

# A Study of Argumentative Characterisations of Preferred Subtheories

Marcello D’Agostino<sup>1</sup> and Sanjay Modgil<sup>2</sup>

<sup>1</sup> Department of Philosophy, University of Milan

<sup>2</sup> Department of Informatics, King’s College London  
marcello.dagostino@unimi.it, sanjay.modgil@kcl.ac.uk

## Abstract

Classical logic argumentation (*Cl-Arg*) under the stable semantics yields argumentative characterisations of non-monotonic inference in Preferred Subtheories. This paper studies these characterisations under both the standard approach to *Cl-Arg*, and a recent dialectical approach that is provably rational under resource bounds. Two key contributions are made. *Firstly*, the preferred extensions are shown to coincide with the stable extensions. This means that algorithms and proof theories for the admissible semantics can now be used to decide credulous inference in Preferred Subtheories. *Secondly*, we show that as compared with the standard approach, the grounded semantics applied to the dialectical approach more closely approximates sceptical inference in Preferred Subtheories.

## 1 Introduction

The impact of Dung’s seminal theory of argumentation [Dung, 1995] is largely due to its dialectical characterisation of non-monotonic inference. Given a base  $\mathcal{B}$  of logical formulae, one defines the arguments related by binary attacks. Preferences over arguments can then be used to determine which attacks succeed as defeats. Given the resulting argument framework (*AF*) consisting of arguments related by attacks or defeats, the winning sets (extensions) of arguments under various semantics are evaluated, and the conclusions of arguments in these extensions identify the inferences defined by the ‘instantiating’  $\mathcal{B}$ . These are then shown to correspond to the inferences defined directly over  $\mathcal{B}$  by various non-monotonic logics.

A number of works [Amgoud and Vesic, 2010; Modgil and Prakken, 2013; Thang and Luong, 2014] provide argumentative characterisations of non-monotonic inference in Brewka’s Preferred Subtheories [Brewka, 1989]. The latter defines credulous and sceptical inference relations over a totally ordered set  $\mathcal{B}$  of classical wff. These are shown to correspond to the conclusions of arguments in at least one, respectively all, *stable* extensions of the *AFs* instantiated by  $\mathcal{B}$  in the aforementioned works. However, these works formalising standard accounts of classical logic argumentation (*Cl-Arg*) do not fully account for practical applications, as

the correspondences and satisfaction of rationality postulates [Caminada and Amgoud, 2007; Caminada *et al.*, 2012] assume construction of *all* classical logic arguments (which are in general infinite) defined by  $\mathcal{B}$ . Moreover, they enforce checking that the arguments’ premises are consistent and ‘subset minimal’ (no proper subset of the premises entail the conclusion). Hence, [D’Agostino and Modgil, 2016; 2018] provide a fully rational *dialectical* account of *Cl-Arg* that makes minimal assumptions as to the arguments that need to be constructed from  $\mathcal{B}$ , and eschews the need to check for subset minimality and consistency of premises. They then give an argumentative characterisation (again by reference to stable extensions) of Preferred Subtheories inference defined under the assumption that agents have limited inferential capabilities.

This paper makes two key contributions to the study of these argumentative characterisations of Preferred Subtheories inference. *Firstly*, we show that for both the standard approach to *Cl-Arg* (as formalised in [Modgil and Prakken, 2013]), and the dialectical approach for use by resource bounded agents, the preferred and stable extensions coincide. Hence the above correspondences with Preferred Subtheories inferences are shown to hold under the preferred semantics. The significance of these results is that argumentative characterisations of non-monotonic inference pave the way for individual agent non-monotonic reasoning via the use of argument game proof theories for establishing membership of arguments in admissible extensions (e.g., [Modgil and Caminada, 2009]), and generalisation of these proof theories to dialogical models of distributed non-monotonic reasoning (e.g., [Prakken, 2005]). However, each admissible extension of an *AF* is not necessarily a subset of a stable extension, but is a subset of a preferred extension. Hence, correspondences for Preferred Subtheories inferences under the preferred semantics enables use of these argument games and dialogues for establishing Preferred Subtheories inference. *Secondly*, we discuss how use of the standard approach to *Cl-Arg* results in: 1) the ‘foreign commitment problem’ when applying argumentation in dialogical contexts [Caminada *et al.*, 2014], and; 2) overly sceptical results when defining argumentation based inferences under the grounded semantics. We show how both these problems are avoided in the dialectical approach. Avoiding the latter problem then means that the grounded semantics applied to the dialectical characterisation

more closely ‘approximates’ sceptical Preferred Subtheories inference (as explained in Section 4).

This paper is organised as follows. Section 2 reviews Preferred Subtheories and its argumentative characterisations in the standard and dialectical approaches to *Cl-Arg*. Sections 3 and Section 4 respectively establish the first and second contributions described above. We conclude in Section 5.

## 2 Background

### 2.1 Preferred Subtheories

*Preferred Subtheories* [Brewka, 1989; 1991] resolves inconsistencies amongst an ordered set of classical formulae so as to formalise both non-monotonic reasoning and belief revision. Henceforth,  $\vdash_c$  denotes the classical consequence relation, lower and upper case greek letters respectively denote individual, and sets of, first order formulae. For simplicity, examples will use a propositional language (lower case roman letters denote propositions).

**Definition 1**  $(\Gamma, \leq)$  is a default theory where  $\leq$  is a total ordering over the set  $\Gamma$  of first order classical wffs (well formed formulas).  $(\Gamma, \leq)$  can be represented as  $(\Gamma_1, \dots, \Gamma_n)$ , where  $\Gamma = \bigcup_{i=1}^n \Gamma_i$  and  $\forall \alpha \in \Gamma_i, \forall \beta \in \Gamma_j : i < j$  iff  $\alpha > \beta$ . Hence  $\alpha, \beta \in \Gamma_i$  iff  $\alpha \simeq \beta$ <sup>1</sup>.

A preferred subtheory (*ps*) is a set  $\Sigma = \Sigma_1 \cup \dots \cup \Sigma_n$  such that for  $i = 1, \dots, n$ ,  $\Sigma_1 \cup \dots \cup \Sigma_i$  is a maximal (under set inclusion) consistent subset of  $\Gamma_1 \cup \dots \cup \Gamma_i$ .

$\phi$  is a sceptical (credulous) *ps*-inference from  $(\Gamma, \leq)$ , denoted  $(\Gamma, \leq) \vdash_{ps}^{sc} \phi$  ( $(\Gamma, \leq) \vdash_{ps}^{cr} \phi$ ) iff  $\phi$  is classically entailed by all (respectively at least one) *ps* of  $(\Gamma, \leq)$

Intuitively, a *ps* is obtained by taking a maximal under set inclusion consistent subset (*mcs*) of  $\Gamma_1$ , extending this to a *mcs* of  $\Gamma_1 \cup \Gamma_2$ , and so on.

**Example 1** The default theory  $(\Gamma_1 = \{-a \vee \neg b\}, \Gamma_2 = \{a, b, a \wedge b \rightarrow \neg c\}, \Gamma_3 = \{c\})$  has two *ps*s:  $\Sigma = \{-a \vee \neg b, a, a \wedge b \rightarrow \neg c, c\}$  and  $\Sigma' = \{-a \vee \neg b, b, a \wedge b \rightarrow \neg c, c\}$ .

### 2.2 Preferred Subtheories as an instance of Classical Logic Argumentation

Given a set of first order wffs  $\Gamma$ , classical logic arguments are tuples  $(\Delta, \alpha)$  such that  $\Delta \subseteq \Gamma$ ,  $\Delta \vdash_c \alpha$ ,  $\Delta$  is ‘subset minimal’ (no proper subset of  $\Delta$  entails  $\alpha$ ) and  $\Delta \not\vdash_c \perp$  [Besnard and Hunter, 2008]. Henceforth,  $\text{conc}(X) = \alpha$ ,  $\text{prem}(X) = \Delta$ , and  $\text{prem}(S) = \bigcup_{X \in S} \text{prem}(X)$ . [Modgil and Prakken, 2013] then use the total ordering  $\leq$  over  $\Gamma$  to define strict preferences over arguments, and hence defeats:

**Definition 2** Let  $\mathcal{A}$  be all arguments defined by  $\Gamma$  in the default theory  $(\Gamma, \leq)$ .  $\forall X, Y \in \mathcal{A}$ :

- $(X, Y) \in \mathcal{C}$  (i.e.,  $X$  attacks  $Y$ ) iff  $\text{conc}(X) = \phi$ ,  $-\phi \in \text{prem}(Y)$ , where for any  $\phi$ , if  $\phi$  is of the form  $\neg\alpha$  then  $-\phi = \alpha$ , else  $-\phi = \neg\phi$ .

$X$  is said to attack  $Y$  on the argument  $Y' = (\{-\phi\}, -\phi)$  (abusing notation we may write ‘ $X$  attacks  $Y$  on  $-\phi$ ’).

- $X \prec Y$  iff  $\exists \alpha \in \text{prem}(X), \forall \beta \in \text{prem}(Y): \alpha < \beta$

<sup>1</sup> $x < y$  iff  $x \leq y$ ,  $y \not\leq x$ , and  $x \simeq y$  iff  $x \leq y$  and  $y \leq x$ .

- $(X, Y) \in \Rightarrow$  (i.e.,  $X$  defeats  $Y$ , denoted  $X \Rightarrow Y$ ) iff  $X$  attacks  $Y$  on  $Y'$ , and  $X \not\prec Y'$ .  $X \Rightarrow Y$  denotes  $(X, Y) \notin \Rightarrow$ .
- The extensions of the *argument framework* (*AF*)  $(\mathcal{A}, \Rightarrow)$  defined by  $(\Gamma, \leq)$  are defined in the usual way [Dung, 1995]:

Let  $E \subseteq \mathcal{A}$ ,  $X \in \mathcal{A}$ . Then  $X$  is acceptable w.r.t.  $E$  iff  $\forall Y$  s.t.  $Y \Rightarrow X$  implies  $\exists Z \in E$  s.t.  $Z \Rightarrow Y$ .

$E \subseteq \mathcal{A}$  is conflict free iff  $\forall X, Y \in E, X \not\Rightarrow Y$ . A conflict free  $E$  is an *admissible* extension iff  $\forall X \in E$ ,  $X$  is acceptable w.r.t.  $E$ . An admissible  $E$  is a *complete* extension if  $X$  is acceptable w.r.t.  $E$  implies  $X \in E$ .  $E$  is a *preferred* extension if  $E$  is a maximal (under set inclusion) complete extension.  $E$  is the *grounded* extension if  $E$  is the minimal (under set inclusion) complete extension.  $E$  is a *stable* extension if  $E$  is conflict free and  $\forall Y \notin E, \exists X \in E$  s.t.  $X \Rightarrow Y$ .

- For  $S \in \{\text{preferred}, \text{grounded}, \text{stable}\}$ , the sceptical and credulous inferences are respectively defined as:  
 $(\mathcal{A}, \Rightarrow) \vdash_S^{sc} \phi$  if every  $S$  extension contains an argument concluding  $\phi$ ;  
 $(\mathcal{A}, \Rightarrow) \vdash_S^{cr} \phi$  if at least one  $S$  extension contains an argument concluding  $\phi$ .

**Example 2** Let  $(\mathcal{A}, \Rightarrow)$  be the *AF* defined by the default theory  $(\Gamma, \leq)$  in Example 1.  $\mathcal{A}$  consists of all classical inferences from consistent subsets of  $\Gamma$ , including:

- $A_1 = (\{b, \neg a \vee \neg b\}, \neg a)$ ,  $A_2 = (\{b, a \wedge b \rightarrow \neg c, c\}, \neg a)$ ,  
 $B_1 = (\{a, \neg a \vee \neg b\}, \neg b)$ ,  $B_2 = (\{a, a \wedge b \rightarrow \neg c, c\}, \neg b)$ ,  
 $C = (\{a, b\}, \neg(\neg a \vee \neg b))$ ,  $D = (\{a, b, c\}, \neg(a \wedge b \rightarrow \neg c))$ ,  
 $E = (\{a, b, a \wedge b \rightarrow \neg c\}, \neg c)$ , and;  
 $Y_1 = (\{\neg a \vee \neg b\}, \neg a \vee \neg b)$ ,  $Y_2 = (\{a\}, a)$ ,  $Y_3 = (\{b\}, b)$ ,  
 $Y_4 = (\{a \wedge b \rightarrow \neg c\}, a \wedge b \rightarrow \neg c)$ ,  $Y_5 = (\{c\}, c)$
- $A_1$  and  $A_2$  attack  $Y_2, B_1, B_2, C, D, E$  on  $Y_2$ .  
 $B_1$  and  $B_2$  attack  $Y_3, A_1, A_2, C, D, E$  on  $Y_3$ .  
 $C$  attacks  $Y_1, A_1, B_1$  on  $Y_1$ .  
 $D$  attacks  $Y_4, A_2, B_2, E$  on  $Y_4$ .  $E$  attacks  $Y_5, D$  on  $Y_5$ .

- Only the attacks from  $A_1, B_1$  and  $E$  succeed as defeats, since  $A_2 \prec Y_2$ ,  $B_2 \prec Y_3$ ,  $C \prec Y_1$  and  $D \prec Y_4$ . There are two stable extensions:  
 $E = \{X \in \mathcal{A} | \text{prem}(X) \subseteq \{-a \vee \neg b, a, a \wedge b \rightarrow \neg c, c\}\}$   
 $E' = \{X \in \mathcal{A} | \text{prem}(X) \subseteq \{-a \vee \neg b, b, a \wedge b \rightarrow \neg c, c\}\}$

We recall Theorem 34 in [Modgil and Prakken, 2013], employing the notation  $\text{args}(\Sigma) = \{X \in \mathcal{A} | \text{prem}(X) \subseteq \Sigma\}$ :

**Theorem 3** Let  $(\mathcal{A}, \Rightarrow)$  be the *AF* defined by  $(\Gamma, \leq)$ . Then: If  $E$  is a stable extension of  $(\mathcal{A}, \Rightarrow)$  then  $\text{prem}(E)$  is a preferred subtheory of  $(\Gamma, \leq)$ ;

If  $\Sigma$  is a preferred subtheory of  $(\Gamma, \leq)$  then  $\text{args}(\Sigma)$  is a stable extension of  $(\mathcal{A}, \Rightarrow)$ .

Hence there is a *ps*  $\Sigma$  of  $(\Gamma, \leq)$  such that  $\alpha$  is classically entailed from  $\Sigma$  iff there is an argument in a stable extension of  $(\mathcal{A}, \Rightarrow)$  that concludes  $\alpha$ . Hence:

$$\begin{aligned} (\Gamma, \leq) \vdash_{ps}^{sc} \phi &\text{ iff } (\mathcal{A}, \Rightarrow) \vdash_{stable}^{sc} \phi \\ (\Gamma, \leq) \vdash_{ps}^{cr} \phi &\text{ iff } (\mathcal{A}, \Rightarrow) \vdash_{stable}^{cr} \phi \end{aligned}$$

### 2.3 Dialectical Characterisation of Preferred Subtheories

The consistency and closure rationality postulates [Caminada and Amgoud, 2007], and results in Theorem 3, are satisfied

assuming that *all* arguments (which are in general infinite) that can be constructed from a set of formulae ('base')  $\Gamma$ , are included in an *AF*, and that the premises of each constructed argument is checked for consistency. These assumptions are clearly not feasible for real-world agents with bounded resources. Moreover, the inconsistency of arguments' premises is typically demonstrated dialectically (as in the classic Socratic move whereby an interlocutor is shown to commit to inconsistent premises). Hence [D'Agostino and Modgil, 2016; 2018] develop a *dialectical* account of *Cl-Arg* that is provably rational under minimal assumptions as to the arguments that need be constructed, and that eschews the need to check for consistency of premises; inconsistency is demonstrated dialectically. The key insight is to account for a distinction ubiquitous in dialectical practice, whereby an argument entails a conclusion from assumptions regarded as premises assumed to be true, and assumptions that are supposed true for the 'sake of argument' (i.e., those premises that an interlocutor commits to).

**Definition 3**  $X = (\Delta, \Phi, \alpha)$  is a *dialectical argument* defined by  $\Gamma$ , if  $(\Delta \cup \Phi) \subseteq \Gamma$ ,  $\Delta \cap \Phi = \emptyset$ , and  $\Delta \cup \Phi \vdash_c \alpha$ .

- $\Delta$ ,  $\Phi$  and  $\alpha$  are respectively  $X$ 's premises, suppositions and conclusion, denoted  $\text{prem}(X)$ ,  $\text{supp}(X)$  and  $\text{conc}(X)$ .  $X$  is *unconditional* if  $\text{supp}(X) = \emptyset$ . Also,  $\text{assum}(X) = \text{prem}(X) \cup \text{supp}(X)$  denotes  $X$ 's assumptions. The preceding notation is generalised to sets of arguments in the obvious way.
- Let  $[X] = \{(\Delta', \Phi', \alpha) \mid \Delta' \cup \Phi' = \Delta \cup \Phi\}$ . Then  $\forall Y, Z \in [X]$ ,  $Y$  and  $Z$  are said to be *logically equivalent*. That is to say,  $[X]$  denotes all 'epistemic variants' in which  $X$ 's assumptions are differentially labelled according to whether they are referenced as premises or suppositions.
- Any strict preference relation  $\prec$  over arguments is then assumed to be *invariant modulo logical equivalence (imle)*:  $\forall X' \in [X], \forall Y' \in [Y], X \prec Y$  iff  $X' \prec Y'$ . (i.e., the strength of an argument is invariant with respect to the distinction between premises and suppositions).

Any  $Z$  challenging the acceptability of  $X$  w.r.t. a set of arguments  $S$ , can suppose the truth of premises of any arguments in  $S \cup \{X\}$ . In turn, some  $Y \in S$  that defends  $X$ , can counter  $Z$  by supposing the truth of any premises in  $Z$ .

For example, suppose  $S$  contains  $X = (\{a\}, \emptyset, a)$  and  $Z = (\{b, \neg a \vee \neg b\}, \emptyset, \neg a)$  defeats  $X$  (assuming  $Z \not\prec X$ ). Then  $Y = (\{a\}, \{\neg a \vee \neg b\}, \neg b) \in S$  can, *without committing to the truth of  $\neg a \vee \neg b$  as a premise*, suppose the truth of  $Z$ 's premise  $\neg a \vee \neg b$  to defeat  $Z$  on  $b$ , and so defend  $X$ . Also, suppose some  $S$  containing arguments  $X_1$  and  $X_2$  that respectively have premises  $\neg a \vee \neg b$  and  $a, b$ . Then  $Z = (\emptyset, \{\neg a \vee \neg b, a, b\}, \perp)$  dialectically demonstrates that  $X_1$  and  $X_2$  in  $S$  make use of inconsistent premises, by attacking  $X_1$  on  $\neg a \vee \neg b$  and  $X_2$  on  $a$  and  $b$ . These attacks succeed as defeats independently of preferences since such a demonstration should clearly not be contingent on preferences. One cannot then defend against  $Z$ 's defeats as  $Z$  cannot be defeated by virtue of its empty premises. Now, suppose  $Z' = (\{\neg a \vee \neg b, a, b\}, \emptyset, \neg c)$  defeats  $(\{c\}, \emptyset, c) \in S$ . The 'contaminating' effect of  $Z'$ , exploiting the explosivity of its inconsistent premises to entail the syntactically disjoint

$\neg c$ , is avoided by  $Z \in S$  defeating  $Z'$  (independently of preferences given that  $Z$  demonstrates the inconsistency of  $Z'$ 's premises). Note that  $Z$  can be included in any admissible extension as it cannot be defeated given its empty premises.

**Definition 4** Let  $X = (\Delta, \Phi, \alpha), Y = (\Pi, \Sigma, \beta) \in \mathcal{A}$ . Then:

- if  $\alpha = \perp$ ,  $X$  attacks  $Y$  on any  $\phi \in \Phi \cap \Pi$ ;
- if  $\alpha \neq \perp$ ,  $X$  attacks  $Y$  on  $-\alpha \in \Pi$ .
- The defeat relation  $\mathcal{D} \subseteq 2^{\mathcal{A}}$  is defined as follows:  $(X, Y) \in \mathcal{D}$  if  $X$  attacks  $Y$  on  $\phi$ , and either  $X$  is of the form  $(\emptyset, \Phi, \perp)$ , or  $X \not\prec (\{\phi\}, \emptyset, \phi)$ .
- Then  $X$  *defeats*  $Y$  with respect to  $S$ , denoted  $X \Rightarrow_S Y$ , if  $(X, Y) \in \mathcal{D}$  and  $\text{supp}(X) \subseteq \text{prem}(S \cup \{Y\})$ .
- $Y$  is acceptable w.r.t.  $S$  if  $\forall X$  s.t.  $X \Rightarrow_S Y$ ,  $\exists Z \in S$  s.t.  $Z \Rightarrow_{\{X\}} X$ . One then defines conflict free sets (with ' $\Rightarrow_E$ ' replacing ' $\Rightarrow$ '), and admissible, complete, preferred and stable (with ' $\Rightarrow_{\{Y\}}$ ' replacing ' $\Rightarrow$ ') extensions of a *dialectical AF*  $(\mathcal{A}, \mathcal{D})$  in the same way as in Definition 2. Then:
- $(\mathcal{A}, \mathcal{D}) \vdash_{\mathcal{S}}^{\text{sc}} \phi$  ( $(\mathcal{A}, \mathcal{D}) \vdash_{\mathcal{S}}^{\text{ct}} \phi$ ) if every (at least one)  $\mathcal{S}$  extension contains an *unconditional* argument concluding  $\phi$ .<sup>2</sup>

Suppose  $\mathcal{A}$  is any set of dialectical arguments defined by  $\Gamma$ , that satisfies the following:

- P1 If  $\alpha \in \Gamma$  then  $(\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$ ;
- P2 If  $X \in \mathcal{A}$  then  $\forall X' \in [X], X' \in \mathcal{A}$  (i.e, if one can construct  $X$  from  $\Gamma$ , then one has access to all logical equivalents of  $X$  that differentially label the assumptions);
- P3 If  $(\Delta, \Gamma, \alpha), (\Pi, \Phi, -\alpha) \in \mathcal{A}$  then  $(\Delta \cup \Pi, \Gamma \cup \Phi, \perp) \in \mathcal{A}$  (i.e, if one can construct arguments with conflicting conclusions then one can combine their assumptions to yield an argument concluding  $\perp$ ).

Then [D'Agostino and Modgil, 2018] show satisfaction of the consistency and closure postulates assuming *any* preference relation. The crash resistance and non-interference rationality postulates [Caminada *et al.*, 2012] are satisfied if: 1) whenever  $X$  in  $\mathcal{A}$  includes premises redundantly used to entail the conclusion (since subset minimality is not enforced) by virtue of these premises being syntactically disjoint from the remaining premises and the conclusion, then resources suffice to construct the non-redundant argument excluding these premises; 2)  $\prec$  is such that adding redundant premises does not strengthen an argument (this is satisfied by  $\prec$  as defined in Definition 2).

One can then use *dialectical Cl-Arg* to give a fully rational characterisation of Preferred Subtheories inference for agents who do not have unlimited inferential capabilities:

**Definition 5** Let  $\vdash_r \subseteq \vdash_c$  be any resource-bounded approximation of classical consequence such that  $\forall \Delta$ : 1) if  $\beta \in \Delta$  then  $\Delta \vdash_r \beta$ ; 2) if  $\Delta \vdash_r \alpha$  and  $\Delta \vdash_r \neg \alpha$  then  $\Delta \vdash_r \perp$ .  $\Delta$  is said to be *r-inconsistent* iff  $\Delta \vdash_r \perp$ ; else *r-consistent*. The *r-preferred subtheories* of  $(\Gamma, \leq)$  are then defined as in Definition 1, with 'r-consistent' replacing 'consistent'.

**Theorem 4** Let  $(\mathcal{A}, \mathcal{D})$  be the dialectical AF defined by  $(\Gamma, \leq)$ , where  $X \in \mathcal{A}$  iff  $\text{assum}(X) \vdash_r \text{conc}(X)$ , and  $\mathcal{D}$

<sup>2</sup>The inferences are conclusions of *unconditional* arguments that commit to the truth of *all* their assumptions as premises.

is defined based on the strict preference relation:  $X \prec Y$  if  $\exists \alpha \in \text{assum}(X), \forall \beta \in \text{assum}(Y), \alpha < \beta$ .

Then (letting  $\text{Args}(\Sigma)$  denote  $\{X \in \mathcal{A} \mid \text{prem}(X) \subseteq \Sigma\}$ ):

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  $(\Gamma, \leq)$ .

If  $\Sigma$  is an  $r$ -preferred subtheory of  $(\Gamma, \leq)$ , then  $E = \text{Args}(\Sigma)$  is a stable extension of  $(\mathcal{A}, \mathcal{D})$ .

Theorem 4 implies correspondences between resource-bounded Preferred Subtheories and argumentation based inference as formulated at the end of Section 2.2 (with ‘ $(\mathcal{A}, \mathcal{D})$ ’ substituting for ‘ $(\mathcal{A}, \Rightarrow)$ ’).

**Example 5** Assuming  $(\Gamma, \leq)$  from Example 1, let  $\vdash_r \subseteq \vdash_c$  be such that  $\mathcal{A}$  is the following set of dialectical arguments:

$A = (\{b, \neg a \vee \neg b\}, \emptyset, \neg a)$ ,  $B = (\{a, \neg a \vee \neg b\}, \emptyset, \neg b)$ ,  
 $C = (\{a, b\}, \emptyset, a \wedge b)$ ,  $D = (\{a, b, \neg a \vee \neg b\}, \emptyset, \perp)$ ,  
 $Z' = (\{a, b, \neg a \vee \neg b\}, \emptyset, \neg c)$ ,  $E = (\{a, b, a \wedge b \rightarrow \neg c\}, \emptyset, \neg c)$ ,  
 $F = (\{a, b, a \wedge b \rightarrow \neg c, c\}, \emptyset, \perp)$ ,  $Y_1 = (\{\neg a \vee \neg b\}, \emptyset, \neg a \vee \neg b)$ ,  
 $Y_2 = (\{a\}, \emptyset, a)$ ,  $Y_3 = (\{b\}, \emptyset, b)$ ,  $Y_4 = (\{a \wedge b \rightarrow \neg c\}, \emptyset, a \wedge b \rightarrow \neg c)$ ,  
 $Y_5 = (\{c\}, \emptyset, c)$  and their logical equivalents (e.g.,  $Z = (\emptyset, \{a, b, \neg a \vee \neg b\}, \perp) \in [D]$ ).

$(A, Y_2) \in \mathcal{D}$  and  $(B, Y_3) \in \mathcal{D}$ , and so both  $(A, X), (B, X) \in \mathcal{D}$  for  $X \in \{C, D, Z', E, F\}$ . Also,  $(E, Y_5), (Z', Y_5) \in \mathcal{D}$  and  $(Z, D), (Z, Z') \in \mathcal{D}$ . There are two stable extensions:

$E = \{X \in \mathcal{A} \mid \text{prem}(X) \subseteq \{\neg a \vee \neg b, a, a \wedge b \rightarrow \neg c, c\}\}$   
 $E' = \{X \in \mathcal{A} \mid \text{prem}(X) \subseteq \{\neg a \vee \neg b, b, a \wedge b \rightarrow \neg c, c\}\}$

Note that each stable extension includes  $Z$ .

### 3 Preferred Equals Stable

The above correspondences allow one to identify credulous  $ps$  inferences with the conclusions of (unconditional) arguments in stable extensions of a (dialectical)  $AF$  defined by  $(\Gamma, \leq)$ . However, it would ideally suffice to identify these inferences with conclusions of arguments in *admissible* extensions, since: 1) there are numerous works establishing argument game proof theories and algorithms for deciding membership of admissible extensions (e.g., [Modgil and Caminada, 2009]), which moreover need not consider *all* arguments and attacks/defeats in an  $AF$  (unlike when deciding membership of a stable extension; see footnote 3); 2) argumentation based dialogues formalising distributed non-monotonic reasoning typically involve showing membership of an argument in an admissible extension (of the  $AF$  incrementally defined by contents of locutions, e.g., [Prakken, 2005]). However, it is not always the case that an admissible extension is a subset of a stable extension<sup>3</sup>. Hence, for practical purposes it would be useful to also establish the previous section’s correspondences for the preferred semantics. That is, if we can show that credulous  $ps$  inferences correspond to the conclusions of (unconditional) arguments in preferred extensions of a (dialectical)  $AF$ , we can then utilise the above argument game proof theories and dialogues for individual

<sup>3</sup>E.g., consider the  $AF$  consisting of  $X$  and  $Y$  where  $Y$  defeats itself. Then  $\{X\}$  is admissible (and preferred) but there is no stable extension. This also illustrates the need to account for the *whole*  $AF$  when deciding membership of a stable extension.

and distributed credulous  $ps$  reasoning, since *any admissible extension is a subset of a preferred extension*<sup>4</sup>. We now establish these correspondences for the standard and dialectical approaches, by showing that in both cases the preferred and stable extensions coincide.

#### 3.1 Result for Standard $Cl$ -Arg Characterisation

We now show that given an  $AF$  defined by  $(\Gamma, \leq)$  in Section 2.2, the preferred and stable extensions coincide. Straightforwardly, any stable extension of a Dung  $AF$  is preferred [Dung, 1995]. It therefore suffices to show that each preferred extension is stable. We make use of the following lemma.

**Lemma 6** Let  $E$  be a complete extension of an  $AF$   $(\mathcal{A}, \Rightarrow)$  defined by  $(\Gamma, \leq)$ . Then  $\forall X \in \mathcal{A}$  such that  $\text{prem}(X) \subseteq \text{prem}(E), X \in E$ .

**Proposition 7** Let  $(\mathcal{A}, \Rightarrow)$  be defined by  $(\Gamma, \leq)$ . If  $E$  is a preferred extension of  $(\mathcal{A}, \Rightarrow)$  then  $E$  is a stable extension of  $(\mathcal{A}, \Rightarrow)$ .

**PROOF.** Let  $E$  be a preferred extension of  $(\mathcal{A}, \Rightarrow)$ . Given Theorem 3, it suffices to show that  $\Sigma = \text{prem}(E)$  is a preferred subtheory ( $ps$ ). Firstly,  $\Sigma$  must be consistent, since if  $\Sigma \vdash_c \perp$ , then  $\forall \alpha: \Sigma \vdash_c \alpha, \neg \alpha$ . Hence  $\exists \Sigma', \Sigma'' \subseteq \Sigma$  such that (by Lemma 6)  $(\Sigma', \alpha), (\Sigma'', \neg \alpha) \in E$ , contradicting *direct consistency* (Theorem 14 in [Modgil and Prakken, 2013]). Now, let  $\Sigma = \Sigma_1, \dots, \Sigma_n$  where for  $i = 1, \dots, n$ ,  $\Sigma_i = \text{prem}(E) \cap \Gamma_i$ . Suppose for contradiction that  $\Sigma$  is not a  $ps$ . Without loss of generality, assume:

$$\text{for } j = 1, \dots, i-1, \Sigma_1 \cup \dots \cup \Sigma_{i-1} \text{ is a} \\ \text{maximal consistent subset of } \Gamma_1 \cup \dots \cup \Gamma_{i-1} \quad (1)$$

$$\left( \bigcup_{j=1}^{i-1} \Sigma_j \right) \cup \Sigma_i \text{ is consistent but not maximally consistent} \quad (2)$$

and so let  $\Lambda \neq \emptyset$  be any maximal subset of  $\Gamma_i \setminus \Sigma_i$  s.t.:

$$\bigcup_{j=1}^i \Sigma_j \cup \Lambda \not\vdash_c \perp. \quad (3)$$

$$\text{Let } E' = \{Y \mid \text{prem}(Y) \subseteq (\Lambda \cup \text{prem}(E)), Y \notin E\} \quad (4)$$

where since  $\Lambda \neq \emptyset, E' \neq \emptyset$ .

We show that  $E \cup E'$  is admissible, contradicting  $E$  is a preferred extension. We first show that  $E \cup E'$  is conflict free.

i) Suppose  $X \in E, Y \in E', X \Rightarrow Y$  on  $\alpha$ . It must be that  $\alpha \in \Lambda$  else  $X$  defeats some  $Z \in E$  on  $\alpha$ , contradicting  $E$  is conflict free. By Eq. 3,  $\bigcup_{j=1}^i \Sigma_j \not\vdash_c \neg \alpha$ . Hence,  $\exists \beta \in \text{prem}(X)$  s.t.  $\beta \in \Sigma_{j>i}$ . But then  $X \prec (\{\alpha\}, \alpha)$ , contradicting  $X \Rightarrow Y$ .

ii) Suppose  $X \in E, Y \in E', Y \Rightarrow X$  on  $\alpha$ . Since  $E$  is admissible,  $\exists Z \in E, Z \Rightarrow Y$ . But then this leads to a contradiction as shown in i).

iii) Suppose  $X \in E', Y \in E', X \Rightarrow Y$  on  $\alpha \in \Lambda$ . By Eq. 3,  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda \not\vdash_c \neg \alpha$ . Hence  $\exists \beta \in \text{prem}(X)$  s.t.  $\beta \in \Sigma_{j>i}$ .

<sup>4</sup>This result is shown for Dung  $AF$ s in [Dung, 1995], and for dialectical  $AF$ s in [D’Agostino and Modgil, 2018].

But then  $X \prec (\{\alpha\}, \alpha)$ , contradicting  $X \Rightarrow Y$ .

iv) Suppose  $X \in E', Y \in E', X \Rightarrow Y$  on  $\alpha \in \text{prem}(E)$ . Then  $\exists Z \in E, X \Rightarrow Z$  and we are in case ii).

We show that all arguments  $Y \in E'$  are acceptable w.r.t.  $E \cup E'$ . Consider the following mutually exclusive cases.

1)  $X \Rightarrow Y$  on  $\alpha$ , where at least one premise of  $X$  is in  $\Gamma_{j>i}$ . If  $\alpha \in \Lambda$ , then  $X \prec (\{\alpha\}, \alpha)$ , contradicting  $X \Rightarrow Y$ . Hence  $\alpha \in \text{prem}(E)$ . Hence  $\exists Z \in E, X \Rightarrow Z$  on  $\alpha$ . Since  $E$  is admissible,  $\exists W \in E$  s.t.  $W \Rightarrow X$ . Hence  $Y$  is acceptable w.r.t.  $E \cup E'$ .

2)  $X \Rightarrow Y$  on  $\alpha$ , where all of  $X$ 's premises are in  $\Gamma_j, j \leq i$ :

2.1) Suppose  $\text{prem}(X) \subseteq \text{prem}(E) \cup \Lambda$ . By Eq. 4,  $X \in E \cup E'$ , contradicting  $E \cup E'$  is conflict free.

2.2) Hence  $\exists \beta \in \text{prem}(X)$  s.t.  $\beta$  in  $\bigcup_{j=1}^i \Gamma_j - (\bigcup_{j=1}^i \Sigma_j \cup \Lambda)$ .

2.2.1) Let  $\beta \in \Gamma_j - \Sigma_j$  for some  $j = 1, \dots, i-1$ . By Eq.1:

$$\bigcup_{k=1}^j \Sigma_k \not\vdash \perp \text{ and } \bigcup_{k=1}^j \Sigma_k \cup \{\beta\} \vdash \perp.$$

Hence  $\bigcup_{k=1}^j \Sigma_k \vdash -\beta$  and  $\exists Z$  s.t.  $\text{prem}(Z) \subseteq \bigcup_{k=1}^j \Sigma_k$ ,  $\text{conc}(Z) = -\beta$ , and  $Z \Rightarrow (\{\beta\}, \beta)$  (it cannot be that  $Z \prec (\{\beta\}, \beta)$  given that  $\beta \in \Gamma_j - \Sigma_j$ ). Hence  $Z \Rightarrow X$ . Moreover, by Lemma 6,  $Z \in E$ . Hence  $Y$  is acceptable w.r.t.  $E \cup E'$ .

2.2.2) Let  $\beta \in \Gamma_i - (\Sigma_i \cup \Lambda)$ . Since  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda$  is maximally consistent (Eq. 3),  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda \vdash -\beta$ . By Eq. 4,  $\exists Z \in E \cup E', \text{prem}(Z) \subseteq \bigcup_{j=1}^i \Sigma_j \cup \Lambda$ ,  $\text{conc}(Z) = -\beta$ , and  $Z \Rightarrow (\{\beta\}, \beta)$  (it cannot be that  $Z \prec (\{\beta\}, \beta)$  since  $\beta \in \Gamma_i$ ). Hence  $Z \Rightarrow X$ , and  $Y$  is acceptable w.r.t.  $E \cup E'$ .

QED

### 3.2 Result for Dialectical Characterisation

We now show that given a dialectical  $AF$   $(\mathcal{A}, \mathcal{D})$  defined by  $(\Gamma, \leq)$  in Section 2.3, where  $\mathcal{A}$  satisfies P1–P3, then the preferred and stable extensions coincide (it is straightforward to show that stable implies preferred for dialectical  $AF$ s). We make use of Lemma 9 in [D'Agostino and Modgil, 2018]:

**Lemma 8** 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  $\text{prem}(X) \subseteq \text{prem}(E), X \in E$ .

**Proposition 9** Let  $(\mathcal{A}, \mathcal{D})$  be defined by  $(\Gamma, \leq)$ , where  $(\Delta, \Phi, \alpha) \in \mathcal{A}$  iff  $\Delta \cup \Phi \vdash_r \alpha$ . If  $E$  is a preferred extension of  $(\mathcal{A}, \mathcal{D})$  then  $E$  is a stable extension of  $(\mathcal{A}, \mathcal{D})$ .

PROOF. Let  $E$  be a preferred extension of  $(\mathcal{A}, \mathcal{D})$ . Given Theorem 4, it suffices to show that  $\Sigma = \text{prem}(E)$  is a  $r$ -preferred subtheory ( $r$ -ps). Firstly,  $\Sigma$  is  $r$ -consistent, since if  $\Sigma \vdash_r \perp$ , then  $\exists \Sigma' \subseteq \Sigma$  such that (by Lemma 8),  $(\Sigma', \emptyset, \perp) \in E$ , contradicting *direct consistency* (Theorem 11 in [D'Agostino and Modgil, 2018]).

Now, let  $\Sigma = \Sigma_1, \dots, \Sigma_n$  where for  $i = 1, \dots, n, \Sigma_i = \text{prem}(E) \cap \Gamma_i$ . Suppose for contradiction that  $\Sigma$  is not a  $r$ -ps. Without loss of generality assume that:

$$\text{for } j = 1, \dots, i-1, \Sigma_1 \cup \dots \cup \Sigma_{i-1} \text{ is a maximal } r\text{-consistent subset of } \Gamma_1 \cup \dots \cup \Gamma_{i-1} \quad (5)$$

$$\left( \bigcup_{j=1}^{i-1} \Sigma_j \right) \cup \Sigma_i \text{ is } r\text{-consistent but not maximally } r\text{-consistent} \quad (6)$$

and so let  $\Lambda \neq \emptyset$  be any maximal subset of  $\Gamma_i \setminus \Sigma_i$  s.t.:

$$\bigcup_{j=1}^i \Sigma_j \cup \Lambda \not\vdash_r \perp. \quad (7)$$

We show that there exists an admissible  $E^* \supset E \cup \{(\{\alpha\}, \emptyset, \alpha) \mid \alpha \in \Lambda\}$ , contradicting  $E$  is a preferred extension of  $(\mathcal{A}, \mathcal{D})$ . Let

$$E^* = E \cup \{(\{\alpha\}, \emptyset, \alpha) \mid \alpha \in \Lambda\} \quad (8)$$

$$E^* = E' \cup \{(\Delta, \Phi, \phi) \mid \Delta \subseteq \text{prem}(E'), \Delta \cup \Phi \vdash_r \phi\} \quad (9)$$

We first show that  $E^*$  is conflict free. Suppose for contradiction that  $X, Y \in E^*, X \Rightarrow_{E^*} Y$  on  $(\{\beta\}, \emptyset, \beta)$ . By construction of  $E^*$ , and  $\text{supp}(X) \subseteq \text{prem}(E^*)$  (given  $Y \in E^*$  and  $X \Rightarrow_{E^*} Y$ ),  $\text{assum}(X) \cup \{\beta\} \subseteq \text{prem}(E')$ .

- Suppose  $\beta \in \bigcup_{j=1}^i \Sigma_j \cup \Lambda$  (recall that  $\Lambda \subseteq \Gamma_i$ ). Then  $\text{assum}(X) \subseteq \bigcup_{j=1}^i \Sigma_j \cup \Lambda$  (else  $X \not\Rightarrow_{E^*} Y$ ). If  $\text{conc}(X) = \perp$  then this contradicts Eq. 7. If  $\text{conc}(X) = -\beta$ , then  $\beta, -\beta$  and so  $\perp$  are all  $\vdash_r$  entailed from  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda$ , contradicting Eq. 7.

- Suppose  $\beta \in \bigcup_{j=i+1}^n \Sigma_j$  and so  $\beta \in \text{prem}(E)$ , and by P1  $B = (\{\beta\}, \emptyset, \beta) \in \mathcal{A}$ . Let  $X = (\Pi, \Omega, \phi)$ .

Suppose that  $(\Pi \cup \Omega) \cap \Lambda = \emptyset$ . Then  $\Pi \cup \Omega \subseteq \text{prem}(E)$ , and by P2,  $X' = (\Pi \cup \Omega, \emptyset, \phi) \in \mathcal{A}$ . By Lemma 8,  $X', B \in E$ . But then  $X' \Rightarrow_E B$ , contradicting  $E$  is conflict free. Hence  $(\Pi \cup \Omega) \cap \Lambda \neq \emptyset$ .

Let  $\Theta = (\Pi \cup \Omega) \cap \Lambda$ . By P2,  $X' = (\Theta, (\Pi \cup \Omega) \setminus \Theta, \phi) \in \mathcal{A}$ . Since  $(\Pi \cup \Omega) \setminus \Theta \subseteq \text{prem}(E)$  and  $\beta \in \text{prem}(E)$ , we have that  $X' \Rightarrow_E Y'$  on  $\beta$ , for some  $Y' \in E$ . Since  $E$  is admissible,  $\exists Z \in E$  s.t.  $Z \Rightarrow_{\{X'\}} X'$  on some  $\alpha \in \Theta \subseteq \Lambda$ . Hence  $Z \not\prec (\{\alpha\}, \emptyset, \alpha)$  and so  $\text{assum}(Z) \subseteq \bigcup_{j=1}^i \Sigma_j \cup \Lambda$ . If  $\text{conc}(Z) = \perp$  then this contradicts Eq. 7. If  $\text{conc}(Z) = -\alpha$ , then  $-\alpha, \alpha$  and so  $\perp$  are  $\vdash_r$  entailed by  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda$ , contradicting Eq. 7.

We now show that if  $Y \in E^*$  then  $Y$  is acceptable w.r.t.  $E^*$ :

1) Suppose  $X = (\Delta, \Psi, \phi) \Rightarrow_{E^*} Y$  on  $\delta \in \text{prem}(E)$ . By P1 and Lemma 8:  $Y' = (\{\delta\}, \emptyset, \delta) \in E$ .

1.1) Suppose  $X \Rightarrow_E Y'$ . By the admissibility of  $E$ ,  $\exists Z \in E$  (and so  $Z \in E^*$ ) such that  $Z \Rightarrow_{\{X'\}} X'$ .

1.2) Suppose  $X \not\Rightarrow_E Y'$ . Since  $\Lambda = \text{prem}(E^*) \setminus \text{prem}(E)$ , then  $(\Psi^\Lambda = \Psi \cap \Lambda) \neq \emptyset$ , and:

$$X' = (\Delta \cup \Psi^\Lambda, \Psi \setminus \Psi^\Lambda, \phi) \in \mathcal{A} \text{ (by P2) and } X' \Rightarrow_E Y' \quad (10)$$

By admissibility of  $E$ ,  $\exists Z = (\Pi, \Phi, \gamma) \in E, Z \Rightarrow_{\{X'\}} X'$ .

1.2.1) Suppose  $Z \Rightarrow_{\{X'\}} X'$  on  $\alpha \in \Psi^\Lambda$ . Hence  $Z \not\prec (\{\alpha\}, \emptyset, \alpha)$ . Since  $\alpha \in \Gamma_i$ , it must be that  $\Pi \subseteq \bigcup_{j=1}^i \Sigma_j$ .

Moreover, if  $\Phi \subseteq \bigcup_{j=1}^i \Sigma_j \cup \Lambda$ , then either  $\perp$  is  $\vdash_r$  entailed from  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda$  (in the case that  $\gamma = \perp$ ), contradicting Eq. 7, or (in the case that  $\gamma = -\alpha$ ),  $-\alpha, \alpha$  and so  $\perp$  are  $\vdash_r$

entailed by  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda$ , contradicting Eq. 7.

Hence,  $\exists \beta \in \Phi$  s.t.  $\beta \in \bigcup_{j=1}^i (\Gamma_j \setminus \Sigma_j) \setminus \Lambda$  (if  $\beta \in \Gamma_{j>i}$  then this would mean  $Z \prec (\{\alpha\}, \emptyset, \alpha)$  and so  $Z \not\Rightarrow_{\{X'\}} X'$ ). Moreover, since  $\beta \notin \Lambda$ , then  $\beta \notin \Psi^\Lambda$ , and so  $\beta \in (\Phi \cap \Delta)$ .

If **1.2.1.1**)  $\beta \notin \Gamma_i$ , then by Eq. 5,  $\bigcup_{j=1}^{i-1} \Sigma_j \cup \{\beta\} \vdash_r \perp$ , else if **1.2.1.2**)  $\beta \in \Gamma_i$ , then by Eq. 7:  $\bigcup_{j=1}^i \Sigma_j \cup \Lambda \cup \{\beta\} \vdash_r \perp$ .

In either case,  $\exists \Omega \subseteq \text{prem}(E^*)$  s.t.  $\Omega \cup \{\beta\} \vdash_r \perp$ , where  $\Omega \subseteq \bigcup_{j=1}^i \Sigma_j \cup \Lambda$ . By construction of  $E^*$ ,  $(\Omega, \{\beta\}, \perp) \in E^*$ . Moreover,  $(\Omega, \{\beta\}, \perp) \Rightarrow_{\{X'\}} X'$  on  $\beta \in \Delta$ . Hence  $(\Omega, \{\beta\}, \perp) \Rightarrow_{\{X\}} X$  on  $\beta \in (\text{prem}(X) = \Delta)$ .

**1.2.2**) Suppose  $Z \Rightarrow_{\{X'\}} X'$  on some  $\alpha \in \Delta$ . By P2,  $Z' = (\Pi \cup \Psi^\Lambda, \Phi \setminus \Psi^\Lambda, \gamma) \in \mathcal{A}$ . By construction of  $E^*$ ,  $Z' \in E^*$ . Since  $(\Phi \setminus \Psi^\Lambda) \subseteq \Delta$ ,  $Z' \Rightarrow_{\{X\}} X$  on  $\alpha$ .

2) Suppose  $X = (\Delta, \Psi, \phi) \Rightarrow_{E^*} Y \in E^*$ , on some  $\delta \notin \text{prem}(E)$ . Hence  $\delta \in \Lambda$ , and so  $\Delta \cup \Psi \subseteq \bigcup_{j=1}^i \Gamma_j$  (else  $X \prec (\{\delta\}, \emptyset, \delta)$ ). Then either  $\Delta \cup \Psi \cup \{\delta\} \vdash_r \perp$  (in the case that  $\phi = \perp$ ) or  $\Delta \cup \Psi \cup \{\delta\} \vdash_r \neg\delta, \delta$  and so  $\perp$  (in the case that  $\phi = \neg\delta$ ). Hence, by Eq. 7, it must be that  $\exists \beta \in \Delta \cup \Psi$  s.t.  $\beta \in (\bigcup_{j=1}^{i-1} \Gamma_j \setminus \Sigma_j) \cup (\Gamma_i \setminus (\Sigma_i \cup \Lambda))$ . Hence, since  $\Psi \subseteq \text{prem}(E^*) \subseteq \bigcup_{i=1}^n \Sigma_i \cup \Lambda$ , then  $\beta \in \Delta$ . Then, as in 1.2.1.1 and 1.2.1.2 above,  $\exists (\Omega, \{\beta\}, \perp) \in E^*$  such that  $(\Omega, \{\beta\}, \perp) \Rightarrow_{\{X\}} X$  on  $\beta \in (\text{prem}(X) = \Delta)$ .

We have shown that if  $\Sigma = \text{prem}(E)$  is not a  $r$ -preferred subtheory, then this contradicts the assumption that  $E$  is a preferred extension QED

## 4 Foreign Commitment, Sceptical Inference and the Grounded Semantics

The distinction between premises and suppositions adopted in dialectical argumentation, solves the so called ‘foreign commitment’ problem which arises in dialogues in which distinct agents make commitments when exchanging arguments. Adapting an example in [Caminada *et al.*, 2014], suppose the default theory  $(\Gamma_1 = \{g, g \rightarrow (\neg a \vee \neg b)\}, \Gamma_2 = \{a, b\})$ , where  $g$  denotes ‘budget’, and  $a$  and  $b$  respectively denote ‘bob will attend conference A’ and ‘bob will attend conference B’. Using the standard approach to *Cl-Arg*, *Ag1* submits  $X = (\{a, b\}, a \wedge b)$ . However, given that the budget does not support attendance at both conferences, *Ag2* would ideally be able to defeat  $X$  on its *conclusion* with  $(\{g, g \rightarrow (\neg a \vee \neg b)\}, \neg(a \wedge b))$ . However, it is well known that attacks and defeats that target conclusions of classical logic arguments results in violation of the rationality postulates [Caminada and Amgoud, 2007; Modgil and Prakken, 2013]. Hence, *Ag2* is forced to commit to one of *Ag1*’s premises, by defeating  $X$  on its premises with either  $Y_1 = (\{a, g, g \rightarrow (\neg a \vee \neg b)\}, \neg b)$  or  $Y_2 = (\{b, g, g \rightarrow (\neg a \vee \neg b)\}, \neg a)$ . In either case *Ag2* is forced to commit to the truth of a proposition that she does not believe in, which is unnatural from the perspective of commitment in dialogue [Walton and Krabbe, 1995]. However, in dialectical *Cl-Arg*, *Ag2* can defeat  $X = (\{a, b\}, \emptyset, a \wedge b)$  with either  $Y_1 = (\{g, g \rightarrow (\neg a \vee \neg b)\}, \{a\}, \neg b)$  or  $Y_2 = (\{g, g \rightarrow (\neg a \vee \neg b)\}, \{b\}, \neg a)$ , neither of which commit to the truth of *Ag1*’s premises, by deploying them as suppositions.

Moreover, notice that the sceptical *ps* inferences in Example 1 include  $c$ , corresponding to  $c$  being an argumentation defined sceptical inference under the stable/preferred semantics in Examples 2 and 5. Now, ideally one would identify the sceptical *ps* inferences of  $(\Gamma, \leq)$  with the conclusions of arguments in the single *grounded* extension of the (dialectical) *AF* defined by  $(\Gamma, \leq)$ , since computing the grounded extension<sup>5</sup> is computationally far less demanding than computing sceptical inference under preferred/stable semantics [Dunne and Wooldridge, 2009]. However, in Example 2  $Y_5 = (\{c\}, c)$  is *not* in the grounded extension  $E_G$ . To see why, observe that although  $Y_1 = (\{\neg a \vee \neg b\}, \neg a \vee \neg b) \in E_G$ ,  $Y_1$  does not defend against  $E$ ’s defeat on  $Y_5$ , since  $Y_1$  does not *individually* target either  $a$  or  $b$  in  $E$ ’s premises. If one were to then allow attacks targeting multiple premises (i.e.,  $a$  and  $b$ ), then this leads to violation of the consistency postulates [Gorogiannis and Hunter, 2011]. Hence one can only defeat  $E$  and so defend  $Y_5$ , with  $A_1$  or  $B_1$ , neither of which are in  $E_G$ . On the other hand, in the dialectical approach one can effectively simulate attacks on multiple premises. In Example 5, the grounded extension includes  $Y_1 = (\{\neg a \vee \neg b\}, \emptyset, \neg a \vee \neg b)$  and the logical equivalents  $A' = (\{\neg a \vee \neg b\}, \{b\}, \neg a)$ ,  $B' = (\{\neg a \vee \neg b\}, \{a\}, \neg b)$  (which only commit to the truth of  $\neg a \vee \neg b$ ) of  $A$  and  $B$  respectively.  $A'$  and  $B'$  defeat  $E = (\{a, b, a \wedge b \rightarrow \neg c\}, \emptyset, \neg c)$  on  $a$  and  $b$  respectively, so defending  $Y_5 = (\{c\}, \emptyset, c)$  which is therefore in the grounded extension. Hence the grounded semantics for dialectical *Cl-Arg* are less sceptical than for the standard approach. That is to say, assuming (dialectical) *AFs*  $(\mathcal{A}, \Rightarrow)$   $((\mathcal{A}, \mathcal{D}))$  defined by default theories  $(\Gamma, \leq)$  (as in Sections 2.2 and 2.3), and writing ‘ $g$ ’ to denote *grounded*:  $\forall (\Gamma, \leq) (\mathcal{A}, \Rightarrow) \vdash_g^{sc} \phi$  implies  $(\mathcal{A}, \mathcal{D}) \vdash_g^{sc} \phi$ ;

$\exists (\Gamma, \leq)$  such that  $(\mathcal{A}, \mathcal{D}) \vdash_g^{sc} \phi$  and  $(\mathcal{A}, \Rightarrow) \not\vdash_g^{sc} \phi$ .

Unfortunately,  $(\Gamma, \leq) \vdash_{ps}^{sc} \phi$  iff  $(\mathcal{A}, \mathcal{D}) \vdash_g^{sc} \phi$  does not in general hold for dialectical *Cl-Arg* because of the well known syntax sensitivity of syntactic approaches to non-monotonic inference.

**Example 10** Suppose:  $\Gamma_1 = \{(\neg a \vee \neg b) \wedge d, (\neg a \vee \neg b) \wedge \neg d\}$ ;  $\Gamma_2 = \{a, b, (a \wedge b) \rightarrow \neg c, c\}$ . Then  $(\Gamma, \leq) \vdash_{ps}^{sc} c$  (there are four preferred subtheories, none of which contain both  $a$  and  $b$  and so entail  $\neg c$ ). However, the grounded extension of the dialectical *AF* no longer contains arguments with premise  $\neg a \vee \neg b$  (and suppositions  $\{a\}$  or  $\{b\}$ ) that then defeat  $(\{a, b, a \wedge b \rightarrow \neg c\}, \emptyset, \neg c)$  and so defend  $(\{c\}, \emptyset, c)$ . This is because any argument that commits to the truth of  $\neg a \vee \neg b$  in the premises, must also commit to  $d$  or  $\neg d$ , and the ordering (preferences) do not arbitrate between the latter. It is this somewhat unnatural coupling of the ‘fate’ of  $\neg a \vee \neg b$  with  $d$  ( $\neg d$ ) through use of the conjunction connective, that disrupts the correspondence between sceptical *ps* inference and sceptical inference under the grounded semantics.

<sup>5</sup>Which in both standard and dialectical approaches can (starting with the empty set) be computed by iteratively adding arguments acceptable w.r.t. the previous iteration, until reaching a fixed point that is the grounded extension.

## 5 Conclusions

Argumentative characterisations of Preferred Subtheories have been developed using both standard approaches to *Cl-Arg*, and dialectical approaches more suited for real-world resource-bounded agents. This paper is the first to establish these characterisations for the preferred semantics, by showing that the preferred and stable extensions of (dialectical) *AFs* instantiated by default theories coincide. These results imply that argument game proof theories and algorithms for deciding membership of arguments in admissible extensions, suffice for realising the practical use of argumentation in deciding credulous Preferred Subtheories inference by single agents, and in distributed reasoning amongst multiple agents. However, it should be noted that argument game proof theories for Dung *AFs* cannot straightforwardly be applied to dialectical *AFs*. In particular, argument games involve a proponent submitting arguments that define an admissible extension *S* (containing the argument whose membership in an admissible extension is being decided), so as to defend against arguments moved by an opponent that defeat proponent's submitted arguments (e.g., [Modgil and Caminada, 2009]). Future work is therefore required to adapt such proof theories so that defeats moved by the opponent are parameterised so as to reference (the premises of arguments in) the set *S* that the proponent has thus far committed to.

Moreover, we have shown that under the grounded semantics, the dialectical approach more closely approximates sceptical Preferred Subtheories inference, as compared with the standard approach. We also conjecture that a full correspondence fails only in 'pathological' cases in which syntactically unrelated formulae are conjoined. Future work will more precisely articulate and verify this conjecture. Finally, we note that [Thang and Luong, 2014] provide argumentative characterisations of Preferred Subtheories inference using assumption based argumentation [Bondarenko *et al.*, 1997], and do so for the more general case in which the preferred subtheories are defined by all total orderings that complete a given *partial* ordering over a set  $\Gamma$  of classical wff. It remains to study characterisations of this more general case using standard and dialectical approaches to *Cl-Arg*.

## References

- [Amgoud and Vesic, 2010] Leila Amgoud and Srdjan Vesic. Handling inconsistency with preference-based argumentation. In *Scalable Uncertainty Management: 4th International Conference, SUM 2010*, pages 56–69. Springer, 2010.
- [Besnard and Hunter, 2008] Phillipe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, 2008.
- [Bondarenko *et al.*, 1997] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.
- [Brewka, 1989] Gerhard Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.
- [Brewka, 1991] Gerhard Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge University Press, 1991.
- [Caminada and Amgoud, 2007] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.
- [Caminada *et al.*, 2012] Martin W.A. Caminada, Walter E. Carnielli, and Paul E. Dunne. Semi-stable semantics. *Logic and Computation*, 22(5):1207–1254, 2012.
- [Caminada *et al.*, 2014] Martin W.A. Caminada, Sanjay Modgil, and Nir Oren. Preferences and unrestricted rebut. In *Proc. International Conference on Computational Models of Argument: COMMA 2014*, pages 209–220, 2014.
- [D'Agostino and Modgil, 2016] Marcello D'Agostino and Sanjay Modgil. A rational account of classical logic argumentation for real-world agents. In *European Conference on Artificial Intelligence (ECAI 2016)*, pages 141–149, 2016.
- [D'Agostino and Modgil, 2018] Marcello D'Agostino and Sanjay Modgil. Classical logic, argument and dialectic. *Artificial Intelligence (to appear)* <https://nms.kcl.ac.uk/sanjay.modgil/AIJSubmission.pdf>, 2018.
- [Dung, 1995] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. *Artificial Intelligence*, 77:321–357, 1995.
- [Dunne and Wooldridge, 2009] Paul E. Dunne and Michael Wooldridge. Chapter 5 : Complexity of abstract argumentation. In I. Rahwan and G. Simari, editors, *Argumentation in AI*, pages 85–104. Springer, 2009.
- [Gorogiannis and Hunter, 2011] Nikos Gorogiannis and Anthony Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, 175(10):1479 – 1497, 2011.
- [Modgil and Caminada, 2009] Sanjay Modgil and Martin Caminada. Chapter 6 : Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in AI*, pages 105–129. Springer, 2009.
- [Modgil and Prakken, 2013] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.
- [Prakken, 2005] Henry Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.
- [Thang and Luong, 2014] Phan M. Thang and Huynh T. Luong. Translating preferred subtheories into structured argumentation. *Journal of Logic and Computation*, 24(4):831–849, 2014.
- [Walton and Krabbe, 1995] Douglas N. Walton and Eric C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, NY, USA, 1995.