ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-141

A Rational Account of Classical Logic Argumentation for Real-World Agents

M. D'Agostino and S.Modgil[1]

Abstract.

Classical logic based argumentation (*ClAr*) characterises single agent non-monotonic reasoning and enables distributed non-monotonic reasoning amongst agents in dialogues. However, features of *ClAr* that have been shown sufficient to ensure satisfaction of rationality postulates, preclude their use by resource bounded agents reasoning individually, or dialectically in real-world dialogue. This paper provides a new formalisation of *ClAr* that is both suitable for such uses and satisfies the rationality postulates. We illustrate by providing a rational dialectical characterisation of Brewka's non-monotonic Preferred Subtheories defined under the assumption of restricted inferential capabilities.

1 Introduction

Context. In Dung's seminal theory of argumentation [13], arguments are built from a possibly inconsistent knowledge base $\mathcal{B}$. Attacks between arguments are defined, and preferences over arguments can then be used to decide whether one argument successfully attacks (defeats) another [1, 18]. The graph of arguments and defeats is then evaluated, based on the intuitive principle that an argument is justified if all its defeaters are themselves defeated by justified arguments. $\mathcal{B}$'s argumentation defined consequences are then the justified arguments' conclusions, and have been shown to correspond to the consequence relations of a number of non-monotonic logics. For example, classical logic arguments [15] are pairs (Δ, α) built from a base $\mathcal{B}$ of classical wff, where the premises Δ are a *consistent* subset of $\mathcal{B}$ that classically entail the conclusion α, and no proper subset of Δ entails α. An argument X attacks Y if X's conclusion negates one of Y's premises. [2, 18] show that given preferences over arguments defined on the basis of a total ordering on $\mathcal{B}$, the argumentation defined consequences correspond to the non-monotonic consequences from $\mathcal{B}$ defined by Preferred Subtheories (*PS*) [4].

Argumentation's dialectical characterisation of non-monotonic consequence, and the intuitive, familiar nature of the evaluative principles, accounts for its widely advocated benefits in enabling individual agent reasoning, and distributed ('dialogical') reasoning amongst computational and/or human agents [19]. However, features of classical logic instantiations of Dung graphs (*ClAr*) posited to ensure satisfaction of rationality postulates [5, 6], preclude its use by resource-bounded agents reasoning dialectically[2], either as individuals or in real-world dialogues. Firstly, the consistency and subset minimality checks on arguments' premises incur prohibitive computational expense. Moreover, the inconsistency of arguments' premises are in real-world argumentation established dialectically, by showing that an interlocutor contradicts herself. On the other hand, the consistency check ensures satisfaction of the *non contamination* postulates [6]. Secondly, exclusively targeting attacks at an argument's premises leads to the so called 'foreign commitment problem' whereby an agent is forced to commit to the premises of his interlocutor when attacking his interlocutor's arguments [16]. However, allowing attacks on the conclusions of arguments results in violation of the *consistency* postulates [5]. Thirdly, consistency may also be violated unless one assumes that a Dung graph includes *all* arguments defined by a base $\mathcal{B}$. However, this further precludes the use of *ClAr* by resource-bounded agents.

Contributions This paper proposes a new account of *ClAr* that is suitable for resource bounded agents reasoning individually and in real-world dialogues, and is provably rational. We review background in Section 2, and then Section 3 presents our first contribution. We propose a new dialectical ontology for *ClAr* arguments that distinguishes amongst premises assumed true, and those assumed true 'for the sake of argument'. Agents are therefore not forced to commit to the premises of their interlocutors despite the fact that attacks are targeted at premises. We also accommodate the use of *ClAr* by resource bounded agents, by not requiring consistency or subset minimality checks on arguments' premises, and, subject to intuitive assumptions on available resources for constructing arguments, we allow for instantiation of Dung graphs by subsets of arguments defined by a base. Our formalisation also accommodates the real-world move whereby the mutual inconsistency of arguments' premises is demonstrated dialectically. We then provide an account of Preferred Subtheories that assumes limited inferential resources, and show that the defined non-monotonic consequence relation corresponds to the argumentation defined consequences obtained by our dialectical formalisation of *ClAr*. Section 4 presents our second contribution. We show that our approach satisfies key results that hold for Dung's theory[3] despite our conservative adaptation of Dung's evaluative principles. We also show that despite satisfaction of the above desiderata for real-world applications of *ClAr*, the consistency and closure postulates [5], as well as the *non contamination* postulates [6], are satisfied. Section 5 concludes by discussing related and future work.

2 Background

We review classical logic instantiations of Dung graphs (*ClAr*) [15, 18] that study [5]'s rationality postulates. We assume the propositional language $\mathcal{L}$ consisting of atoms $\bot, a, b, c, \ldots$ with the

[1] University of Milan, email: marcello.dagostino@unimi.it, and King's College London, email: sanjay.modgil@kcl.ac.uk

[2] By 'dialectic' we mean 'a method of examining and discussing opposing ideas in order to find the truth' (`www.merriam-webster.com`) .

[3] We will refer the reader to [11] where space limitations preclude full details of proofs in this paper.

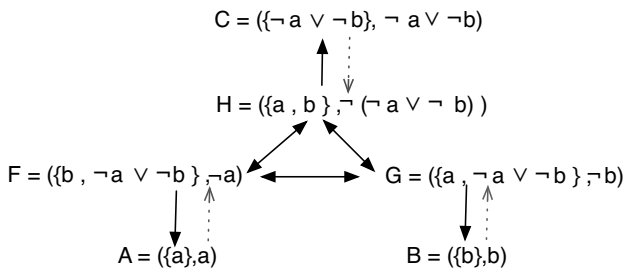

Figure 1. Attacks on premises are solid arrows. Dotted arrows are additional attacks if attacks can target conclusions.

usual connectives and definition of classical *wff*. Lower case and upper case Greek letters (as well as the symbol $\mathcal{B}$) respectively refer to arbitrary classical wff and finite sets of classical wff. We assume the complement function:

$\overline{\phi} = \psi$ if ϕ is of the form $\neg\psi$; else $\overline{\phi} = \neg\phi$

and let $Cn(\Delta)$ denote $\{\alpha \mid \Delta \vdash \alpha\}$, where $\vdash$ is the classical consequence relation. The arguments $\mathcal{A}$ defined by a base $\mathcal{B}$ of classical propositional wff, are pairs (Δ, α) where $\Delta \subseteq \mathcal{B}$, and: 1) the premises Δ are consistent; 2) $\Delta \vdash \alpha$; 3) no strict subset of Δ satisfies 2.

[15] study *ClAr* assuming variously defined notions of attacks. [18] additionally assume a strict argument preference relation $\prec \subseteq \mathcal{A} \times \mathcal{A}$ ($X \prec Y$ denotes Y strictly preferred to X), and define attacks and defeats as follows. For any $X = (\Delta, \alpha), Y = (\Gamma, \beta) \in \mathcal{A}$:

- X attacks Y (denoted $(X, Y) \in \mathcal{C}$) if $\alpha = \overline{\gamma}$ for some $\gamma \in \Gamma$, in which case X is said to attack Y on γ (or on $(\{\gamma\}, \gamma)$).
- X defeats Y (denoted $(X, Y) \in \mathcal{D}$) if X attacks Y on γ, and $X \not\prec (\{\gamma\}, \gamma)$ ($X \Rightarrow Y$ ($X \not\Rightarrow Y$) denotes that X does (not) defeat Y).

[18] study well known preference relations over arguments that are defined by an ordering $\leq$ over the formulae in $\mathcal{B}$ (where $<$ and $\sim$ are defined in the usual way). In particular, for any $X = (\Delta, \alpha), Y = (\Gamma, \beta)$:

$X \prec_{El} Y$ if $\exists \delta \in \Delta, \forall \gamma \in \Gamma$: $\delta < \gamma$ (*Elitist* preference)

A Dung argumentation framework (AF) is then a tuple $(\mathcal{A}, \mathcal{D})$. For any $E \subseteq \mathcal{A}$, $X \in \mathcal{A}$ is *acceptable* w.r.t. (i.e., *defended by*) E if $\forall Y$ s.t. Y defeats X, $\exists Z \in E$ s.t. Z defeats Y. The extensions (sets of justified arguments) can then be defined under Dung's semantics [13]:

Definition 1 Let $(\mathcal{A}, \mathcal{D})$ be a AF. Then $E \subseteq \mathcal{A}$ is conflict free if $\forall X, Y \in E$: $X \not\Rightarrow Y$. For any conflict free $E \subseteq \mathcal{A}$:

E is said to be an extension that is: *admissible* if every $X \in E$ is acceptable w.r.t. E; *complete* if admissible and every $X \in A$ that is acceptable w.r.t. E, is in E; *grounded* if E is the minimal (under set inclusion) complete extension; *preferred* if E is a maximal (under set inclusion) complete extension; *stable* if every $Y \notin E$ is defeated by an argument in E.

A correspondence then holds between the stable extensions of a $AF = (\mathcal{A}, \mathcal{D})$ defined by $(\mathcal{B}, \leq)$ (where $\leq$ is total, and $\mathcal{A}, \mathcal{C}, \prec_{El}, \mathcal{D}$ are defined as above), and the widely studied Preferred Subtheories (*PS*) non-monotonic consequence relations defined over $(\mathcal{B}, \leq)$ [4].

Definition 2 Let $(\mathcal{B}_1, \ldots, \mathcal{B}_n)$ be the stratification of $(\mathcal{B}, \leq)$ such that $\alpha \in \mathcal{B}_i, \beta \in \mathcal{B}_j, i < j$ iff $\beta < \alpha$. A preferred subtheory (ps) Σ is a set $\Sigma_1 \cup \ldots \cup \Sigma_n$ such that for $i = 1 \ldots n$, $\Sigma_1 \cup \ldots \cup \Sigma_i$ is a $\subseteq$-maximal consistent subset of $\mathcal{B}_1 \cup \ldots \cup \mathcal{B}_i$.

Intuitively, a ps is obtained by taking a $\subseteq$-maximal consistent subset of $\mathcal{B}_1$, extending this with a $\subseteq$-maximal consistent subset of $\mathcal{B}_2$, and so on. For example, $\Sigma = \{a, \neg a \vee \neg b\}$ and $\Sigma' = \{a, b\}$ are the ps of ($\mathcal{B}_1 = \{a\}, \mathcal{B}_2 = \{\neg a \vee \neg b, b\}$). [18] then show that:

Σ is a preferred subtheory of $(\mathcal{B}, \leq)$ iff E is a stable extension of the AF defined by $(\mathcal{B}, \leq)$, where $\Delta \subseteq \Sigma$ iff $(\Delta, \alpha) \in E$.

One then obtains a correspondence between the PS non-monotonic consequences and the argumentation defined consequences. These can be defined either sceptically:

$$\{\alpha \mid \forall \Sigma : \Sigma \vdash \alpha\} = \{\alpha \mid \forall E : \exists (\Delta, \alpha) \in E\}$$

or credulously, which involves selecting the classical consequences of a single *ps*, equivalently the conclusions of arguments in a single stable extension. For example, Figure 1 shows some of the arguments and attacks (represented as solid arrows) defined by ($\{a, b, \neg a \vee \neg b\}, b \sim \neg a \vee \neg b < a$). The ordering determines that $F \prec_{El} A$, hence $F \not\Rightarrow A, F \not\Rightarrow H, F \not\Rightarrow G$, and the remaining attacks succeed as defeats. Then $\{A, C, G\}$ and $\{A, H, B\}$ are, respectively, subsets of the two stable extensions E and E', and α is a conclusion of an argument in E iff $\alpha \in Cn(\{a, \neg a \vee \neg b\})$, α is a conclusion of an argument in E' iff $\alpha \in Cn(\{a, b\})$.

A number of features of *ClAr* ensure that rationality postulates for argumentation are satisfied.

Firstly, the consistency check on arguments' premises ensures satisfaction of the *non contamination* postulates [6]. Essentially, these postulates state that arguments built from syntactically disjoint subsets of $\mathcal{B}$ should not impact on each other's justification status. To illustrate, suppose $\mathcal{B} = \{p, \neg p, s\}$, and suppose we allow the 'inconsistent argument' $Y = (\{p, \neg p\}, \neg s)$ which attacks $X = (\{s\}, s)$. Then (assuming $\prec = \emptyset$) there is a complete extension (the grounded extension) that does not include X[4]. However, intuitively the status of X should not be affected by arguments built from the syntactically disjoint $\{p, \neg p\}$.

Secondly, exclusively targeting attacks on arguments' premises ensures satisfaction of the consistency postulates [5]. To see why, observe that if attacks on conclusions are additionally permitted (as illustrated by the dotted attacks in Figure 1), then (assuming $\prec = \emptyset$), one obtains an additional stable extension containing $\{A, B, C\}$, whose conclusions are mutually inconsistent.

Thirdly, in *ClAr* it is tacitly assumed that all arguments defined by $\mathcal{B}$ are included in the AF for evaluation. In particular:

$$\text{if } (\Gamma, \alpha) \in \mathcal{A}, \text{ then } \forall \gamma \in \Gamma, (\Gamma \setminus \{\gamma\} \cup \{\overline{\alpha}\}, \overline{\gamma}) \in \mathcal{A} \quad \text{(\textit{contraposition})}$$

is posited as sufficient for satisfaction of the consistency postulates. To illustrate, suppose in Figure 1's example, that the AF only includes F and A, and let $F \prec A$. Then neither A or F defeat each other, and both are contained in a complete extension E that violates consistency. However, assuming contraposition the AF must additionally include G and H. Moreover, [18] also show that if $F \prec A$, and $\prec$ satisfies properties that are deemed 'reasonable', then it must be that either $G \not\prec B$ or $H \not\prec C$, and so either G defeats F on B, or H defeats F on C. But then E cannot be complete. To see why, if

[4] In general, *all* arguments in an AF will be attacked by arguments built from inconsistent premises (given the *ex-falso* principle), and it is well known that $\emptyset$ is the grounded extension of an AF that contains no un-attacked arguments.

either G or H defeats F, then by assumption of E being complete, there must be an argument in E that defends F by defeating G or H. But then any such argument must also defeat A or F, and so E would not be conflict free. Hence, contraposition and reasonable preference relations are shown to guarantee satisfaction of consistency (note that [18] show that the *Elitist* preference is reasonable).

3 Dialectical Classical Logic Argumentation

3.1 Motivation

Apart from its intuitive characterisation of single agent reasoning, a key advantage of argumentation [19] is that it provides a formal basis for dialogue amongst computational and/or human agents [17]. Given argumentative characterisations of non-monotonic consequence relations (e.g. Logic Programming [13], Prioritised Default Logic [23] and Preferred Subtheories [18]), such dialogues effectively enable distributed non-monotonic reasoning amongst communicating agents. Agents submit arguments[5] in order to establish acceptance of the initial claim (a belief or decision option). Intuitively, the agent advocating the initial claim, attempts to build an admissible extension that includes an argument concluding the claim. In these dialogues, a base $\mathcal{B}_p$ is incrementally defined by the agents' 'public commitments'; that is, the contents of exchanged locutions (rather than assuming a given initial base in the case of single agent reasoning), and agents can construct arguments from premises in their private bases *and* the incrementally defined $\mathcal{B}_p$.

However, we argue that the three features of *ClAr* shown to ensure satisfaction of rationality postulates (as discussed in the previous section): 1) preclude use of *ClAr* by resource bounded agents reasoning individually or in dialogues, and; 2) preclude modelling features of dialectical reasoning that are ubiquitous in real-world dialogue.

Firstly, the tacit assumption that an AF is instantiated by all arguments defined by a base $\mathcal{B}$, is clearly not feasible for resource bounded agents, given that deciding whether $\Delta \vdash \alpha$ is in general NP-hard (hence most likely intractable). One would thus want to *identify as 'undemanding' a set of assumptions as possible on available resources for constructing arguments, such that rationality is preserved when AFs are not instantiated by all definable arguments* (**D1**).

In particular one would want to relax the contraposition condition. To illustrate, suppose arguments are classical *Intelim* natural deduction (*I-ND*) proofs [9]. *I-ND* allows parameterisation of proofs by the depth of nesting of discharged assumptions, such that stepwise increments in depth define a hierarchy of tractable inference relations, and each depth bounded system can be used to reflect the assumed inferential capabilities of real-world agents. Now, suppose $Ag1$ submits arguments whose premises include the inconsistent $\Pi = \{p, p \rightarrow \neg q, p \rightarrow q\}$. $Ag2$ can respectively attack these premises with $A = (\{p \rightarrow q, p \rightarrow \neg q\}, \neg p)$ or $B = (\{p, p \rightarrow q\}, \neg(p \rightarrow \neg q))$, or $C = (\{p, p \rightarrow \neg q\}, \neg(p \rightarrow q))$. Assuming $\prec$ is reasonable, one such attack must be a defeat. Hence $Ag1$ must defend itself by submitting an argument that defeats A or B or C. But then this argument must defeat one of $Ag1$'s own arguments on a premise in Π, and so $Ag1$ cannot construct an admissible set containing the arguments with premises Π. However, suppose neither of the attacks by B and C succeed, so that $Ag2$ must defeat with A. But then it may be that $Ag2$ has insufficient resources to construct A (indeed, in *I-ND*, constructing A requires greater nesting of discharged assumptions than B or C) and so $Ag1$ may be able to construct an admissible extension containing arguments with mutually inconsistent premises.

Furthermore, the computational non-viability of *ClAr* is further exacerbated by the checks on arguments' premises. Checking for consistency is of course as computationally demanding as deciding $\Delta \vdash \alpha$[6]. Moreover, the subset minimality check implies that for every constructed argument (Δ, α), one must in the worst case check that $\forall \Delta' \subset \Delta$, $\Delta' \nvdash_c \alpha$. Hence, for resource bounded agents one would want *a rational account of ClAr that does not require checking for consistency or subset minimality of premises* (**D2**).

Moreover, in real-world dialogues, the inconsistency of arguments' premises is typically established *dialectically*, via the well known Socratic move of demonstrating that an opponent's argument(s) rests on inconsistent premises [7, 21]. Also, in real-world dialogues one wants to avoid the anomaly of an agent being forced to commit to the premises of his interlocutor (known as the 'foreign commitment problem' [16]), which arises due to restricting attacks to targeting premises. To illustrate, consider an agent $Ag1$ submitting A in Figure 1. $Ag2$ counters with F. $Ag1$ cannot now counter F with A, but rather has to publically commit to a premise of his opponent (either b or $\neg a \vee \neg b$), by defending A with either H or G, and so having to possibly defend these premises from challenges by other agents. Hence, one would want *an account of ClAr that accommodates the dialectical demonstration that arguments' premises are inconsistent* (**D3**) and *avoids the foreign commitment problem* (**D4**).

3.2 Defining Dialectical Argumentation

We now formalise an account of *ClAr* that satisfies the desiderata **D1** – **D4**. Our starting point is the observation that when interlocutors construct arguments, they typically distinguish their own premises 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". This pattern is pervasive in real argumentation practice, and motivates the following definition of arguments in which we also drop the consistency and subset minimality checks. Attacks are then targeted only at premises and *not* suppositions. Also, arguments may now conclude $\perp$, and these can target any premise. However, letting $\mathsf{atoms}(\mathcal{B})$ denote the set of propositional atoms in $\mathcal{B}$, we henceforth assume finite bases $\mathcal{B}$ such that $\perp \notin \mathsf{atoms}(\mathcal{B})$ (i.e., $\perp$ is reserved as a notational device to express that an inconsistency has been reached in the course of constructing an argument).

Definition 3 A dialectical argument X defined by $\mathcal{B}$ is a triple (Δ, Γ, α) such that $(\Delta \cup \Gamma) \subseteq \mathcal{B}$ and $\alpha \in Cn(\Delta \cup \Gamma)$.

We say that Δ, Γ and α are, respectively, X's premises, suppositions and conclusion. We let $\mathsf{prem}(X) = \Delta$, $\mathsf{supp}(X) = \Gamma$, and generalise this notation to sets of arguments in the obvious way.

Also, if $Cn(\Delta \cup \Gamma) = \mathcal{L}$ then X is said to be *inconsistent*; else X is *consistent*. Finally, if $\mathsf{supp}(X) = \emptyset$ then X is said to be *unconditional*; else X is *conditional*.

Let $\mathcal{A}$ be the dialectical arguments defined by $\mathcal{B}$. Then:

$\mathcal{C} = \{(X, Y) \mid X, Y \in \mathcal{A}, X = (\Delta, \Gamma, \phi)(\phi = \alpha \text{ or } \phi = \perp), Y = (\Pi, \Sigma, \psi)$ and if $\phi = \alpha$ then $\overline{\alpha} \in \Pi\}$.

If $\phi = \alpha$, X is said to attack Y on premise $\overline{\alpha}$; equivalently, on the

[5] Arguments may be defined implicitly, e.g., $(\{q, q \rightarrow p\}, p)$ obtained by claiming q and (responding to a 'why' locution) asserting 'since q, $q \rightarrow p$'.

[6] Consistency checking is computationally hard not just for isolated or artificially constructed examples, but also in typical cases, as shown in [8].

argument $Y' = (\{\overline{\alpha}\}, \emptyset, \overline{\alpha})$. If $\phi = \bot$, X attacks Y on any $\beta \in \Pi$ (any $Y' = (\{\beta\}, \emptyset, \beta)$).

In dialogues, agents can suppose the truth of premises in their interlocutors' argument(s), when attacking their interlocutors' arguments. This motivates the following notion of *dialectical* attacks:

Definition 4 Let $S \subseteq \mathcal{A}$. Then $X \in \mathcal{A}$ *dialectically attacks* $Y \in \mathcal{A}$, *with respect to* S (denoted $X \rightarrow_S Y$) iff $(X, Y) \in \mathcal{C}$ **and** $\mathsf{supp}(X) \subseteq \mathsf{prem}(S)$.

Example 1 The dialectical arguments $\mathcal{A}$ defined by $\mathcal{B} = \{a, b, \neg a \vee \neg b\}$ include:

$A_1 = (\{a\}, \emptyset, a)$	$B_1 = (\{b\}, \emptyset, b)$
$C_1 = (\{\neg a \vee \neg b\}, \emptyset, \neg a \vee \neg b)$	$F_1 = (\{b, \neg a \vee \neg b\}, \emptyset, \neg a)$
$G_1 = (\{a, \neg a \vee \neg b\}, \emptyset, \neg b)$	$H_1 = (\{a, b\}, \emptyset, \neg(\neg a \vee \neg b))$
$F_2 = (\{b\}, \{\neg a \vee \neg b\}, \neg a)$	$G_2 = (\{a\}, \{\neg a \vee \neg b\}, \neg b)$
$H_2 = (\{a\}, \{b\}, \neg(\neg a \vee \neg b))$	$I_1 = (\{a, b, \neg a \vee \neg b\}, \emptyset, \bot)$
$I_2 = (\emptyset, \{a, b, \neg a \vee \neg b\}, \bot)$	

Notice that G_1 and G_2 are epistemically distinguished by the partitioning of premises and suppositions, but are 'logically equivalent'[7]. In what follows, we refer to preference relations that are *invariant modulo logical equivalence*.

Definition 5 Let $X = (\Delta, \Gamma, \alpha)$. 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.
• $\prec \subseteq \mathcal{A} \times \mathcal{A}$ is invariant modulo logical equivalence (*imle*) if $Y \prec X$ implies $\forall X' \in [X], \forall Y' \in [Y] : Y' \prec X'$

We now define dialectical defeat, acceptability and extensions under Dung's semantics. Any defeating argument Y challenging the acceptability of X w.r.t. E, can suppose the truth of premises in arguments in E, and any defense against Y can suppose the truth of premises in Y:

Definition 6 Let $(\mathcal{A}, \mathcal{C}, \prec)$ be a Dialectical Classical Framework (DCF) where $\mathcal{A}, \mathcal{C}$ are defined as in Definition 3, and $\prec$ is *imle*.
• $X \in \mathcal{A}$ defeats $Y \in \mathcal{A}$, *with respect to* $S \subseteq \mathcal{A}$ (denoted $X \Rightarrow_S Y$) iff $X \rightarrow_S Y$ on Y', and $X \not\prec Y'$.
• Let $E \subseteq \mathcal{A}$. Then $X \in \mathcal{A}$ is acceptable w.r.t. E iff $\forall Y \in \mathcal{A}$ s.t. $Y \Rightarrow_{E \cup \{X\}} X, \exists Z \in E$ s.t. $Z \Rightarrow_{\{Y\}} Y$.

Conflict free sets and extensions of DCFs are defined as in Definition 1, where $E \subseteq \mathcal{A}$ is now conflict free if for no $X, Y \in E$, $X \Rightarrow_{E \cup \{Y\}} Y$, and an extension is *stable* if $\forall Y \notin E, \exists X \in E$ s.t. $X \Rightarrow_{\{Y\}} Y$. The argumentation defined consequences are then the conclusions of *unconditional* arguments[8] in extensions of a DCF.

Example 2 (Example 1 cont.) Suppose $\{A_1, G_1, G_2, C_1\} \subseteq E$, $F_1 \not\prec A_1$. Then:
$F_1 \Rightarrow_{E \cup \{A_1\}} A_1$. Also, $F_2 \Rightarrow_{E \cup \{A_1\}} A_1$ since $\{\neg a \vee \neg b\} \subseteq \mathsf{prem}(E \cup \{A_1\})$ and $\prec$ is *imle*.
G_1 attacks F_1 and F_2 on b (on B_1), and $G_2 \rightarrow_{\{F_1\}} F_1$ on B_1, but $G_2 \not\rightarrow_{\{F_2\}} F_2$ since $\mathsf{supp}(G_2) \not\subseteq \mathsf{prem}(F_2)$.
Suppose $G_1 \not\prec B_1$. Hence, $G_1 \Rightarrow_{\{F_1\}} F_1, G_1 \Rightarrow_{\{F_2\}} F_2$, and since $\prec$ is *imle*, $G_2 \Rightarrow_{\{F_1\}} F_1$.

[7] In the sense that they are identical proofs distinguished only by the distinction between premises and suppositions. This is a stronger notion of equivalence than that which would apply to $(\{a, b\}, \emptyset, a \wedge b)$ and $(\{a \wedge b\}, \emptyset, a \wedge b)$.

[8] Since their conclusions are based only on premises assumed true, and not premises supposed true for the sake of argument.

Continuing with this example, suppose the Elitist preference $\prec_{El}$ defined by an ordering $\leq$ on $\mathcal{B}$. We illustrate how the desiderata **D1** – **D4** are satisfied. We will formally show satisfaction of the rationality postulates in Section 4

(**D4**) Suppose an admissible extension E_1 containing A_1, such that $F_1 \Rightarrow_E A_1$ (when a defeat is on $X \in E$ we will index the defeat with E rather than $E \cup \{X\}$). Now, rather than defending A_1 with G_1, it suffices to include G_2 in E_1 in order to defeat F_1. G_2 does not include as a premise (and so does not imply commitment to and the potential need to defend) the opponent's premise $\neg a \vee \neg b$.

(**D2**) Note that any argument in E_1 will be attacked (on any premise) by the inconsistent I_1. However, I_2, which has empty premises and so cannot be attacked (and is therefore said to be *unassailable*), is trivially acceptable w.r.t. any set of arguments, and so can be included in E_1. I_2 attacks I_1 (since $\mathsf{supp}(I_2) \subseteq \mathsf{prem}(I_1)$) on each of I_1's premises (i.e., on A_1, B_1 and C_1), and at least one of these attacks must succeed as a defeat. To suppose otherwise would mean that $I_2 \prec_{El} A_1$, $I_2 \prec_{El} B_1$ and $I_2 \prec_{El} C_1$. But it is easy to verify that these preferences hold only if we assume $\alpha < \alpha$ for some $\alpha \in \{a, b, \neg a \vee \neg b\}$, contradicting the irreflexivity of $<$. This illustrates how *non-contamination* is satisfied, despite arguments with inconsistent premises. Recall the example base $\{p, \neg p, s\}$ in Section 3.1. Then $X = (\{s\}, \emptyset, s)$ is in every complete extension E, since even though $I = (\{p, \neg p\}, \emptyset, \neg s)$ may defeat X, any such E will include the unassailable $(\emptyset, \{p, \neg p\}, \bot)$ which must (by the same reasoning as above) defeat I on p or $\neg p$, and so defend X.

(**D3**) Furthermore, we can now formalise the dialectical move whereby one shows that an interlocutor contradicts himself. Recall Section 3.1 and the arguments with inconsistent premises $p, p \rightarrow \neg q, p \rightarrow q$. Any E that includes these arguments cannot be admissible since given $I = (\emptyset, \{p, p \rightarrow \neg q, p \rightarrow q\}, \bot)$, then $I \rightarrow_E X$ for any $X \in E$, and (reasoning as above) either $I \Rightarrow_E (\{p\}, \emptyset, p)$ or $I \Rightarrow_E (\{p \rightarrow \neg q\}, \emptyset, p \rightarrow \neg q)$ or $I \Rightarrow_E (\{p \rightarrow q\}, \emptyset, p \rightarrow q)$. Since I is unassailable, no argument in E can defend against I.

(**D1**) The above illustrates that consistency is preserved, despite not having to assume all arguments defined under contraposition. We now define the notion of a *partially instantiated* DCF ($pDCF$), which makes a relatively undemanding (in terms of the required resources) set of assumptions as to the arguments that must be included in a DCF and that suffice to guarantee satisfaction of the rationality postulates. One can thus, for example, assume instantiation by a finite subset of the arguments defined by a base, and so accommodate uses of argumentation by real-world agents with limited resources. Before defining $pDCF$s, we introduce the following required notation:

Notation 3 $\mathcal{B} \| \mathcal{B}'$ denotes $\mathsf{atoms}(\mathcal{B}) \cap \mathsf{atoms}(\mathcal{B}') = \emptyset$ ($\mathcal{B}$ and $\mathcal{B}'$ are said to be *syntactically disjoint*). Also, $\mathcal{B}_{|At} = \{\alpha \in \mathcal{B} \mid \mathsf{atoms}(\{\alpha\}) \subseteq At\}$ (e.g., $\{\neg a \vee \neg b, c \wedge a\}_{|\{a,b\}} = \{\neg a \vee \neg b\}$).

Definition 7 $(\mathcal{A}, \mathcal{C}, \prec)$ is a *partially instantiated* DCF ($pDCF$) if $\mathcal{A}$ is any subset of the set of all arguments defined by a base $\mathcal{B}$, such that:

P1 $\forall \alpha \in \mathcal{B} : (\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$
P2 If $X \in \mathcal{A}$ then $\forall X' \in [X] : X' \in \mathcal{A}$
P3 If $(\Delta_1, \Gamma_1, \alpha) \in \mathcal{A}$, $(\Delta_2, \Gamma_2, \overline{\alpha}) \in \mathcal{A}$, then $(\Delta_1 \cup \Delta_2, \Gamma_1 \cup \Gamma_2, \bot) \in \mathcal{A}$.
P4 If $(\Delta \cup \Gamma, \emptyset, \alpha) \in \mathcal{A}$ and $\Delta \| \Gamma \cup \{\alpha\}$, then either $(\Delta, \emptyset, \bot) \in \mathcal{A}$ or $(\Gamma, \emptyset, \alpha) \in \mathcal{A}$.

P1 is self-explanatory. P2 expresses that given some X, additional resources are not required to assume construction of logically equivalent arguments (since these differ only in terms of the epistemic distinction between premises and suppositions). P3 is key for showing

consistency. It expresses that given arguments with conflicting conclusions, then resources suffice to combine their premises and suppositions to yield inconsistent arguments. To illustrate, in Example 1, suppose we only assume construction of the conflicting $A1$ and $F1$, and $F1 \prec A1$. By P3 and P2, we have the unassailable I_2 which must (reasoning as described earlier) defeat $F1$ or $A1$. Hence no admissible extension can include the conflicting $F1$ and $A1$, despite the absence of arguments defined under contraposition.

Finally, P4 is required to show satisfaction of the non-contamination postulates. To elaborate, since standard accounts of *ClAr* make no reference to specific proof theories for constructing arguments, they employ subset minimality as a somewhat 'blunt instrument' for ensuring that premises are relevant to deriving the argument's conclusion[9]. However, in practice agents clearly do not check for subset minimality. Rather, the proof theoretic means by which one entails a conclusion from premises, may ensure to varying degrees, the relevance of the premises for deriving the conclusion. Now, let us identify a notion of relevance that in principle can be satisfied by specific proof theories. Observe that by the properties of classical logic, if $\Delta \cup \Gamma \vdash \alpha$, $\Delta \| \Gamma \cup \{\alpha\}$, and α is not a tautology, then either α is provable from Γ (in which case Δ is redundant) or α must be provable from the inconsistent Δ by the explosivity of classical logic (in which case Γ is redundant). Of course, if α is a tautology, then $\Gamma \vdash \alpha$. Indeed, the *I-ND* natural deduction proof theory of [9] allows for a notion of proof that does not make use of syntactically disjoint premises; thus irrelevant proofs of this kind cannot be constructed (see [10, Definition 15, Theorem 9]). However, for proof theories that do allow such proofs, P4 simply states that if resources suffice to construct an argument that redundantly uses premises, then resources suffice to construct their non-redundant versions [10].

We now show that Dialectical *ClAr* characterises Preferred Subtheories, where the latter is now defined under the assumption that resources may not suffice to infer all classical consequences from a base.

Definition 8 Let $\vdash_r \subseteq \vdash$ be any resource bounded classical consequence relation, such that: 1) for any Δ, if $\beta \in \Delta$ then $\Delta \vdash_r \beta$; 2) if $\Delta \vdash_r \alpha$ and $\Delta \vdash_r \neg\alpha$ then $\Delta \vdash_r \bot$.

We say Δ is r-inconsistent iff $\Delta \vdash_r \bot$; r-consistent otherwise. A r-preferred subtheory of $(\mathcal{B}, \leq)$ is then defined as in Definition 2, with 'r-consistent' substituting for 'consistent'.

The following uses the notation $\mathtt{Args}(\Sigma) = \{X \mid \mathsf{prem}(X) \subseteq \Sigma\}$.

Theorem 4 *Let $(\mathcal{A}, \mathcal{C}, \prec_{El})$ be a* pDCF *defined by $(\mathcal{B}, \leq)$, such that $(\Delta, \Gamma, \alpha) \in \mathcal{A}$ iff $\Delta \cup \Gamma \vdash_r \alpha$. Then:*

1) Σ is a r-preferred subtheory of $(\mathcal{B}, \leq)$ implies $E = \mathtt{Args}(\Sigma)$ is a stable extension of $(\mathcal{A}, \mathcal{C}, \prec_{El})$.

2) E is a stable extension of $(\mathcal{A}, \mathcal{C}, \prec_{El})$ implies $\Sigma = \bigcup_{X \in E} \mathtt{Prem}(X)$ is a r-preferred subtheory of $\mathcal{B}$.

PROOF.

***Proof of 1)**: Suppose for contradiction that E is not conflict free. Then $X, Y \in E$, $X = (\Delta, \Gamma, \phi)$ ($\phi = \bot$ or β), $X \Rightarrow_E Y$ on $\overline{\beta} \in \mathsf{prem}(Y)$. We have $\Sigma \vdash_r \overline{\beta}$. Since $\Gamma \subseteq \mathsf{prem}(Y) \subseteq \mathsf{prem}(E)$, then $\Sigma \vdash_r \bot$ or β. Either case contradicts Σ is r-consistent.*

Suppose $Y \in \mathcal{A} \setminus E$. Hence $\exists \gamma \in \mathsf{prem}(Y)$, $\gamma \notin \Sigma$. We show $\exists X \in E, X \Rightarrow_{\{Y\}} Y$. By construction, $\Sigma = \Sigma_1 \cup \ldots \cup \Sigma_n$ such that for $i = 1 \ldots n$, $\Sigma_1 \cup \ldots \cup \Sigma_i$ is a maximal r-consistent subset of $\mathcal{B}_1, \ldots, \mathcal{B}_i$. Hence, suppose $\gamma \in \mathcal{B}_j$ for some $j = 1 \ldots n$. Then $\Sigma_1 \cup \ldots \cup \Sigma_j \cup \{\gamma\} \vdash_r \bot$. Hence $X = (\Delta, \{\gamma\}, \bot) \in \mathtt{Args}(\Sigma_1 \cup \ldots \cup \Sigma_j) \subseteq E$ s.t. $X \rightharpoonup_{\{Y\}} Y$. Since $\gamma \in \mathcal{B}_j$, and $\Delta \subseteq \bigcup_{k=1}^{j} \mathcal{B}_k$, $X \not\prec_{El} (\{\gamma\}, \emptyset, \gamma)$. Hence $X \Rightarrow_{\{Y\}} Y$.

***Proof of 2)**: Suppose for contradiction that $\Sigma = \bigcup_{X \in E} \mathtt{Prem}(X)$ is not r-consistent (i.e., $\Sigma \vdash_r \bot$). Then $Z = (\emptyset, \Sigma, \bot) \in \mathcal{A}$. By properties of $\prec_{El}$ (see Section 3.2) $\exists \alpha \in \Sigma$, $Z \not\prec (\{\alpha\}, \emptyset, \alpha)$. Hence $\exists B \in E$, $Z \Rightarrow_E B$ on α. No argument in E can defeat the unassailable Z, contradicting E is stable.*

Suppose for contradiction that Σ is not $\subseteq$-maximal r-consistent. Let $\Sigma_1, \ldots, \Sigma_n$ partition Σ s.t. for $i = 1 \ldots n$, Σ_i is a (possibly empty) subset of $\mathcal{B}_i$. Then, for some i, for $k = 1 \ldots i-1$, $\Sigma_1, \ldots, \Sigma_k$ is a $\subseteq$-maximal r-consistent subset of $\mathcal{B}_1, \ldots, \mathcal{B}_{i-1}$, and $\exists \alpha \in \mathcal{B}_i$ s.t.:

***i)** $\alpha \notin \Sigma_i$ **ii)** $\Sigma_1 \cup \ldots \cup \Sigma_{i-1} \cup \Sigma_i \cup \{\alpha\} \nvdash_r \bot$.*

Given i), $\exists Y = (\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$, $Y \notin E$. Since E is stable, $\exists X \in E$, $X \Rightarrow_{\{Y\}} Y$, hence $X \not\prec_{El} Y$. Consider two cases:

• *Suppose X concludes $\bot$. It cannot be that $\mathsf{supp}(X) = \emptyset$, since this would imply $\mathsf{prem}(X) \vdash \bot$, contradicting the r-consistency of Σ. Hence $X = (\Delta, \{\alpha\}, \bot)$.*

• *Suppose X concludes $\overline{\alpha}$, $\mathsf{prem}(X) = \Delta$, $\mathsf{supp}(X) = \emptyset$ or $\{\alpha\}$. By P3 and P2 (Def.7), $\exists X' = (\Delta, \{\alpha\}, \bot)$. Since X and X' have the same premises, and E is complete, then (by [11, Lemma 14]) $X' \in E$. Since $X \not\prec_{El} Y$ then $\forall \beta \in \Delta$, $\beta \not\prec \alpha$, and so $X' \not\prec_{El} Y$ and $X' \Rightarrow_{\{Y\}} Y$.*

*Given **ii)**, it must be that $\exists \beta \in \Delta$, s.t. $\beta \in E_j$, $j > i$. But then $X \prec_{El} Y$, respectively $X' \prec_{El} Y$, contradicting $X \not\prec_{El} Y$, respectively $X' \not\prec_{El} Y$.*

QED

As in Section 2, this result establishes a correspondence between the *PS* and argumentation consequence relations, where the latter are conclusions of *unconditional* arguments in stable extensions.

4 Properties and Postulates

4.1 Dung's Fundamental Lemma and Monotonicity of the Characteristic Function

We now study two key properties of AFs [13] as they apply to $pDCF$s. Firstly, the Fundamental Lemma (*FL*) states that:

> if X, X' are acceptable w.r.t. an admissible E, then $E \cup \{X\}$ is admissible and X' is acceptable w.r.t. $E \cup \{X\}$.

Secondly, an AF's characteristic function $\mathcal{F}$ is defined as:

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

Hence, the fixed points of $\mathcal{F}$ are an AF's complete extensions. Then $\mathcal{F}$ is shown to be monotonic: $E \subseteq E'$ implies $\mathcal{F}(E) \subseteq \mathcal{F}(E')$.

For $pDCF$s, the *FL* and monotonicity of a $pDCF$'s characteristic function cannot straightforwardly be shown, since proofs of these properties rely on the fact that attacks and defeats on any argument X is fixed and independent of the premises in a given set E. However, we can show similar properties for 'epistemically closed' sets E that enjoy the following property:

> if $W = (\Pi, \Sigma, \beta) \in E$, then for any $\Sigma' \subseteq \Sigma$ such that $\Sigma' \subseteq \mathsf{prem}(E)$, E also includes $W' = (\Pi \cup \Sigma', \Sigma \setminus \Sigma', \beta)$.

[9] Clearly arguments may not be subset minimal and yet use all the premises to derive a conclusion, e.g., two applications of modus ponens deriving q from p, $p \rightarrow q$, $p \rightarrow ((p \rightarrow q) \rightarrow q)$.

[10] For example, consider r provable from $(\Gamma = \{p, p \rightarrow r\}) \cup (\Delta = \{q\})$. Assuming natural deduction rules, one could by $\wedge_{\mathcal{I}}$ obtain $p \wedge q$, then by $\wedge_{\mathcal{E}}$, p, and then by $\rightarrow_{\mathcal{E}}$, r from p and $p \rightarrow r$. Clearly, such a proof, which redundantly makes use of q, implies sufficient resources for a proof of r from Γ (by a single application of $\rightarrow_{\mathcal{E}}$).

Epistemically closed sets are so named, as commitment to premises Σ' in E implies commitment to the logically equivalent W'.

Definition 9 Let $Cl_{ec}(E) = E \cup \{W' \mid W \in E, W' \in [W], \mathsf{prem}(W) \subseteq \mathsf{prem}(W'), \mathsf{prem}(W') \subseteq \mathsf{prem}(E)\}$.
Then E is *epistemically closed* (*ec*) if $E = Cl_{ec}(E)$.

Proofs of the following two propositions are shown in [11, Lemma 19] and [11, Lemma 23] respectively .

Proposition 5 Let X, X' be acceptable w.r.t. an admissible extension E of a $pDCF$ $(\mathcal{A}, \mathcal{C}, \prec)$. Then:

1. $Cl_{ec}(E \cup \{X\})$ is admissible.
2. X' is acceptable w.r.t. $Cl_{ec}(E \cup \{X\})$.

Proposition 6 Let E, E' be two *ec* admissible extensions of $(\mathcal{A}, \mathcal{C}, \prec)$ such that $E \subseteq E'$. Then $\mathcal{F}(E) \subseteq \mathcal{F}(E')$.

We sketch a key step in the proof of Proposition 5 that illustrates the importance of assuming epistemically closed sets.

Suppose X acceptable w.r.t. an admissible E where $Y \in E$. Inclusion of X in E may mean $Z \Rightarrow_{E \cup \{X\}} Y$, but $Z \not\Rightarrow_E Y$, since:

$$Z = (\Delta, \Gamma, \phi), \Phi \subseteq \Gamma \text{ and } \Phi \subseteq \mathsf{prem}(X), \Phi \nsubseteq \mathsf{prem}(E).$$

Since $Z \not\Rightarrow_E Y$, we cannot assume that the admissibility of E implies some Q in E (and hence $E \cup \{X\}$) defeating Z. Hence, we cannot immediately assume that Y is acceptable w.r.t. $E \cup \{X\}$, and so $E \cup \{X\}$ is admissible.

However we can show that there is an argument in $Cl_{ec}(E \cup \{X\})$ that defeats Z. Consider the following line of reasoning:

- $Z' \Rightarrow_E Y$ where $Z' = (\Delta \cup \Phi, \Gamma \setminus \Phi, \phi)$
- Hence $\exists W = (\Pi, \Sigma, \beta) \in E, W \Rightarrow_{\{Z'\}} Z'$
- Note: $\Sigma \subseteq \Delta \cup \Phi$ and $\Pi \cup \Phi \subseteq \mathsf{prem}(E \cup \{X\})$ **(1)**

Consider two cases:

a) Suppose W defeats Z' on $\alpha \in \Phi$.

We have the logically equivalent $W' = (\Sigma \cap \Delta, \Pi \cup (\Sigma \cap \Phi), \beta)$.

Given (1), $W' \Rightarrow_{E \cup \{X\}} X$. Since X is acceptable w.r.t. E, $\exists Q \in E$ s.t. $Q \Rightarrow_{\{W'\}} W'$. Since $\mathsf{prem}(W') \subseteq \mathsf{prem}(Z)$, then $Q \Rightarrow_{\{Z\}} Z$.

b) Suppose W defeats Z' on a premise in Δ. We have $W' = (\Pi \cup (\Sigma \cap \Phi), \Sigma \cap \Delta, \beta)$, $W' \Rightarrow_{\{Z\}} Z$. Given (1) *and the assumption that* $E \cup \{X\}$ *is epistemically closed*, then $W' \in E \cup \{X\}$.

Proposition 5 suffices to prove a key result implied by the *FL*:

Proposition 7 Every admissible extension of a $pDCF$ is a subset of a preferred extension.

See [11, Proposition 22] for proof of the above. Proposition 6, together with the fact that every fixed point of $\mathcal{F}$ is epistemically closed, facilitates proof of a key result following from the monotonicity of $\mathcal{F}$ (the proof of which is shown in [11, Proposition 25]):

Proposition 8 The characteristic function $\mathcal{F}$ of a $pDCF$ has a unique least fixed point (the grounded extension).

4.2 Rationality Postulates

We now show that the rationality postulates in [5] and [6] are satisfied, under some intuitive assumptions on preference relations.

Recall that in Example 1, no admissible extension contains the conflicting A_1 and F_1 since the unassailable I_2 must defeat either A_1, or F_1 on B_1, or F_1 on C_1. To suppose otherwise implies $I_2 \prec A_1, I_2 \prec B_1$, and $I_2 \prec C_1$. But such a preference relation would be incoherent as one would be rejecting the dialectical demonstration that A_1 and F_1 make use of mutually inconsistent premises, and effectively prefers arguments built from inconsistent premises. Indeed, in general, a strict preference $Y \prec X$, where $Y = (\Delta, \Gamma, \phi)$ attacks $X = (\{\alpha\}, \emptyset, \alpha)$, can be interpreted as:

> from amongst the inconsistent $\Delta \cup \Gamma \cup \{\alpha\}$, one preferentially accepts arguments constructed from α and rejects arguments constructed from $\Delta \cup \Gamma$.

Hence $I_1 \prec A_1, I_1 \prec B_1$, and $I_1 \prec C_1$ collectively indicate preferentially accepting arguments built from the inconsistent $\{a, b, \neg a \vee \neg b\}$ and rejecting arguments built from $\{a, b, \neg a \vee \neg b\}$. Contradiction.

Given the above interpretation of $Y \prec X$ it should follow that $Y' \prec X$, where $Y' = (\Delta, \Gamma \cup \{\alpha\}, \phi)$ ($\phi = \overline{\alpha}$ or $\phi = \bot$). Hence, if $F_1 \prec A_1$ then $(\{b, \neg a \vee \neg b\}, \{a\}, \neg a) \not\prec A_1$ would be incoherent. Similarly, if $(\{b, \neg a \vee \neg b\}, \{a\}, \neg a) \prec A_1$ then $F_1 \not\prec A_1$ would be incoherent. We therefore assume that preference relations satisfy the following properties:

Definition 10 Let $(\mathcal{A}, \mathcal{C}, \prec)$ be a $pDCF$. Then $\prec$ is dialectically coherent iff:

$\forall (\emptyset, \Delta, \bot) \in \mathcal{A}$: $\exists \alpha \in \Delta$ such that $(\emptyset, \Delta, \bot) \not\prec (\{\alpha\}, \emptyset, \alpha)$. (*Pref1*)

$\forall X = (\{\alpha\}, \emptyset, \alpha), Y = (\Delta, \Gamma, \phi), Y' = (\Delta, \Gamma \cup \{\alpha\}, \phi)$ ($\phi = \overline{\alpha}$ or $\phi = \bot$): $Y \prec X$ iff $Y' \prec X$. (*Pref2*)

We prove [5]'s closure and consistency rationality postulates for *pDCF*s, under the assumption that $\prec$ is dialectically coherent (note, one can straightforwardly show that the *Elitist* $\prec_{El}$ is dialectically coherent). [5] state theses postulates with reference to complete extensions. Also, recall (Section 3 and footnote 8), that the argumentation based consequences are the conclusions of *unconditional* arguments. Hence we define:

$$\mathsf{conc}(E) = \{\phi \mid (\Delta, \emptyset, \phi) \in E\}.$$

[5]'s postulates are stated with respect to a general framework for argumentation logics that integrate deductive and defeasible reasoning, so that arguments are trees whose links denote application of strict and defeasible inference rules, and sub-arguments correspond to subtrees. [18] extend the framework to accommodate *ClAr*, in which arguments are constructed from premises (the leaf nodes) that entail a conclusion (root node), via application of a single strict inference rule encoding the classical entailment. Hence, for *ClAr*, an argument X is a tree of depth 1, whose leaves are the 'elementary' arguments associated with the premises of X, and are X's sub-arguments. Therefore, we have the following formulation of the *sub-argument postulate* which [5] state as: if X is in a complete extension E, then all sub-arguments of X are in E:

Theorem 9 ***[Sub-argument Closure]*** *Let E be a complete extension of a $pDCF$ $(\mathcal{A}, \mathcal{C}, \prec)$, and $X \in E$. Then for all $\alpha \in \mathsf{prem}(X)$: $(\{\alpha\}, \emptyset, \alpha) \in E$.*

PROOF. *By P1 (Definition 7), $X' = (\{\alpha\}, \emptyset, \alpha) \in \mathcal{A}$. If $Y \Rightarrow_{E \cup \{X'\}} X'$, then $Y \Rightarrow_{E \cup \{X\}} X$ (on X'). Since E is complete, X is acceptable w.r.t. E and so $\exists Z \in E$ s.t. $Z \Rightarrow_{\{Y\}} Y$. Hence X' is acceptable w.r.t. E, and since E is complete, $X' \in E$.* QED

We now prove that *direct consistency* holds more generally for *admissible*, and not just complete, extensions.

Theorem 10 ***[Direct Consistency]*** *Let E be an admissible extension of a $pDCF$ $(\mathcal{A}, \mathcal{C}, \prec)$. Then $\forall \alpha, \beta \in \mathsf{conc}(E)$, $\alpha \neq \bot$ and $\alpha \neq \overline{\beta}$.*

PROOF. *Suppose for contradiction that E contains $X = (\Delta, \emptyset, \alpha)$ and $Y = (\Gamma, \emptyset, \beta)$, and 1) $\alpha = \bot$, or 2) $\alpha = \overline{\beta}$. Suppose 1) is the case. By P2, $Z = (\emptyset, \Delta, \bot) \in \mathcal{A}$. Suppose 2) is the case. By P3, $\exists Z' = (\Delta \cup \Gamma, \emptyset, \bot) \in \mathcal{A}$. By P2, $Z = (\emptyset, \Delta \cup \Gamma, \bot) \in \mathcal{A}$. In either case, $\mathsf{supp}(Z) \subseteq \mathsf{prem}(E)$. Hence $\forall \beta \in \mathsf{supp}(Z)$, $\exists W \in E$ s.t. $Z \rightarrow_E W$ on $(\{\beta\}, \emptyset, \beta)$. By Pref1 (Definition 10), at least one such attack succeeds as a defeat. Since the unassailable Z cannot itself be defeated by an argument in E, then this contradicts $W \in E$ is acceptable w.r.t. E.* QED

[5]'s *closure under strict rules* postulate states that if $\mathsf{conc}(E) \vdash \alpha$, then there is an argument X in E that concludes α. This postulate is stated for $pDCF$s, under the assumption that resources suffice to construct such an X. We refrain from mentioning [5]'s *indirect consistency* postulate as this immediately follows from direct consistency and closure under strict rules. Note however that (together with P2 and P3) the proofs of direct consistency and closure indicate that if resources suffice to recognise inconsistency in a set of premises, either through use of these premises in constructing an argument concluding $\bot$, or arguments with conflicting conclusions, then this suffices to ensure satisfaction of the rationality postulates.

Theorem 11 ***[Closure under Strict Rules]*** *Let E be a complete extension of a $pDCF$ $(\mathcal{A}, \mathcal{C}, \prec)$, $E' \subseteq E$, and $\mathsf{conc}(E') \vdash \alpha$. Suppose there exists a $X = (\Delta, \emptyset, \phi) \in \mathcal{A}$ such that $\Delta = \mathsf{prem}(E')$. Then $X \in E$.*

PROOF. *Suppose $Y \Rightarrow_{E \cup \{X\}} X$ on some $X' = (\{\alpha\}, \emptyset, \alpha)$. $\mathsf{prem}(X) \subseteq \mathsf{prem}(E)$ implies $\mathsf{supp}(Y) \subseteq \mathsf{prem}(E)$. Hence since $\alpha \in \mathsf{prem}(E)$, $\exists X'' \in E$ s.t. $Y \Rightarrow_{E \cup \{X''\}} X''$ on X'. Since E is complete $\exists Z \in E$ s.t. $Z \Rightarrow_{\{Y\}} Y$. Hence X is acceptable w.r.t. E. Since E is complete, $X \in E$.* QED

Finally, the contamination postulates – *non-interference* and *crash resistance* [6] – essentially state that the conclusions of arguments in complete extensions of an AF defined by $\mathcal{B}_1$ are preserved when unioning some $\mathcal{B}_2$ with $\mathcal{B}_1$, *such that the propositional atoms in $\mathcal{B}_2$ are disjoint from those in $\mathcal{B}_1$*. However, [6] does not account for the use of preferences. For *pDCF*s, we also need to refer to the preferences over arguments defined by the union of $\mathcal{B}_1$ and $\mathcal{B}_2$, such that the preference relations over arguments defined for each of $\mathcal{B}_1$ and $\mathcal{B}_2$ are preserved. In what follows, we define a composition operator for $pDCF$s, and assume the same resources are available for constructing arguments from $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$, so that for $\mathcal{A}$ defined by $\mathcal{B}$, $(\mathcal{A}_1 \cup \mathcal{A}_2) \subseteq \mathcal{A}$. Also, if resources suffice to construct a 'tautological' argument $X = (\emptyset, \emptyset, \alpha)$ from $\mathcal{B}_1$ ($\mathcal{B}_2$), then X can also be constructed from $\mathcal{B}_2$ ($\mathcal{B}_1$) and $\mathcal{B}$.

Definition 11 Let $(\mathcal{A}_1, \mathcal{C}_1, \prec_1)$ be defined by $\mathcal{B}_1$, $(\mathcal{A}_2, \mathcal{C}_2, \prec_2)$ defined by $\mathcal{B}_2$. Then $(\mathcal{A}, \mathcal{C}, \prec) = (\mathcal{A}_1, \mathcal{C}_1, \prec_1) \oplus (\mathcal{A}_2, \mathcal{C}_2, \prec_2)$, iff:

1. $\mathcal{A}_1 \cup \mathcal{A}_2 \subseteq \mathcal{A}$ (it is obvious to see that $(\mathcal{C}_1 \cup \mathcal{C}_2) \subseteq \mathcal{C}$).
2. $\forall X = (\emptyset, \emptyset, \alpha) : X \in \mathcal{A}_1$ iff $X \in \mathcal{A}_2$ iff $X \in \mathcal{A}$.
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 $(X_2, Y_2) \in \prec$

In Section 3.2 we informally described how the contaminating effect of inconsistent arguments is avoided in our approach. However, contamination may also arise as a result of dropping the subset minimality check on arguments.

Example 12 Let $\mathcal{B}_1 = \{p, \neg p\}$ and $\mathcal{B}_2 = \{s\}$, and :

- $\mathcal{A}_1 =$
  $\{X_1 = (\{p\}, \emptyset, p), X_1' = (\emptyset, \{p\}, p), Y_1 = (\{p\}, \{\neg p\}, \bot),$
  $X_2 = (\{\neg p\}, \emptyset, \neg p), X_2' = (\emptyset, \{\neg p\}, \neg p), Y_2 = (\{\neg p\}, \{p\}, \bot),$
  $Z = (\{\neg p, p\}, \emptyset, \bot), U = (\emptyset, \{\neg p, p\}, \bot)\}$.

Suppose also that $X_2 \prec_1 X_1$. Then $X_2 \not\Rightarrow_{E_1} X_1$, and $Y_2 \not\Rightarrow_{E_1} X_1$ (since by *Pref2* $Y_2 \prec_1 X_1$), and $Z \not\Rightarrow_{E_1} X_1$ ($Z \prec_1 X_1$ since $\prec_1$ is *imle*). $E_1 = \{X_1, X_1', X_2', Y_1, U\}$ is the single complete (grounded and preferred) extension.
$E_2 = \{X_2, X_2', Y_2, U\}$ is *not* admissible, since X_2 and Y_2 are both defeated by X_1 and Y_1, and neither defeats can be defended.

- $\mathcal{A}_2 = \{S = (\{s\}, \emptyset, s), S' = (\emptyset, \{s\}, s)\}$, and $\prec_2 = \emptyset$.
- $(\mathcal{A}, \mathcal{C}, \prec) = (\mathcal{A}_1, \mathcal{C}_1, \prec_1) \oplus (\mathcal{A}_2, \mathcal{C}_2, \prec_2)$, where $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \{C = (\{\neg p, s\}, \emptyset, \neg p), Z' = (\{\neg p, s, p\}, \emptyset, \bot)\}$ and their logically equivalent arguments, and $\prec = \prec_1$[11]. Now, we obtain two preferred extensions:

  E_1' that is a superset of E_1 and contains S and S';

  E_2' that is a superset of E_2 and contains S, S' and C.

We obtain the additional E_2' because X_2 and Y_2 are now defended by C, since $C \not\prec X_1$ and so $C \Rightarrow_{\{X_1\}} X_1$. Furthermore, the grounded extension E now contains S (recall that $U \in E$ will defend against Z's (and Z''s) defeat on S) *but not* P. Hence, contamination has taken place since adding the syntactically disjoint s has changed the credulously and sceptically defined consequences of $(\mathcal{A}_1, \mathcal{C}_1, \prec_1)$.

The problem here is that by adding premise s to X_2 to obtain C, X_2 has been strengthened, since (given $\prec = \prec_1$) $X_2 \prec X_1$ but $C \not\prec X_1$. However, the strengthening of X_2 is clearly counter-intuitive, since s is an irrelevant premise in C. Thus we would expect that $C \prec X_1$. Given this latter preference, we then obtain that E_1' is the single complete (grounded and preferred) extension of $(\mathcal{A}, \mathcal{C}, \prec)$.

Hence, preference relations must satisfy the following property in order to prevent contamination

Definition 12 Let $(\mathcal{A}, \mathcal{C}, \prec)$ be a $pDCF$. Then $\prec$ is *relevance coherent* iff $\forall X, Y, Y'$ such that

$Y = (\Gamma, \emptyset, \alpha)$, $Y' = (\Delta \cup \Gamma, \emptyset, \alpha)$, and $\Delta \| \Gamma \cup \{\alpha\}$ (Δ syntactically disjoint from $\Gamma \cup \{\alpha\}$): if $Y \prec X$ then $Y' \prec X$.

Of course, relevance coherence is trivially satisfied by proof theories that preclude construction of arguments with syntactically disjoint premises [10]. Note also that one can straightforwardly show that the *Elitist* $\prec_{El}$ is relevance coherent.

The following results assume $pDCF$s whose preference relations are dialectically and relevance coherent. In [6, 22], the *crash resistance* and *non-interference* postulates are formulated w.r.t. the 'consequences' of an *AF*. We analogously define the consequences of $pDCF$s, and state satisfaction of the postulates for the complete (and hence grounded, preferred and stable) semantics (recall that $\mathsf{conc}(E)$ denotes the conclusions of unconditional arguments).

[11] Note that all three $pDCF$s satisfy P1 – P4 in Definition 7.

Definition 13 Let $(\mathcal{A}, \mathcal{C}, \preceq)$ be a $pDCF$. Then $Cn((\mathcal{A}, \mathcal{C}, \preceq)) = \{\mathsf{conc}(E_1), \ldots, \mathsf{conc}(E_n)\}$ where $E_1, \ldots, E_n$ are the complete extensions of $(\mathcal{A}, \mathcal{C}, \preceq)$.

Non-interference states that the consequences of a $pDCF$ defined by a base $\mathcal{B}_1$, restricted to the atoms in $\mathcal{B}_1$, remain unchanged in the $pDCF$ defined by the union of $\mathcal{B}_1$ and a syntactically disjoint $\mathcal{B}_2$.

Theorem 13 *[Non Interference] Let* $\mathcal{B}_1 \| \mathcal{B}_2$, $(\mathcal{A}, \mathcal{C}, \preceq) = (\mathcal{A}_1, \mathcal{C}_1, \preceq_1) \oplus (\mathcal{A}_2, \mathcal{C}_2, \preceq_2)$. *Then:*

$$Cn((\mathcal{A}_1, \mathcal{C}_1, \preceq_1))_{|\mathsf{atoms}(\mathcal{B}_1)} = Cn((\mathcal{A}, \mathcal{C}, \preceq))_{|\mathsf{atoms}(\mathcal{B}_1)}{}^{12}.$$

PROOF. *See [11, Theorem 37].* QED

Referring to Example 12, $Cn((\mathcal{A}_1, \mathcal{C}_1, \preceq_1))_{|\mathsf{atoms}(\mathcal{B}_1)} = \{\{p\}\}$. If $\preceq$ is not relevance coherent then $Cn((\mathcal{A}, \mathcal{C}, \preceq))_{|\mathsf{atoms}(\mathcal{B}_1)} = \{\{p\}, \{\neg p\}\}$. However, assuming relevance coherence, then $C \preceq X_1, C \not\approx_{\{X_1\}} X_1, E_2'$ is not a complete extension of $(\mathcal{A}, \mathcal{C}, \preceq)$, and so $Cn((\mathcal{A}, \mathcal{C}, \preceq))_{|\mathsf{atoms}(\mathcal{B}_1)} = \{\{p\}\}$.

Definition 14 A base $\mathcal{B}_1$ is said to be contaminating iff there exists a $(\mathcal{A}_1, \mathcal{C}_1, \preceq_1)$ defined by $\mathcal{B}_1$, such that for any $\mathcal{B}_2$ and $(\mathcal{A}_2, \mathcal{C}_2, \preceq_2)$ defined by $\mathcal{B}_2$, where $\mathcal{B}_1 \| \mathcal{B}_2$: $Cn((\mathcal{A}_1, \mathcal{C}_1, \preceq_1)) = Cn((\mathcal{A}, \mathcal{C}, \preceq))$, where $(\mathcal{A}, \mathcal{C}, \preceq) = (\mathcal{A}_1, \mathcal{C}_1, \preceq_1) \oplus (\mathcal{A}_2, \mathcal{C}_2, \preceq_2)$.

Theorem 14 *[Crash Resistance] There does not exist a contaminating base* $\mathcal{B}$.

PROOF. *See [11, Theorem 39].* QED

Referring to Example 12, $Cn((\mathcal{A}_1, \mathcal{C}_1, \preceq_1)) = \{\{p\}\} \neq Cn((\mathcal{A}, \mathcal{C}, \preceq)) = \{\{p, s\}\}$.

5 Conclusions

This paper has argued that features of propositional classical logic instantiations of AFs (*ClAr*) that suffice to ensure satisfaction of rationality postulates, preclude uses of argument characteristic of real-world dialectical reasoning by resource bounded agents. Our solution has been to provide an account of *ClAr* in which the ontology of classical logic arguments explicitly distinguishes between an argument's premises assumed true, and those supposed true for the sake of argument. In so doing, we obviate the need for checking consistency and subset minimality of premises, and identify an intuitive set of assumptions on the available resources for constructing arguments for inclusion in a framework, and show that the resulting formalism satisfies the closure, consistency, non-interference and crash resistance postulates. We thus provide a rational account of *ClAr* that is suitable for use by resource bounded agents. Our account also avoids the foreign commitment problem, and formalises the real-world use of argument in dialectically demonstrating that an agent's premises are inconsistent. We have shown that key properties of Dung's theory are preserved, and we provide an argumentative characterisation of the Preferred Subtheories non-monotonic logic, under the assumption that agents have limited inferential capabilities.

[14] also identify requirements for practical applications of argumentation. They stipulate that the computational cost of validating the legitimacy of a constructed argument should be at most polynomial (in the size of the arguments), and whether an argument attacks another should be at most linear (in the size of the argument's conclusion). Both are satisfied by our approach (the former trivially since we drop checks on premises). [14] also argue that an argument's premises should be relevant to its conclusion. Although we drop the subset minimality check, we suggest that the issue of relevance should be addressed by the specific proof theoretic means for constructing arguments.

Pragmatic considerations also motivate dropping consistency and subset minimality checks on arguments in [3]. 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 [5] are not studied and neither is there consideration of argumentation under resource bounds. Finally, [7] also distinguish between premises and suppositions, but in a restricted logical setting (arguments are constructed from literals and defeasible rules). This work studies only the grounded semantics, does not consider preferences or investigate satisfaction of the rationality postulates.

A number of works show satisfaction of the non-interference and crash resistance postulates for argumentation formalisms that integrate deductive and defeasible reasoning. [6] show that logic programming and Default Logic instantiations of Dung frameworks satisfy these postulates under the *semi-stable* semantics. In [22], arguments are built from a set of *classically consistent* propositional formulae, and defeasible and strict inference rules. [22] do not consider the use of preferences, and show satisfaction of the postulates under the assumption that inconsistent arguments (identified as those whose contained premises together with the conclusions of defeasible rules are classically inconsistent) are excluded from the argumentation framework. Finally, [12] define a version of the *ASPIC*$^+$ framework [18] in which the strict inference rules encode inference in [20]'s paraconsistent logic. The focus of [12] is on showing satisfaction of the closure and consistency postulates, and satisfaction of non-interference and crash resistance is not formally shown (the authors state that satisfaction of these postulates can be taken for granted given the absence of the *Ex Falso* principle). Finally, we have identified that contamination may result if one does not implement the subset minimality check on an argument's premises. While we drop the subset minimality check, we identify a notion of relevance that is more readily addressed by classical proof theories. For proof theories that do not exclude construction of arguments making use of irrelevant premises, we show that contamination is avoided if preference relations do not strengthen arguments upon addition of irrelevant premises (one such preference relation being the widely used Elitist preference). This result is closely related to a result shown in [18], which states that the argumentation defined consequences of a framework remain unchanged if one additionally includes non-subset minimal arguments, provided that they are not stronger than their subset minimal counterparts.

With regard to future research directions, we recognise that while we accommodate agents who resources are bounded with respect to the *construction* of arguments, we need to investigate the complexity of computing semantics (i.e, *evaluation* of arguments) given our dialectical definition of attacks (defeats) on arguments. Finally, we are currently extending our dialectical formulation of arguments and acceptability to the *ASPIC*$^+$ framework [18]. *ASPIC*$^+$ is a general framework for structured argumentation that accommodates arguments built from strict inference rules that encode the inference relations of deductive logics, as well as defeasible inference rules. We thus aim to provide a general account of structured argumentation for use by real-world resource bounded agents.

Acknowledgements: We thank the reviewers whose comments helped improve this paper.

[12] Recall Notation 3

REFERENCES

[1] L. Amgoud and C. Cayrol, 'A reasoning model based on the production of acceptable arguments', *Annals of Mathematics and Artificial Intelligence*, **34**(1-3), 197–215, (2002).

[2] L. Amgoud and S. Vesic, 'Handling inconsistency with preference-based argumentation', in *Scalable Uncertainty Management: 4th International Conference, SUM 2010*, pp. 56–69. Springer, (2010).

[3] O. Arieli and C. Straßer, 'Sequent-based logical argumentation', *Argument and Computation*, **6**(1), 73–99, (2015).

[4] G. Brewka, 'Preferred subtheories: An extended logical framework for default reasoning', in *International Joint Conference on Artificial Intelligence*, pp. 1043–1048, (1989).

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

[6] M. Caminada, W. Carnielli, and P. Dunne, 'Semi-stable semantics', *Logic and Computation*, **22**(5), 1207–1254, (2012).

[7] Martin Caminada, 'Dialogues and HY-arguments', in *Non-Monotonic Reasoning*, pp. 94–99, (2004).

[8] V. Chvátal and E. Szemerédi, 'Many hard examples for resolution', *Journal of the ACM*.

[9] M. D'Agostino, 'An informational view of classical logic', *Theoretical Computer Science*, **606**, 79–97, (2015).

[10] M. D'Agostino, D.Gabbay, and S. Modgil, 'Normal proofs and non-contamination in classical natural deduction', *Technical Report*, `https://dl.dropboxusercontent.com/u/5626429/CND_TR.pdf`, (2016).

[11] M. D'Agostino and S. Modgil, 'Classical logic, argumentation and dialectic: Technical report', *Technical Report*, www.dcs.kcl.ac.uk/staff/smodgil/ECAITechnicalReport.pdf, (2016).

[12] D.Grooters and H.Prakken, 'Combining paraconsistent logic with argumentation', in *Computational Models of Argument. Proceedings of COMMA 2014*, pp. 301–312. IOS Press, (2014).

[13] P. M. 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).

[14] P.M. Dung, F. Toni, and P. Mancarella, 'Some design guidelines for practical argumentation systems', in *Proc. Conference on Computational Models of Argument: COMMA 2010*, pp. 183–194, (2010).

[15] N. Gorogiannis and A. Hunter, 'Instantiating abstract argumentation with classical logic arguments: Postulates and properties', *Artificial Intelligence*, **175**(910), 1479 – 1497, (2011).

[16] M.Caminada, S. Modgil, and N.Oren, 'Preferences and unrestricted rebut', in *Computational Models of Argument: Proceedings of COMMA 2014*, pp. 209–220. IOS Press, (2014).

[17] P. McBurney and S. Parsons, 'Chapter 13: Dialogue games for agent argumentation', in *Argumentation in AI*, 261–280, Springer, (2009).

[18] S. Modgil and H. Prakken, 'A general account of argumentation and preferences', *Artificial Intelligence*, **195(0)**, 361 – 397, (2013).

[19] S. Modgil, F. Toni, F. Bex, I. Bratko, C. Chesñevar, W. Dvořák, M.A. Falappa, X. Fan, S. Gaggl, A.J. García, M.P. González, T. Gordon, J. Leite, M. Možina, C. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran, 'Chapter 21: The added value of argumentation', in *Agreement Technologies*, ed., S. Ossowski, 357–403, Springer, (2013).

[20] N. Rescher and R. Manor, 'On inference from inconsistent premises', *Journal of Theory and Decision*, **1**, 179–219, (1970).

[21] G. Vlastos, 'The socratic elenchus', *The Journal of Philosophy*, **79**(11), 711–714, (1982).

[22] Y. Wu and M. Podlaszewski, 'Implementing crash-resistance and non-interference in logic-based argumentation', *Journal of Logic and Computation*, **25**, 303–333, (2015).

[23] A.P. Young, S. Modgil, and O. Rodrigues, 'Prioritised default logic as rational argumentations', in *To appear in: Proc. International Joint Conference on Autonomous Agents and Multi-Agents Systems (AAMAS'2016)*, (2016).