left in each of Fig. 5 (a), (b), (c) and (d) and to generate the normal ones on the right instead. In particular, notice the redundant proof in Fig. 5(a), which makes contrived use of an assumption ( $\neg q$ ) that is syntactically disjoint from the other assumptions and from the conclusion.

#### 4.1.2. Depth-bounded instantiations of pDCFs

One of the main advantages of C-intelim deduction is that, unlike standard natural deduction, it immediately provides a sharp measure of the *depth* of an argument that can be associated with the resources required for its construction.

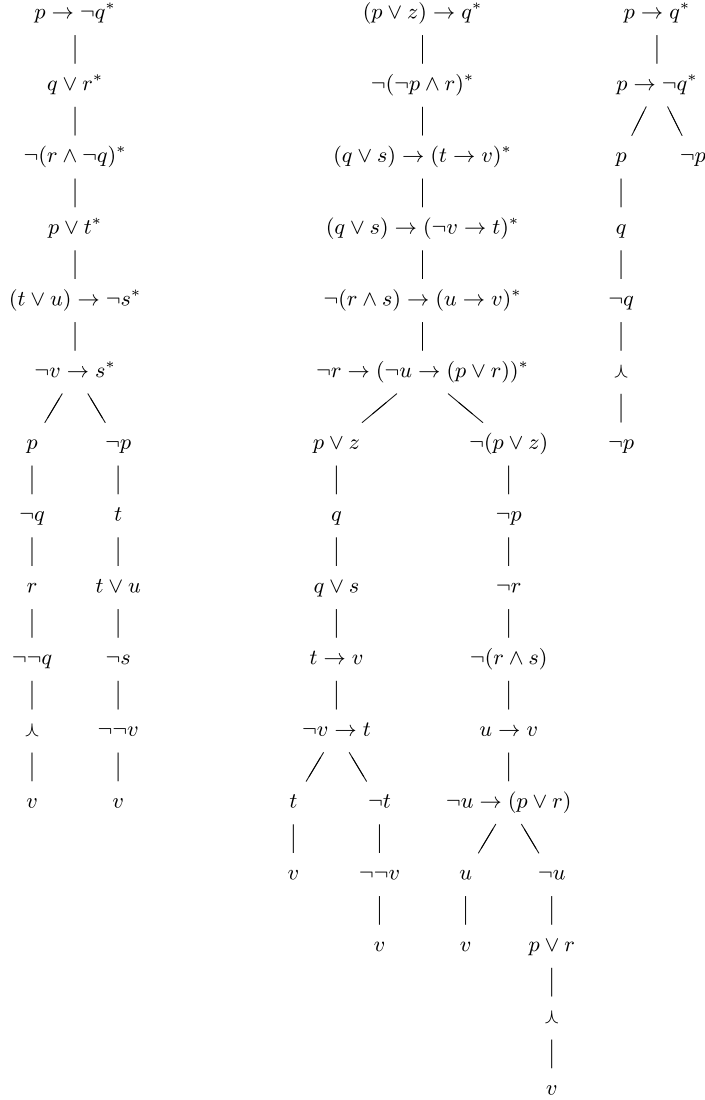
In the sequel we shall restrict our attention to *normal* C-intelim proofs.

<sup>24</sup> By contrast, in Gentzen-style natural deduction,  $\vee$  elimination and  $\rightarrow$ -introduction are discharge rules.

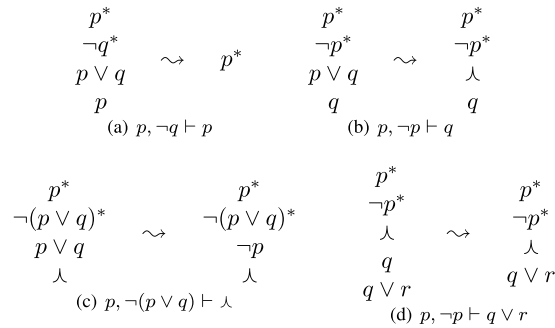
<sup>25</sup> Whenever a formula inferred from assumptions is used more than once as a premise of further inferences, its proof tree has to be replicated.

<sup>26</sup> See [24, Section 8].

<sup>27</sup> Every formula occurring in the proof is a subformula of the premises or of the conclusion, or the negation of such a subformula.



**Fig. 4.** C-intelim proofs. Actual assumptions are marked with "\*".



**Fig. 5.** Convoluted proofs are shown on the left, and their non-convoluted versions are shown on the right. Connection links between nodes have been removed for ease of presentation.

**Definition 31** (*Depth of C-intelim proofs*). The *depth* of a C-intelim proof  $\mathcal{T}$  is the maximum number of virtual assumptions occurring in a branch of  $\mathcal{T}$ .

A C-intelim proof of depth 0 contains no virtual assumptions, i.e., no application of RB (and is therefore a C-intelim sequence). A C-intelim proof of depth  $k$  contains at most  $k$  nested applications of RB. Consider the rightmost C-intelim proof in Fig. 4 which makes use of one application of RB. A reasoning agent (whether human or artificial) with actual assumptions  $p \rightarrow q$  and  $p \rightarrow \neg q$  considered true, does not possess the information that  $\neg p$  is true, unless it is able to simulate information states containing virtual assumptions. This is a non-trivial step both from the computational (and arguably cognitive) viewpoint, and the depth at which the iterated use of such virtual assumptions is required is an interesting measure of the computational effort involved in extracting the information ( $\neg p$ ) implicitly contained in the actual assumptions. Let C-Intelim $_k$  be the system allowing only (normal) C-intelim proofs of depth  $\leq k$ .

**Definition 32** (C-Intelim $_k$ ). Let  $\vdash_k \subseteq 2^{\mathcal{L}} \times \mathcal{L}$  be defined as follows:

$\Gamma \vdash_k \varphi$  if and only if there is a C-intelim $_k$  proof of  $\varphi$  depending on some  $\Delta \subseteq \Gamma$ .

**Proposition 47.** For each  $k \in \mathbb{N}$ , whether or not  $\Gamma \vdash_k \varphi$  can be decided in time  $O(n^{k+2})$ , where  $n$  is the total number of occurrences of symbols in  $\Gamma \cup \{\varphi\}$ .

For a proof see [23].

The relations  $\vdash_k$  of Definition 32 form a sequence of increasingly powerful deducibility relations that are all (by Proposition 47) tractable and converge to classical propositional logic, which is the limit of the sequence for  $k$  approaching infinity.

**Definition 33** ( $k$ -depth C-intelim arguments). Let  $\mathcal{T}$  be any normal C-intelim $_k$  proof of  $\varphi$  depending on  $\Sigma$  (recall – Definition 30 – that  $\Sigma$  are the actual, and not virtual assumptions in  $\mathcal{T}$ ), where  $\forall \alpha \in \Sigma$ ,  $\alpha$  is annotated in  $\mathcal{T}$  by  $\textcircled{D}$  or  $\textcircled{S}$ , where  $\textcircled{D}$  stands for “premise” and  $\textcircled{S}$  stands for “supposition”.

Then  $(\Delta, \Gamma, \varphi)$  is a C-intelim $_k$  argument, where  $\Delta \cup \Gamma = \Sigma$ ,  $\alpha \in \Delta$  iff  $\alpha$  is annotated by  $\textcircled{D}$ , and  $\alpha \in \Gamma$  iff  $\alpha$  is annotated by  $\textcircled{S}$ . We call  $(\Delta, \Gamma, \varphi)$  the C-intelim $_k$  argument associated with  $\mathcal{T}$ .

A set  $\mathcal{A}$  of C-intelim $_k$  arguments is defined by  $\mathcal{B}$  if  $\forall X \in \mathcal{A}$ ,  $\text{assumptions}(X) \subseteq \mathcal{B}$ .

The following key result is shown in [24]:

**Proposition 48.** If  $\mathcal{T}$  is a normal C-intelim $_k$  proof of  $\varphi$  depending on  $\Gamma$ , then the C-intelim $_k$  argument associated with  $\mathcal{T}$  is not R-contaminated.

Moreover [24] shows that if an argument associated with  $\mathcal{T}$  is contaminated, and so given Proposition 48, can only by E-contaminated (e.g. a proof of  $q$  from  $p, \neg p$ ), then:

**Proposition 49.** If an argument associated with a normal C-intelim $_k$  proof  $\mathcal{T}$  depending on  $\Gamma$  is E-contaminated, then the proof is said to be improper in that every branch in the downward-branching tree ends with an application of XFQ (and so  $\Gamma \vdash_c \perp$ ).

**Remark 50.** [24] shows that it is easy to turn an improper proof into a proof of  $\perp$  (i.e., a refutation) of the same depth.

Proposition 48 shows that normal proofs satisfy a basic relevance requirement, which is, however, weaker than the standard requirement that an argument's assumptions are subset-minimal. For example, recalling the discussion in Section 3.2, and in particular Footnote 8 on p. 22, one could build an argument for  $q$  depending on  $\{p, p \rightarrow r, p \rightarrow q, r \rightarrow ((p \rightarrow q) \rightarrow q)\}$  which is not R-contaminated and in which all the assumptions are actually used (i.e., it is relevant in the sense of relevance logic), although it is not subset-minimal.

We now formally instantiate non-redundant  $pdAF$ s (recall Definition 25) with C-intelim $_k$  arguments. Firstly, we show that given a non-redundant  $pdAF$  instantiated by C-intelim $_k$  arguments, then the conditions P1–P4' and P5 on the arguments in the  $pdAF$  can be satisfied. This follows from the following result:

**Proposition 51.** Let  $\mathcal{A}$  be any set of C-intelim $_k$  arguments defined by  $\mathcal{B}$ . Then there exists a set  $\mathcal{A}' \supseteq \mathcal{A}$  of C-intelim $_k$  arguments defined by  $\mathcal{B}$  such that  $\mathcal{A}'$  satisfies P1–P4' and P5.

**Proof.** Proposition 48 states that P5 is satisfied by any set of C-intelim $_k$  arguments. P1 is trivially satisfied, since for any depth  $n \geq 0$ ,  $\alpha \in \mathcal{B}$  implies  $(\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}'$ . P2 is trivially satisfied given that logically equivalent arguments are of equal depth. P3 is satisfied given the following result:

Let  $(\Delta, \emptyset, \alpha)$  and  $(\Gamma, \emptyset, \bar{\alpha})$  be C-intelim $_k$  arguments. Then either  $(\Delta, \emptyset, \perp)$  or  $(\Gamma, \emptyset, \perp)$  or  $(\Delta \cup \Gamma, \emptyset, \perp)$  are C-intelim $_k$  arguments.

(10)

Given the structure of (normal) C-intelim<sub>k</sub> arguments, the proof of (10) follows again from results in [24]. In particular, if one of the two arguments is improper, then it is easy to turn its proof into a normal refutation of the same depth (Remark 50). If neither is improper, then it is easy to combine the two arguments via a single application of the RNC rule to obtain an argument of the same depth concluding to  $\lambda$ . For P4', observe that given satisfaction of P5 and Remark 33, it suffices to show that:

Let  $(\Delta, \Gamma, \alpha)$  be a C-intelim<sub>k</sub> argument such that  $\Delta \cup \Gamma \parallel \alpha$ . Then there exists a C-intelim<sub>k</sub> argument  $(\Delta, \Gamma, \lambda)$ . (11)

This is again straightforward given that such an argument cannot be R-contaminated, hence is E-contaminated, improper, and easily turned into an argument concluding  $\lambda$ .  $\square$

Proposition 51 allows instantiation of a non-redundant pdAF by C-intelim<sub>k</sub> arguments:

**Definition 34** (C-intelim<sub>k</sub> non-redundant pdAF). A C-intelim<sub>k</sub> non-redundant pdAF defined by  $\mathcal{B}$  and  $<$  is a tuple  $(\mathcal{A}, \mathcal{D})$  where  $\mathcal{A}$  is any set of C-intelim<sub>k</sub> arguments satisfying P1, P2, P3, P4' and P5.

The results in Sections 3.4 and 3.7.1 show that the consistency and closure postulates are satisfied by C-intelim<sub>k</sub> non-redundant pdAFs defined by any  $\mathcal{B}$  and  $<$ , and the non-interference and crash resistance postulates are satisfied under the assumption that  $<$  satisfy Pref1 (given that P5 is satisfied, it is not necessary that  $<$  satisfies relevance-coherence).

**Example 52.** We conclude by recalling the example in Section 3.2, on p. 22, in which [43] requires that  $X$ ,  $Y$  and  $Z = (\{p \rightarrow q, p \rightarrow \neg q\}, \emptyset, \neg p)$  can be constructed, to ensure no admissible extension can contain arguments with premises  $p, p \rightarrow q$  and  $p \rightarrow \neg q$ . However, suppose a 0-depth C-intelim pdAF  $(\mathcal{A}, \mathcal{D})$ . Then  $\mathcal{A}$  does not contain the 1-depth argument  $Z$  (the rightmost proof in Fig. 4), but does contain the 0-depth arguments  $(\{p, p \rightarrow \neg q\}, \emptyset, \neg q)$ ,  $(\{p, p \rightarrow q\}, \emptyset, q)$  and  $(\emptyset, \{p, p \rightarrow q, p \rightarrow \neg q\}, \lambda)$ .

#### 4.2. A resource-bounded dialectical characterisation of preferred subtheories

Section 2.1 reviewed [43]'s argumentation-based characterisation of non-monotonic inference relations defined by Preferred Subtheories [13]. Recall that arguments are defined by a totally ordered base, and the Elitist preference relation is used. We now provide a dialectical resource-bounded characterisation of Brewka's non-monotonic inference relations. First, we show that the Elitist preference relation, defined for the more general case of a base equipped with a *partial* preordering  $\leq$  (where as usual  $<$  is defined as  $\alpha < \beta$  iff  $\alpha \leq \beta$  and  $\beta \not\leq \alpha$ ) is a coherent strict partial ordering.

**Definition 35** (Elitist preference ordering). Let  $X, Y$  be dialectical classical logic arguments defined by a base  $\mathcal{B}$ , and  $\leq$  a partial preordering over  $\mathcal{B}$ . Then  $X <_E Y$  iff  $\exists \alpha \in \text{assumptions}(X)$  such that  $\forall \beta \in \text{assumptions}(Y), \alpha < \beta$ .

**Proposition 53** (Properties of elitist preferences).  $<_E$  is a strict partial ordering, invariant modulo logical equivalence (imle), dialectically coherent, and relevance-coherent.

**Proof.** In this proof we abuse notation writing  $\alpha \in A$  instead of  $\alpha \in \text{assumptions}(A)$ .

•  $<_E$  is a strict partial ordering:

*Irreflexivity:* Suppose for contradiction that  $X <_E Y$  and  $Y <_E X$ . Then  $\exists \alpha \in X$  s.t.  $\forall \beta \in Y, \alpha < \beta$ , and  $\exists \beta \in Y$  s.t.  $\forall \alpha \in X, \beta < \alpha$ . But then for some  $\alpha, \beta, \alpha < \beta$  and  $\beta < \alpha$ , contradicting the asymmetry of the strict partial ordering  $<$ .

*Asymmetry:* Suppose for contradiction that  $X <_E X$ . Then  $\exists \alpha \in X$  s.t.  $\alpha < \alpha$ , contradicting the irreflexivity of  $<$ .

*Transitivity:* Suppose  $X <_E Y <_E Z$ . Firstly,  $Y <_E Z$  implies  $\text{assumptions}(Y) \neq \emptyset$  and  $\exists \beta \in Y$  s.t.  $\forall \gamma \in Z, \beta < \gamma$ . By assumption of  $X <_E Y$ ,  $\exists \alpha \in \text{assumptions}(X)$  s.t.  $\alpha < \beta$ . Hence by transitivity of  $<$ ,  $\forall \gamma \in Z, \alpha < \gamma$ . Hence  $X <_E Z$ .

•  $<_E$  is imle. Suppose  $X <_E Y$ , and  $X' \in [X], Y' \in [Y]$ . Then  $\text{assumptions}(X') = \text{assumptions}(X)$  and  $\text{assumptions}(Y') = \text{assumptions}(Y)$ . Hence  $X' <_E Y'$ .

•  $<_E$  is dialectically coherent. For the proof of Pref1, consider any  $X = (\emptyset, \Delta, \lambda)$ . By the properties of classical logic,  $\Delta$  is a non-empty finite (since  $\mathcal{B}$  is finite) set  $\{\alpha_1, \dots, \alpha_n\}$ . Suppose for contradiction that  $\forall i: X <_E (\{\alpha_i\}, \emptyset, \alpha_i)$ . Abusing notation by representing arguments by their assumptions, we have the sequence:

$$S_1 = \{\alpha_1, \dots, \alpha_n\} < \{\alpha_1\} \dots S_n = \{\alpha_1, \dots, \alpha_n\} < \alpha_n$$

1. For  $S_1$ , by irreflexivity of  $<$ ,  $\alpha_1 \not< \alpha_1$ . If  $n = 1$ , then this contradicts  $S_1$ . Else we can assume without loss of generality that  $\alpha_2 < \alpha_1$ .
2. For  $S_2$ , by asymmetry and irreflexivity of  $<$ ,  $\alpha_1 \not< \alpha_2, \alpha_2 \not< \alpha_2$ . If  $n = 2$ , then this contradicts  $S_2$ . Else, we can assume without loss of generality that  $\alpha_3 < \alpha_2$ .

3. For  $S_3$ , by transitivity of  $<$ ,  $\alpha_3 < \alpha_1$ , and so by asymmetry and irreflexivity  $\alpha_1 \not< \alpha_3$ ,  $\alpha_2 \not< \alpha_3$ ,  $\alpha_3 \not< \alpha_3$ . If  $n = 3$ , then this contradicts  $S_3$ . Else, we can assume without loss of generality that  $\alpha_4 < \alpha_3$ .
4. It is easy to see that one can continue reasoning in the same way, until for  $S_n$ ,  $\alpha_i \not< \alpha_n$  for  $i = 1 \dots n$ , contradicting  $S_n$ .

For the proof of *Pref2*, let  $X = (\{\alpha\}, \emptyset, \alpha)$ ,  $Y = (\Delta, \emptyset, \bar{\alpha})$  and  $Y' = (\Delta \cup \{\alpha\}, \emptyset, \phi)$ . Suppose  $Y <_E X$ . Then  $\exists \beta \in \text{assumptions}(Y)$  s.t.  $\beta < \alpha$ . Since  $\text{assumptions}(Y) \subset \text{assumptions}(Y')$ ,  $Y' <_E X$ .

•  $<_E$  is relevance-coherent. Suppose  $Y = (\Delta, \emptyset, \alpha) <_E X$ . Hence  $\exists \beta \in \Delta$ , and so  $\beta \in \Delta \cup \Gamma$  for any  $Y' = (\Delta \cup \Gamma, \emptyset, \alpha)$ , s.t.  $\forall \alpha \in \text{assumptions}(X)$ :  $\beta < \alpha$ .  $\square$

We now adapt [13]'s Preferred Subtheories so as to accommodate agents whose classical reasoning is resource-bounded:

**Definition 36** (*Resource bounded preferred subtheories*). Let  $\vdash_r \subseteq \vdash_c$  be a resource-bounded classical consequence relation, such that: 1) for any  $\Delta$ , if  $\beta \in \Delta$  then  $\Delta \vdash_r \beta$ ; 2) if  $\Delta \vdash_r \alpha$  and  $\Delta \vdash_r \neg \alpha$  then  $\Delta \vdash_r \perp$ . We say that  $\Delta$  is  $r$ -inconsistent iff  $\Delta \vdash_r \perp$ ;  $r$ -consistent otherwise.

Let  $(\mathcal{B}, \leq)$  be a totally ordered set of propositional wff and let  $\mathcal{B}_1, \dots, \mathcal{B}_n$  be a partition of  $\mathcal{B}$  such that for all  $\alpha \in \mathcal{B}_i$ , and all  $\beta \in \mathcal{B}_j$ ,  $i < j$  iff  $\beta < \alpha$ . An  $r$ -preferred subtheory  $\Sigma$  is a set  $\Sigma_1 \cup \dots \cup \Sigma_n$  such that for  $i = 1, \dots, n$ ,  $\Sigma_1 \cup \dots \cup \Sigma_i$  is a  $\subset$ -maximal  $r$ -consistent subset of  $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_i$ .

We now show a correspondence between the  $r$ -preferred subtheories of a base  $\mathcal{B}$  and the stable extensions of a *pdAF* defined by  $\mathcal{B}$ , and using Elitist preferences. Note that in what follows we use the notation  $\text{Args}(\Sigma) = \{X \mid \text{prem}(X) \subseteq \Sigma\}$ .

**Theorem 54.** Let  $(\mathcal{A}, \mathcal{D})$  be defined by  $\mathcal{B}$  and  $<_E$  be defined on the basis of a total ordering  $\leq$  over  $\mathcal{B}$ , and such that  $(\Delta, \Gamma, \alpha) \in \mathcal{A}$  iff  $\Delta \cup \Gamma \vdash_r \alpha$ . Then:

- 1) If  $\Sigma$  is an  $r$ -preferred subtheory of  $\mathcal{B}$ , then  $E = \text{Args}(\Sigma)$  is a stable extension of  $(\mathcal{A}, \mathcal{D})$ .
- 2) If  $E$  is a stable extension of  $(\mathcal{A}, \mathcal{D})$ , then  $\Sigma = \bigcup_{X \in E} \text{Prem}(X)$  is an  $r$ -preferred subtheory of  $\mathcal{B}$ .

**Proof.**

*Proof of 1):* We show  $E$  is conflict free. Since  $\Sigma$  is  $r$ -consistent,  $\Sigma \not\vdash_r \perp$ . Suppose for contradiction that  $E$  is not conflict free, in which case  $\exists X, Y \in E$  s.t.  $X = (\Delta, \Gamma, \phi)$  ( $\phi = \perp$  or  $\beta$ ) and  $X \Rightarrow_E Y$  on  $\bar{\beta} \in \text{prem}(Y)$ . Since  $\Delta \subseteq \text{prem}(E)$  and  $\Gamma \subseteq \text{prem}(E)$  (given  $Y \in E$ ), then  $\phi = \perp$  implies  $\Sigma \vdash_r \perp$ .  $\phi = \beta$  implies  $\Sigma \vdash_r \beta$  and since  $\bar{\beta} \in \text{prem}(Y)$ ,  $\Sigma \vdash_r \bar{\beta}$ . Hence  $\Sigma \vdash_r \perp$ . Contradiction. We show  $\forall Y \in \mathcal{A} \setminus E$ ,  $\exists X \in E$  s.t.  $X \Rightarrow_{\{Y\}} Y$ . Consider any such  $Y$  ( $Y$  cannot be an argument of the form  $(\emptyset, \Delta, \phi)$  since any such  $Y$  is in  $E = \text{Args}(\Sigma)$ ). Then  $\exists \gamma \in \text{prem}(Y)$ ,  $\gamma \notin \Sigma$ . By construction,  $\Sigma = \Sigma_1 \cup \dots \cup \Sigma_n$  such that for  $i = 1 \dots n$ ,  $\Sigma_1 \cup \dots \cup \Sigma_i$  is a maximal  $r$ -consistent subset of  $\mathcal{B}_1, \dots, \mathcal{B}_i$ . Hence, suppose  $\gamma \in \mathcal{B}_j$  for some  $j = 1 \dots n$ . Then  $\Sigma_1 \cup \dots \cup \Sigma_j \cup \{\gamma\} \vdash_r \perp$ . Hence  $\exists X = (\Delta, \{\gamma\}, \perp) \in \text{Args}(\Sigma_1 \cup \dots \cup \Sigma_j)$ , and so  $X \in E$ . Since  $\gamma \in \mathcal{B}_j$ , and  $\Delta \subseteq \bigcup_{k=1}^j \mathcal{B}_k$ , then  $X \not\Rightarrow_E (\{\gamma\}, \emptyset, \gamma)$ . Hence  $X \Rightarrow_{\{Y\}} Y$  on  $\gamma$ .

*Proof of 2):* We show that  $\Sigma = \bigcup_{X \in E} \text{Prem}(X)$  is  $r$ -consistent. Suppose for contradiction that  $\Sigma \vdash_r \perp$ . Then  $Z = (\emptyset, \Delta, \perp) \in \mathcal{A}$  for some  $\Delta \subseteq \Sigma$ . Hence  $\forall \alpha \in \Delta$ ,  $\exists B \in E$  s.t.  $Z \Rightarrow_E B$  on  $\alpha$ ,<sup>28</sup> contradicting  $E$  is stable.

Now let  $\Sigma_1, \dots, \Sigma_n$  be the partition of the  $r$ -consistent  $\Sigma$  s.t. for  $i = 1 \dots n$ ,  $\Sigma_i$  is a (possibly empty) subset of  $\mathcal{B}_i$  in the stratification  $\mathcal{B}_1, \dots, \mathcal{B}_n$  of  $\mathcal{B}$ . Suppose for contradiction that for some  $i$ ,  $\Sigma_1 \cup \dots \cup \Sigma_i$  is not a  $\subset$ -maximal  $r$ -consistent subset of  $\mathcal{B}_1 \cup \dots \cup \mathcal{B}_i$ . Without loss of generality we can assume that for  $k = 1 \dots i - 1$ ,  $\Sigma_1, \dots, \Sigma_k$  is a  $\subset$ -maximal  $r$ -consistent subset of  $\mathcal{B}_1, \dots, \mathcal{B}_{i-1}$ . Then  $\exists \alpha \in \mathcal{B}_i$  s.t.

- i)  $\alpha \notin \Sigma_i$       ii)  $\Sigma_1 \cup \dots \cup \Sigma_{i-1} \cup \Sigma_i \cup \{\alpha\} \not\vdash_r \perp$ .

Given i),  $Y = (\{\alpha\}, \emptyset, \alpha) \notin E$ . Since  $E$  is stable,  $\exists X \in E$ ,  $X \Rightarrow_{\{Y\}} Y$ , where:

- a)  $X = (\Delta, \{\alpha\}, \perp)$  or b)  $X = (\Delta, \{\alpha\}, \bar{\alpha})$  or c)  $X = (\Delta, \emptyset, \bar{\alpha})$ .

Note that since  $E$  is stable then  $E$  is complete (see immediately following Definition 15 and Footnote 15). Suppose either  $Z = (\Delta, \emptyset, \perp) \in \mathcal{A}$  or  $(\{\alpha\}, \emptyset, \perp) \in \mathcal{A}$ . The former contradicts  $X \in E$  is acceptable w.r.t.  $E$  (given that  $Z' = (\emptyset, \Delta, \perp) \in \mathcal{A}$  and  $Z' \Rightarrow_E X$ ), and the latter implies  $\{\alpha\} \vdash_r \perp$ , contradicting ii).

Let us now rename  $X$  in c) to  $X'$ , and in b) we have (by P2)  $X' = (\Delta \cup \{\alpha\}, \emptyset, \bar{\alpha}) \in \mathcal{A}$ .

For both instances of  $X'$ , and having shown  $(\Delta, \emptyset, \perp) \notin \mathcal{A}$  and  $(\{\alpha\}, \emptyset, \perp) \notin \mathcal{A}$ , we have by P3 that  $(\Delta \cup \{\alpha\}, \emptyset, \perp) \in \mathcal{A}$ . These cases and case a) imply  $\Delta \cup \{\alpha\} \vdash_r \perp$ . But then given ii) and  $\Delta \subseteq \Sigma_1, \dots, \Sigma_n$ , it must be that  $\exists \beta \in \Delta$ , s.t.  $\beta \in \Sigma_j$ ,  $j > i$ . But then in cases a), b) and c),  $X <_E Y$ , contradicting  $X \Rightarrow_{\{Y\}} Y$ .  $\square$

The above correspondence yields a resource bounded dialectical characterisation of non-monotonic inference in Preferred Subtheories (both sceptical and credulous inference relations, as described in Section 2.1).

<sup>28</sup> Note also that by virtue of  $<_E$  satisfying *Pref1* it must be that  $\exists \alpha \in \Delta$  s.t.  $Z \not\Rightarrow_E (\{\alpha\}, \emptyset, \alpha)$  and so  $\exists B \in E$  s.t.  $Z \Rightarrow_E B$  on  $\alpha$ .



## 5. Related work

Design guidelines for theoretical models of argumentation suitable for real world applications<sup>29</sup> have been proposed by Dung et al. in [27]. Firstly, submitted arguments and attacks should be *transparent* (Postulate 2.1) in that the computational cost of verifying: 1) the legitimacy of submitted arguments should be at most polynomial in the size of the (constructed) arguments; 2) that an argument attacks another should be at most linear (in the size of the argument's conclusion). Our approach satisfies the first desideratum,<sup>30</sup> since we drop the consistency and subset minimality checks on arguments' assumptions, which is what makes legitimacy checks intractable under a widely accepted conjecture in computational complexity. By contrast, the time required to check the correctness of a constructed proof is bounded above by a polynomial in the size of the proof.<sup>31</sup> As for the second desideratum, we note that it is satisfied by the purely “declarative” notion of attack in Definition 11. However, we also make essential use of the more complex notion of *dialectical defeat* in Definition 15, that allows the defeating argument  $X$  to include, as suppositions, any premises of arguments in the extension  $S$  with respect to which the acceptability of the defeated argument  $Y$  is challenged. This notion of defeat can be checked in polynomial time with respect to  $\text{suppositions}(X) \cup \text{prem}(S \cup \{Y\}) \cup \{\text{conclusion}(X)\}$ .<sup>32</sup> Transparency guarantees that *constructed* arguments and defeats can be understood by any parties, independently of their level of sophistication [27], with a use of resources which is polynomial with respect to the complexity of the input.<sup>33</sup> We go one step further in this paper, arguing against the assumption that an agent constructs all arguments defined by a base; in particular because they may be resource bounded. Hence we drop the consistency and subset minimality checks on arguments' assumptions, and formally study partially instantiated frameworks. Moreover, for depth-bounded instantiations (see Section 4.1.2), it is worth noting that even *finding* a C-intelim<sub>k</sub> proof takes time at most polynomial in the size of the input assumptions and conclusion (see Proposition 47).

Secondly, an argument's assumptions should be *relevant* to its conclusion (Postulate 2.2), at least in the “weaker sense [...] that the argument is a defeasible proof of its claim from its support, without any *obvious* redundancy of any parts of the support” [27, p. 187, our emphasis]. In this paper we propose an analogous notion of “obvious” redundancy in terms of the syntactic unrelatedness of some part of the assumptions to the other assumptions and to the conclusion, which unlike that imposed by subset minimality, can be enforced proof theoretically (as shown in Section 4).

Thirdly, the *no dismissal postulate* (Postulate 2.3) proposes that legitimate arguments should not be dismissed without reason, and if dismissed (i.e., not submitted), their dismissal should not change the semantics of the given argumentation framework. We concur in the sense that all arguments constructed must be included in a *pdAF*, unless their exclusion does not change the semantics. Related to the no dismissal postulate, ‘redundant’ arguments are distinguished from ‘non-redundant’ arguments in [27], where some  $Y$  is redundant if there is an  $X$  that attacks every  $Z$  that is attacked by  $Y$ , and any  $Z$  attacking  $X$  also attacks  $Y$ . One can then, having partitioned all the arguments defined by a base, exclude those that are redundant, while guaranteeing that the status of non-redundant arguments will remain unchanged.

A key feature of our approach is the distinction between premises accepted as true, and suppositions assumed true for the sake of argument. One can thus model the dialectical move of showing that an interlocutor's supposed premises, together with premises accepted as true, imply a contradiction. Caminada [14] is similarly motivated to develop accounts of argumentation for application in dialogues involving human and computational agents. He formalises what he terms ‘hang yourself’ (*HY*) arguments built from a language consisting of premises that are literals – atomic propositions and their negations – as well as defeasible rules  $L_0 \wedge \dots \wedge L_{n-1} \Rightarrow L_n$  and ‘foreign commitments’  $\rightsquigarrow L$ . The latter refer to premises and conclusions of rules in attacked arguments, by way of supposing these to be true for the sake of argument. *HY* arguments can only reference premises or conclusions in an attacked argument  $Y$  (rather than in additional arguments used to defend  $Y$ ). Clearly, Caminada is concerned with formalising intuitions similar to ours, albeit in a more restricted logical setting, only under the grounded semantics, and without accommodating the use of preferences. Satisfaction of rationality postulates is not studied in [14].

It is worth noting at this point that the distinction between premises and suppositions in  $(\Delta, \{\phi_1, \dots, \phi_n\}, \psi)$  cannot equivalently be made using the standard ontology, by an argument  $(\Delta, \phi_1, \dots, \phi_n \rightarrow \psi)$ . This does not faithfully represent the intuitive meaning of the conditional claim because of the well-known qualms concerning the use of material implication to express a kind of entailment [5]. For example, most reasoners would not be prepared to accept that the falsity of such a conditional claim implies that all the  $\phi_i$  are factually true, nor that the factual falsity of one of the premises in  $\phi_i$  implies that the claim is true. Moreover, one would then either only be able to use such arguments as attacks on sets of premises rather than individual premises, or require that the premises of the attacked argument contain  $\neg(\phi_1, \dots, \phi_n \rightarrow \psi)$ .

<sup>29</sup> The guidelines are stated in the context of applications involving computational and/or human agents.

<sup>30</sup> This requirement was first put forward by Cook and Rechow, in their seminal [18].

<sup>31</sup> This depends on the proof system adopted, but is certainly true for natural deduction systems like the one presented in Section 4. Indeed, we are unaware of any conventional proof systems for classical logic whose proofs cannot be verified in polynomial time.

<sup>32</sup> Note also, that in practice, the check can be restricted to the subset  $S'$  of  $S$  consisting of the arguments of  $S$  whose premises are actually referenced in  $\text{suppositions}(X)$  (recall Footnote 14 on page 18).

<sup>33</sup> For dialectical defeats, the input includes, besides the conclusion, the suppositions of the defeating argument, the defeated argument  $Y$ , and the set  $S$  with respect to which it is intended as a defeat.

Recently, pragmatic considerations also motivate dropping consistency and subset minimality checks on propositional arguments in [6,7]. Arguments are Gentzen style sequents and arguments with inconsistent premises are attacked by sequents with empty antecedents. In this work, the distinction between premises and suppositions, and the use of preferences are not considered. The postulates in [15] are not studied and neither is there consideration of partially instantiated frameworks. Types of redundancy in arguments are also identified in the context of *assumption based argumentation* (ABA), and are excluded by switching from a conception of arguments as trees to graphs [19]. While ABA does not capture classical logic argumentation, insights from [19] may potentially usefully be studied in the setting of *Cl-Arg*.

A key study of propositional classical logic instantiations of Dung *AFs* [33] considers various definitions of attacks amongst a (in general infinite) set of arguments defined by a base, and study (without use of preferences) consistency postulates closely related to those in [15] and the premise consistency postulate described here in Section 3.4. The two attack relations shown to be well behaved are those in which the attacking argument's conclusion is classically equivalent to, respectively classically entails, the negation of a premise in the attacked argument.<sup>34</sup> Note that neither of these two attack definitions satisfy condition 2) of the transparency postulate described above.

[43]'s study of classical logic argumentation with preferences, as an instance of the *ASPIC*<sup>+</sup> framework, has been reviewed (in Section 2) and compared with our approach at various points in this paper. In summary, the *ASPIC*<sup>+</sup> instantiation requires checks on the consistency of arguments' premises and does not model the dialectical demonstration that a set of arguments' premises are inconsistent. While the foreign commitment problem has been addressed in [17], by allowing attacks on arguments' conclusions in a version of *ASPIC*<sup>+</sup> that accommodates classical logic instantiations, the consistency and closure rationality postulates of [15] only then hold for the grounded semantics, and then only if argument preferences are defined by a totally ordered base *B* under the Elitist principle [17]. In this paper we have argued that subset minimality is a blunt and prohibitively expensive<sup>35</sup> instrument for ensuring non-redundancy of an argument's premises. As already discussed, *ASPIC*<sup>+</sup> also eschews subset minimality checks on arguments' premises, and show that the argumentation defined inferences remain unchanged under the assumption that preference relations do not strengthen non-subset minimal arguments. We have advocated a notion of non-redundancy defined in terms of the syntactic relatedness of formulae; one that can be enforced through appropriate restrictions on applications of proof rules in which case no such assumption on preference relations is required.

Satisfaction of the non-interference and crash resistance postulates [16] is not studied for the *ASPIC*<sup>+</sup> formalisation of *Cl-Arg*. Consistency and closure are satisfied by complete extensions, for a restricted class of (reasonable) preference relations, and assuming all arguments defined by a base are included in an *AF*. However we show direct consistency for admissible extensions, and (modified) closure postulates for complete extensions, assuming arbitrary preference relations and for partially instantiated frameworks. Note that recent work [28] studies propositional instantiations of *AFs* that combine strict and defeasible rules, and without use of preferences. Dung and Thang [28] show that to guarantee closure and consistency, it is not necessary that the strict inference rules satisfy the property of *contraposition* (see [15,43]). Their aim is different to ours. We consider first order classical logic instantiations with preferences in which *all* arguments are definable by a base, while we drop the assumption that all such arguments are constructed and included in a framework in order to ensure satisfaction of closure, consistency and the non-contamination postulates.

A number of works attempt to show satisfaction of the non-interference and crash resistance postulates for *propositional* argumentation formalisms that integrate deductive and defeasible reasoning. In [16], propositional logic programming and Default Logic [54] instantiations of Dung frameworks are shown to satisfy these postulates under the semi-stable semantics. In [61], arguments are built from a set of *classically consistent* propositional formulae *P*, and defeasible (*D*) and strict inference rules. The latter encode only those classical inferences  $\Delta \vdash_c \alpha$  such that the atoms in  $\Delta \cup \{\alpha\}$  are a subset of the atoms appearing in *P* and *D*. Wu and Podlaszewski [61] relate the constructed arguments by attacks and so do not consider the use of preferences. Inconsistent arguments are identified as those whose contained premises together with the conclusions of defeasible rules are classically inconsistent. They then show satisfaction of the consistency, closure, non-interference and crash resistance postulates for 'inconsistency cleansed' Dung frameworks in which inconsistent arguments are excluded. Finally, Grooters and Prakken [34] define a version of the *ASPIC*<sup>+</sup> framework in which the strict inference rules encode inference in [55]'s paraconsistent logic. Essentially, arguments are built from premises, defeasible inference rules and strict inference rules, where the latter encode classical inferences from only consistent sets of formulae. Grooters and Prakken [34] focus on showing that the adapted *ASPIC*<sup>+</sup> framework satisfies closure and consistency postulates. Satisfaction of non-interference and crash resistance is not shown (although the authors state that satisfaction of these postulates can be taken for granted given the absence of the Ex Falso principle).

Finally, note that this paper revises and substantially extends [25]. We have revised the definition of attacks and defeats originating from arguments concluding  $\perp$ , modified the conditions on arguments in *pdAFs*, and specified a new class of non-redundant frameworks that exclude R-contaminated arguments. The closure and consistency postulates are satisfied

<sup>34</sup> Both these attacks are shown to be equivalent to the undermine attacks used in this paper, in the sense that the complete extensions are the same (even when assuming use of preferences) [43].

<sup>35</sup> Note that there are attempts to ameliorate the computational expense of the consistency and subset minimality checks; most notably [29], in which algorithms are implemented for generating arguments whose claims are clauses, from propositional knowledge bases in clausal form, where connection graphs are used to reduce the computational expense. Tests on these implementations yield encouraging results, and the authors indicate that a comprehensive empirical evaluation will be the subject of future work.

for a restricted class of preference relations in [25], whereas this paper shows satisfaction of these postulates for arbitrary preference relations. This paper also additionally: 1) formally proves the Fundamental Lemma and existence of a unique least fixed point of a *pdAF*'s characteristic function; 2) provides a comprehensive study of two types of contaminated arguments; 3) formally proves the non-interference and crash resistance postulates, generalised here to first order argumentation; 4) studies instantiation of *pdAFs* by natural deduction proof theories that exclude R-contaminated arguments and accommodate agents that have bounds on their inferential capabilities.

## 6. Conclusions

Dung's theory of argumentation has been proposed as a basis for applications, in large part because of its provision of intuitive dialectical characterisations of non-monotonic inference for individual and distributed (dialogical) reasoning accommodating computational and human agents. We have argued that such applications require that the theory account for resource bounded agents and features of real-world dialectical reasoning. Our focus has been on classical logic instantiations of Dung argumentation frameworks (*CI-Arg*) that yield non-monotonic inference relations over first order knowledge bases. We have shown that features of current formulations of *CI-Arg* that suffice to ensure satisfaction of rationality postulates, preclude satisfaction of these practical desiderata. We have thus been motivated to formalise an account of *CI-Arg* that is both practical and rational.

Our solution is to essentially account for the dialectical use of argument by refining the ontology of arguments so as to make explicit the epistemic distinction between an argument's premises assumed true, and those supposed true for the sake of argument. Our approach can thus accommodate a ubiquitous feature of argumentative practice, whereby the inconsistency of arguments' premises is demonstrated dialectically. We eschew the computationally unfeasible checks on the consistency and subset minimality of arguments' premises that are used to verify the legitimacy of arguments prior to inclusion in an *AF*, and enumerate minimal and intuitive assumptions as to the arguments constructed from a base for inclusion in an *AF*, such that they suffice to ensure rational outcomes. In particular: 1) arguments with conflicting conclusions are obviously indicative of the inconsistency of their supporting assumptions, as recognised by inclusion of an argument that combines the assumptions to conclude  $\perp$ ; 2) construction of an argument whose support makes use of redundant premises that contaminate an argument, implies construction of the non-contaminated argument excluding the redundant premises. We have shown that assuming only partially instantiated frameworks (*pdAFs*), key properties of Dung frameworks are satisfied. Also, the consistency and closure postulates are satisfied assuming arbitrary preference relations, with consistency being shown for admissible (and not just complete) extensions; a result that we argue is important for practical applications in which it often suffices to defend an argument by showing membership in an admissible extension. We have also identified a notion of redundancy of an argument's assumptions that does not appeal to subset minimality, but rather to the syntactic relatedness of the assumptions and conclusion, such that redundant arguments that make use of syntactically disjoint assumptions can be excluded proof theoretically. Thus non-redundant *pdAFs* instantiated by proof theories that exclude redundant arguments, trivially satisfy condition 2) above. Moreover, we have shown that despite dropping the legitimacy checks on arguments, *pdAFs* satisfy the non-interference and crash resistance postulates (which in this paper are generalised to argument frameworks instantiated by first order theories) under the assumption that preference relations are dialectically and redundancy coherent. For non-redundant *pdAFs* it suffices that preference relations are dialectically coherent.

Existing accounts of *CI-Arg* typically neglect the study of the proof theoretic means for constructing arguments. We have argued that such accounts provide 'schemata' for classical logic argumentation, as an argument properly constituted should include the reasoning steps employed in entailing the conclusion from the premises (in keeping with the *transparency* postulate advocated in [27]). We have therefore proposed a natural deduction proof theory for propositional *CI-Arg* that in keeping with practical desiderata for applications: i) provides a more perspicuous account of classical reasoning (in comparison with Gentzen style natural deduction); ii) allows definition of a hierarchy of tractable depth-bounded deducibility relations equating with stepwise increments in agents' assumed inferential resources, and such that *pdAFs* instantiated up to any given depth meet conditions for satisfaction of the rationality postulates; iii) can be adapted so that only non-redundant arguments are constructed. We also provided an argumentative characterisation of the non-monotonic Preferred Subtheories [13], under the assumption that agents may have limited inferential capabilities.

This paper suggests an ambitious agenda for future work. We are currently generalising our approach to the *ASPIC*<sup>+</sup> framework [43] in which arguments are built from strict inference rules that encode the inference relations of deductive logics, as well as defeasible inference rules and premises. We believe this will pose few serious challenges, and will involve:

1. Differentially labelling the constituents of *ASPIC*<sup>+</sup> tree structured arguments so as to distinguish rules and premises that are assumed true and supposed true, and adapting this paper's notions of attack defeat and acceptability. Hence, for example, an argument *X* concluding  $\perp$  would target those defeasible rules and premises in the attacked argument that are labelled as suppositions in *X*.
2. Identifying syntactically disjoint subsets of premises and rules in an *ASPIC*<sup>+</sup> argument so as to identify explosively and redundantly contaminated arguments, and adapting *P1*, ..., *P5* and *P4'* so as to define (non-redundant) *pdAFs*. Note that

this may require stipulating assumptions as to the strict rules used in constructing arguments, such that the property satisfied by classical logic (in Proposition 30) applies to *ASPIC*<sup>+</sup> arguments.<sup>36</sup>

With appropriate adaptations of dialectical and redundancy coherent preference relations, we then intend showing that partially instantiated frameworks satisfy the consistency and closure postulates (without assumptions on preference relations), and the non-contamination postulates. Notice that violation of the non-contamination postulates arises because of the encoding of classical logic inference in the strict rules, and as discussed in Section 5 when reviewing [61], avoiding this problem involves significantly compromising the generality of *ASPIC*<sup>+</sup>, and 'cleansing' the framework of inconsistent arguments. We therefore aim for satisfaction of these postulates for the fully general version of *ASPIC*<sup>+</sup> and without any need for cleansing.

Secondly, we intend adapting for *pdAFs*, argument game proof theories developed for Dung *AFs* [42]. We will then generalise these argument games to persuasion dialogues in which agents submit arguments constructed from their own bases and the premises of arguments that are incrementally added to a public commitment store. We believe these to be important steps in achieving our long term aim of formalising accounts of monological reasoning and distributed non-monotonic reasoning through dialogue, suitable for real-world rather than idealised agents. Moreover, this paper has investigated instantiation of *pdAFs* by *propositional* resource bounded proof theories that exclude redundantly contaminated arguments. An important future research focus will be the development of *first order* resource bounded proof theories that exclude redundantly contaminated arguments.

## Acknowledgements

We would like to thank the anonymous reviewers whose detailed and constructive comments have helped to improve this paper.

## References

- [1] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, *Ann. Math. Artif. Intell.* 34 (1–3) (2002) 197–215.
- [2] L. Amgoud, N. Maudet, S. Parsons, Modelling dialogues using argumentation, in: *Proc. 4th Int. Conference on MultiAgent Systems*, 2000, pp. 31–38.
- [3] L. Amgoud, H. Prade, Using arguments for making and explaining decisions, *Artif. Intell.* 173 (34) (2009) 413–436.
- [4] L. Amgoud, S. Vesic, Handling inconsistency with preference-based argumentation, in: *4th International Conference Scalable Uncertainty Management: 4th International Conference, SUM 2010, 2010*, pp. 56–69.
- [5] A.R. Anderson, N.D. Belnap Jr., *Entailment: The Logic of Relevance and Necessity*, vol. 1, Princeton University Press, Princeton, 1975.
- [6] O. Arieli, C. Straßer, Sequent-based logical argumentation, *Argum. Comput.* 6 (1) (2015) 73–99.
- [7] O. Arieli, C. Straßer, Deductive argumentation by enhanced sequent calculi and dynamic derivations, *Electron. Notes Theor. Comput. Sci.* 323 (2016) 21–37.
- [8] K. Atkinson, T. Bench-Capon, P. McBurney, Computational representation of practical argument, *Synthese* 152 (2005) 2006.
- [9] T.J.M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, *J. Log. Comput.* 13 (3) (2003) 429–448.
- [10] T.J.M. Bench-Capon, S. Doutre, P.E. Dunne, Audiences in argumentation frameworks, *Artif. Intell.* 171 (1) (2007) 42–71.
- [11] T.J.M. Bench-Capon, P.E. Dunne, Argumentation in artificial intelligence, *Artif. Intell.* 171 (2007) 10–15.
- [12] P. Besnard, A. Hunter, *Elements of Argumentation*, MIT Press, 2008.
- [13] G. Brewka, Preferred subtheories: an extended logical framework for default reasoning, in: *Proc. 11th International Joint Conference on Artificial Intelligence*, 1989, pp. 1043–1048.
- [14] M. Caminada, Dialogues and HY-arguments, in: *Non-Monotonic Reasoning*, 2004, pp. 94–99.
- [15] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, *Artif. Intell.* 171 (5–6) (2007) 286–310.
- [16] M. Caminada, W. Carnielli, P. Dunne, Semi-stable semantics, *J. Log. Comput.* 22 (5) (2012) 1207–1254.
- [17] M. Caminada, S. Modgil, N. Oren, Preferences and unrestricted rebut, in: *Proc. Computational Models of Argument, COMMA'14*, 2014, pp. 209–220.
- [18] S.A. Cook, R. Rehow, On the length of proofs in the propositional calculus, in: *Proc. 6th Annual Symposium on the Theory of Computing*, 1974, pp. 135–148.
- [19] R. Craven, F. Toni, Argument graphs and assumption-based argumentation, *Artif. Intell.* 233 (C) (2016) 1–59.
- [20] M. D'Agostino, Classical natural deduction, in: S.N. Artëmov, et al. (Eds.), *We Will Show Them!*, vol. 1, College Publications, 2005, pp. 429–468.
- [21] M. D'Agostino, Analytic inference and the informational meaning of the logical operators, *Log. Anal.* 57 (227) (2014) 407–437.
- [22] M. D'Agostino, An informational view of classical logic, *Theor. Comput. Sci.* 606 (2015) 79–97.
- [23] M. D'Agostino, M. Finger, D.M. Gabbay, Semantics and proof-theory of depth-bounded Boolean logics, *Theor. Comput. Sci.* 480 (2013) 43–68.
- [24] M. D'Agostino, D.M. Gabbay, S. Modgil, Normality, Non-Contamination and Logical Depth in Classical Natural Deduction, Technical Report, Parts I and II, 2017, [http://www.filosofia.unimi.it/dagostino/wp-content/uploads/2018/05/NNC\\_TR1.pdf](http://www.filosofia.unimi.it/dagostino/wp-content/uploads/2018/05/NNC_TR1.pdf), [http://www.filosofia.unimi.it/dagostino/wp-content/uploads/2018/05/NNC\\_TR2.pdf](http://www.filosofia.unimi.it/dagostino/wp-content/uploads/2018/05/NNC_TR2.pdf).
- [25] M. D'Agostino, S. Modgil, A rational account of classical logic argumentation for real-world agents, in: *European Conference on Artificial Intelligence, ECAI 2016*, 2016, pp. 141–149.
- [26] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games, *Artif. Intell.* 77 (2) (1995) 321–358.
- [27] P.M. Dung, F. Toni, P. Mancarella, Some design guidelines for practical argumentation systems, in: *Proc. Computational Models of Argument, COMMA'10*, 2010, pp. 183–194.
- [28] P.M. Dung, P.M. Thang, Closure and consistency in logic-associated argumentation, *J. Artif. Intell. Res.* 49 (1) (2014) 79–109.
- [29] V. Efstathiou, A. Hunter, Algorithms for generating arguments and counterarguments in propositional logic, *Int. J. Approx. Reason.* 52 (6) (2011) 672–704.

<sup>36</sup> By substituting  $\vdash_{A^+}$  for  $\vdash_c$  in the statement of this property, where the former is a relation between sets of rules and premises, and the conclusions of arguments defined by these sets.

- [30] T. Eiter, G. Gottlob, The complexity of logic-based abduction, *J. ACM* 42 (1) (1995) 3–42.
- [31] X. Fan, F. Toni, A general framework for sound assumption-based argumentation dialogues, *Artif. Intell.* 216 (2014) 20–54.
- [32] G. Galilei, *Dialogues Concerning Two New Sciences* [1638], Dover, NY, 1954.
- [33] N. Gorogiannis, A. Hunter, Instantiating abstract argumentation with classical logic arguments: postulates and properties, *Artif. Intell.* 175 (2011) 1479–1497.
- [34] D. Grooters, H. Prakken, Combining paraconsistent logic with argumentation, in: *Proc. Computational Models of Argument, COMMA'14*, 2014, pp. 301–312.
- [35] A. Hunter, Making argumentation more believable, in: *Proceedings of AAAI 2004*, 2004, pp. 269–274.
- [36] A. Kakas, P. Mancarella, F. Toni, On argumentation logic and propositional logic, *Stud. Log.* 106 (2) (2018) 237–279.
- [37] J. Lawrence, F. Bex, C. Reed, Dialogues on the argument web: mixed initiative argumentation with arvina, in: *Proc. Computational Models of Argument, COMMA'12*, 2012, pp. 513–514.
- [38] P. McBurney, S. Parsons, Chapter 13: dialogue games for agent argumentation, in: I. Rahwan, G. Simari (Eds.), *Argumentation in AI*, Springer, 2009, pp. 261–280.
- [39] S. Modgil, An argumentation based semantics for agent reasoning, in: *Proc. Workshop on Languages, Methodologies and Development Tools for Multi-Agent Systems, LADS 07*, 2007, pp. 37–53.
- [40] S. Modgil, Revisiting abstract argumentation frameworks, in: *Theory and Applications of Formal Argumentation, TFAFA'14*, 2014, pp. 1–15.
- [41] S. Modgil, T.J.M. Bench-Capon, Integrating dialectical and accrual modes of argumentation, in: *Proc. Computational Models of Argument, COMMA'10*, 2010, pp. 335–346.
- [42] S. Modgil, M. Caminada, Chapter 6: proof theories and algorithms for abstract argumentation frameworks, in: I. Rahwan, G. Simari (Eds.), *Argumentation in AI*, Springer, 2009, pp. 105–129.
- [43] S. Modgil, H. Prakken, A general account of argumentation and preferences, *Artif. Intell.* 195 (2013) 361–397.
- [44] S. Modgil, F. Toni, F. Bex, I. Bratko, C.I. Chesnevar, X. Fan, S. Gaggl, A.J. Garcia, M.P. Gonzalez, T. Gordon, J. Leite, M. Mozina, C. Reed, G.R. Simari, S. Szeider, P. Torroni, S. Woltran, The added value of argumentation, in: *Handbook of Agreement Technologies*, Springer-Verlag, 2013, Chapter 21.
- [45] S. Parsons, M. Wooldridge, L. Amgoud, On the outcomes of formal inter-agent dialogues, in: *Proc. 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2003, pp. 683–688.
- [46] C. Perelman, L. Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation*, University of Notre Dame Press, 1969.
- [47] J.L. Pollock, Defeasible reasoning, *Cogn. Sci.* 11 (1987) 481–518.
- [48] H. Prakken, Coherence and flexibility in dialogue games for argumentation, *J. Log. Comput.* 15 (2005) 1009–1040.
- [49] H. Prakken, A study of accrual of arguments, with applications to evidential reasoning, in: *Proc. International Conference on AI and Law*, 2005, pp. 85–94.
- [50] H. Prakken, G. Sartor, Argument-based extended logic programming with defeasible priorities, *J. Appl. Non-Class. Log.* 7 (1997) 25–75.
- [51] I. Rahwan, L. Amgoud, An argumentation-based approach for practical reasoning, in: N. Maudet, S. Parsons, I. Rahwan (Eds.), *Argumentation in Multi-Agent Systems*, in: *Lecture Notes in Computer Science*, vol. 4766, Springer, Berlin Heidelberg, 2007, pp. 74–90.
- [52] I. Rahwan, M. Madakkatel, J. Bonnefon, R.N. Awan, S. Abdallah, Behavioral experiments for assessing the abstract argumentation semantics of reinstatement, *Cogn. Sci.* 34 (8) (2010) 1483–1502.
- [53] I. Rahwan, G. Simari (Eds.), *Argumentation in AI*, Springer-Verlag, 2009.
- [54] R. Reiter, A logic for default reasoning, *Artif. Intell.* 13 (1980) 81–132.
- [55] N. Rescher, R. Manor, On inference from inconsistent premises, *Theory Decis.* 1 (1970) 179–219.
- [56] R. Stalnaker, The problem of logical omniscience, *Synthese* 89 (3) (1991) 425–440.
- [57] P.M. Thang, H.T. Luong, Translating preferred subtheories into structured argumentation, *J. Log. Comput.* 24 (4) (2014) 831–849.
- [58] P. Tolchinsky, S. Modgil, K. Atkinson, P. McBurney, U. Cortes, Deliberation dialogues for reasoning about safety critical actions, *J. Auton. Agents Multi-Agent Syst.* 25 (2012) 209–259.
- [59] G. Vlastos, The Socratic Elenchus, *J. Philos.* 79 (11) (1982) 711–714.
- [60] D.N. Walton, E.C.W. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, State University of New York Press, 1995.
- [61] Y. Wu, M. Podlaszewski, Implementing crash-resistance and non-interference in logic-based argumentation, *J. Log. Comput.* 25 (2015) 303–333.