

A Fully Rational Account of Structured Argumentation Under Resource Bounds

Marcello D’Agostino¹ and Sanjay Modgil²

¹University of Milan

²Kings’s College London

marcello.dagostino@unimi.it, sanjay.modgil@kcl.ac.uk

Abstract

ASPIC⁺ is an established general framework for argumentation and non-monotonic reasoning. However *ASPIC*⁺ does not satisfy the non-contamination rationality postulates, and moreover, tacitly assumes unbounded resources when demonstrating satisfaction of the consistency postulates. In this paper we present a new version of *ASPIC*⁺ – *Dialectical ASPIC*⁺ – that is *fully rational under resource bounds*.

1 Introduction

Context. Dung’s seminal argumentation theory [Dung, 1995] enables formalisation of single agent, and distributed (dialogical), non-monotonic reasoning in terms of the exchange of arguments. *ASPIC*⁺ is a widely used general framework for defining the arguments and defeats that comprise a Dung Argumentation Framework (*AF*) [Modgil and Prakken, 2013]. Arguments are constructed from a knowledge base (*KB*) of formulae in some logical language, and defeasible and strict inference rules; the latter typically encoding inference in a deductive logic. Attacks are defined between arguments, and a preference relation over arguments decides which attacks succeed as defeats. The claims of the winning (justified) arguments (under Dung’s semantics) define the non-monotonic inferences from the underlying *KB*. *ASPIC*⁺ allows considerable degrees of freedom in the choice of the logical language, the defeasible and strict inference rules, and how preferences are defined. Hence it is shown (see review in [Modgil and Prakken, 2018]) to capture other argumentation formalisms, provide dialectical characterisations of well known non-monotonic logics (e.g., (*Prioritised*) *Default Logic* and *Preferred Subtheories*), and enable formulation of novel non-monotonic logics (e.g., [Dong *et al.*, 2019]).

ASPIC⁺ specifies guidelines that constrain the above choices, so guaranteeing rational outcomes when evaluating justified arguments; in particular *consistency* and *closure* [Caminada and Amgoud, 2007]. However, two key limitations of *ASPIC*⁺ preclude its use in providing dialectical characterisations of non-monotonic reasoning for real world agents reasoning as individuals or via dialogue. Firstly, consistency is satisfied under the tacit assumption that agents have unbounded resources (logical omniscience is assumed

in that *all* deductive inferences are assumed to be generated). Secondly, guidelines for satisfying the ‘non-contamination’ postulates [Caminada *et al.*, 2012] are not specified. Hence, inferences (the claims of justified arguments) may be invalidated by adding syntactically disjoint information to the premises. This is particularly problematic when evaluating arguments under the computationally feasible *grounded* semantics which (sceptically) identifies a unique set of non-monotonic inferences.

Contributions. This paper is the first to provide a formalisation of *ASPIC*⁺ – *Dialectical ASPIC*⁺ (*D-ASPIC*⁺) – that is *fully rational under resource bounds*: consistency, closure and non-contamination are satisfied, while *making only minimal (i.e., relatively undemanding) assumptions as to the resources available for constructing arguments*. Hence, *D-ASPIC*⁺ represents an important step towards enabling deployment of argumentation formalisms and non-monotonic logics in real-world settings by resource bounded agents, while satisfying the full gamut of rationality criteria.

Outline of Paper. In Section 3 we formalise the *D-ASPIC*⁺ framework for structured argumentation. We adapt the approach in [D’Agostino and Modgil, 2018], which establishes rationality under resource bounds for classical logic argumentation. *D-ASPIC*⁺ refines an argument’s ontology to reflect the distinction in real-world dialectical uses of argument, between arguments that an agent commits to, and those that are supposed (conceded) for the sake of argument (and that an interlocutor may commit to in dialogical contexts). Unlike *ASPIC*⁺, consistency and closure hold without assuming full deductive inferential capabilities, or assuming that preference relations satisfy any properties. Moreover, *D-ASPIC*⁺ solves the *foreign commitment problem* [Caminada *et al.*, 2014]. Section 4 then identifies the conditions – properties of preferences relations and assumptions on available resources for constructing arguments – under which *D-ASPIC*⁺ satisfies non-contamination. Section 5 highlights other works aiming at full rationality for restricted versions of *ASPIC*⁺. We also point to future work: we propose that strict inferences in arguments be generated by proof theories which do not licence use of ‘redundant’ (syntactically disjoint) premises (e.g., the natural deduction system in [D’Agostino *et al.*, 2019]). In so doing, non-contamination is then satisfied irrespective of the properties of preference relations.

2 ASPIC⁺ and its Limitations.

An **ASPIC⁺ argumentation system** $(AS, -, \mathcal{R}, n)$ consists of some logical language \mathcal{L} that includes the constant symbol \perp (falsum), and a function $- : \mathcal{L} \mapsto 2^{\mathcal{L}}$ specifying when formulae are in conflict:

- φ is a **contrary** of ψ if $\varphi \in \bar{\psi}$, $\psi \notin \bar{\varphi}$ (e.g., $\forall \phi \in \mathcal{L}$, φ is a contrary of $NOT \varphi$, where NOT is ‘negation as failure’);
 - φ is a **contradictory** of ψ (denoted by ‘ $\varphi = -\psi$ ’), if $\varphi \in \bar{\psi}$, $\psi \in \bar{\varphi}$ (e.g., $\forall \varphi \in \mathcal{L}$, φ and $\neg\varphi$ are contradictories).
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict (\mathcal{R}_s) and defeasible (\mathcal{R}_d) inference rules of the form $\varphi_1, \dots, \varphi_n \rightarrow \varphi$, respectively $\varphi_1, \dots, \varphi_n \rightrightarrows \varphi$ ($\varphi \neq \perp$) where φ_i, φ are meta-variables ranging over wff in \mathcal{L} . Typically, strict rules encode inference in some deductive logic, e.g.,

$$\begin{aligned} \varphi_1, \dots, \varphi_n \rightarrow \varphi \in \mathcal{R}_s &\text{ iff } \varphi_1, \dots, \varphi_n \vdash_{CL} \varphi, \\ &\text{ in which case we may simply write } \mathcal{R}_s = \vdash_{CL} \\ (\vdash_{CL} &\text{ is the consequence relation of classical logic).} \end{aligned}$$

Defeasible rules typically specify domain-specific defeasible (default) inferences (e.g., $bird(X) \rightrightarrows fly(X)$). n is a partial function assigning names (formulae in \mathcal{L}) to defeasible inference rules. Then, any argument concluding a contrary or contradictory of α , invalidates use of the defeasible rule r named α (and so attacks an argument X that uses r).

An **ASPIC⁺ theory** $\mathcal{T} = (AS, \mathcal{K})$ consists of an argumentation system and a knowledge base $\mathcal{K} \subseteq \mathcal{L} \setminus \{\perp\}$ consisting of two disjoint subsets \mathcal{K}_n (the **axiom premises**) and \mathcal{K}_p (the **ordinary premises**). \mathcal{K}_n represents infallible information and/or axioms in some deductive logic, and \mathcal{K}_p fallible information. In this paper we assume any \mathcal{T} satisfies only two of the five conditions identifying ‘well defined’ theories in [Modgil and Prakken, 2013]: 1) the axiom premises are consistent, and 2) whenever φ is a contrary of ψ then $\psi \notin \mathcal{K}_n$ and ψ is not the consequent of a strict rule (e.g., $NOT\varphi$ cannot be an axiom or conclusion of a strict rule). We additionally assume that the following is satisfied by \mathcal{R}_s :

$$\text{if } \varphi = -\psi \text{ then } \varphi, \psi \rightarrow \perp \in \mathcal{R}_s$$

ASPIC⁺ arguments are upside down trees, whose leaves are premises, yielding the argument’s claim (the root node) via application of strict and/or defeasible rules. In this paper, the defeasible, but not strict, rules are explicitly represented as nodes. Figures 1i–iii show such arguments. (In each case $\mathcal{R}_s = \vdash_{CL}$ and $\mathcal{K}_n = \emptyset$.) Application of defeasible rules is indicated by dashed lines (note that ‘ \supset ’ denotes material implication). Application of strict rules is indicated by unbroken lines linking the rule’s conclusion to its antecedents. In i), we assume arguments built from $\mathcal{K}_p = \{k, r, \neg p, q \supset \neg s, t\}$ and $\mathcal{R}_d = \{k, r \rightrightarrows p; t \rightrightarrows s\}$. In ii) and iii), $\mathcal{K}_p = \{e, s\}$, $\mathcal{R}_d = \{e \rightrightarrows \neg s\}$ and $\mathcal{K}_p = \{g, e, s\}$, $\mathcal{R}_d = \{g \rightrightarrows p; e \rightrightarrows \neg s\}$ respectively. We may write arguments as lists of premises and defeasible rules preceding a claim (e.g., $A = [k; r; k, r \rightrightarrows p; \neg p; q \supset \neg s] : \neg s$ in Fig. 1).

The **sub-arguments** of an argument X (denoted $Sub(X)$) are the sub-trees of X , each of whose root node is a node in X other than a defeasible rule. In Fig. 1i, $Sub(A) = A, A_{1..6}$, and the **proper** sub-arguments of A are $A_{1..6}$. Note that we may also represent an argument by its top rule (the last rule applied to obtain the argument’s claim), and with its im-

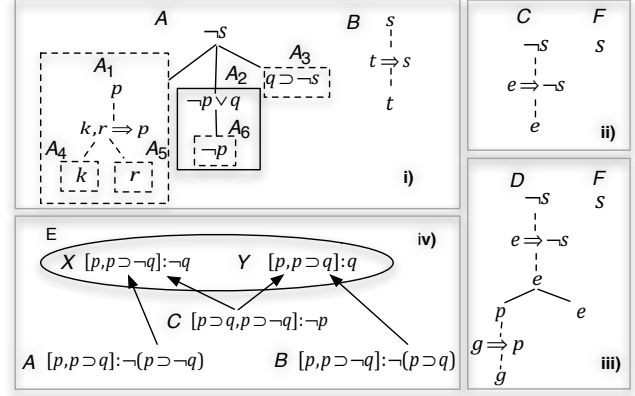


Figure 1: ASPIC⁺ arguments.

mediate proper sub-arguments substituting for the rule’s antecedents; e.g., $A = A_1, A_2, A_3 \rightarrow \neg s$. Also, any $X' \in Sub(X)$ such that X' ’s top rule is defeasible or X' is an ordinary premise, is said to be a **fallible sub-argument** of X .

An argument X **attacks** Y on $Y' \in Sub(Y)$, if the claim α of X is a contrary or contradictory of: i) an ordinary premise Y' (an **undermine attack**), or ii) conclusion of a defeasible rule that is the top rule of Y' (a **rebut attack**), or iii) the name $(n(r))$ of the defeasible rule r that is the top rule of Y' (an **undercut attack**). If X undermine or rebut attacks Y on Y' , then in the case that the claim α is a contradictory of its target, X **defeats** Y on Y' only if $X \not\prec Y'$, where \prec is a strict partial preference ordering over arguments. Else undercut attacks and attacks targeting contraries, succeed as defeats irrespective of preferences (e.g., when X claims α , the contrary of an ordinary premise $Y' = [NOT\alpha] : \alpha \in Sub(Y)$).

One can then define the extensions of a Dung AF $(\mathcal{A}, \rightsquigarrow)$ consisting of an ASPIC⁺ theory’s arguments \mathcal{A} and defeats $(\rightsquigarrow \subseteq \mathcal{A} \times \mathcal{A})$. A set S of arguments is **admissible** if it is conflict free ($\forall X, Y \in S: X \not\prec Y$) and defends all its arguments ($\forall X \in S, \forall Y$ s.t. $Y \rightsquigarrow X, \exists Z \in S$ s.t. $Z \rightsquigarrow Y$). An admissible set S is: **complete** if every argument defended by S is in S ; **grounded/preferred** if it is the *smallest*, respectively *a largest* (under set inclusion), complete extension; **stable** if $\forall X \notin S, \exists Y \in S$ s.t. $Y \rightsquigarrow X$.

Limitations of ASPIC⁺. The **consistency** postulate [Camina and Amgoud, 2007] states that no complete extension should contain arguments with claims φ and ψ such that $\varphi \in \bar{\psi}$. ASPIC⁺ is shown to satisfy consistency ([Modgil and Prakken, 2013]) if \prec is ‘reasonable’ (i.e., \prec satisfies certain properties as illustrated below), and under the tacit assumption that arguments are constructed given unlimited resources for generating all deductive inferences encoded in the strict rules. For example, if $\mathcal{R}_s = \vdash_{CL}$, then in general there will be an infinite number of arguments. Moreover first order classical logic is undecidable; even in the propositional case deciding whether $\Delta \vdash_{CL} \alpha$ is in general NP-hard, and therefore most likely intractable. This is clearly unfeasible for agents with limited resources. To illustrate, suppose $\mathcal{R}_s = \vdash_{CL}$ and

$\mathcal{K}_p = \{p, p \supset \neg q, p \supset q\}$. Note that attacks cannot target conclusions of strict rules (else this would result in violation of *strict closure* and *indirect consistency* rationality postulates [Caminada and Amgoud, 2007]). Hence, to ensure that no complete extension E contains X and Y in Fig. 1iv, one needs to assume resources suffice to additionally construct A and B and C . Then, if \prec is reasonable (i.e., $A \not\prec [p \supset \neg q] : p \supset \neg q$ or $B \not\prec [p \supset q] : p \supset q$ or $C \not\prec [p] : p$), at least one of the attacks from A , B or C succeeds as a defeat. Hence, defending against any such defeat requires some $Z \in E$ that defeats A or B or C , which necessarily implies that Z defeats either X or Y , contradicting E is conflict free and so complete.

Non-contamination [Caminada *et al.*, 2012] may be violated by $ASPIC^+$ if an argument makes use of inconsistent premises and/or rules, or premises/rules that are redundantly used in deriving a claim.

Firstly, suppose $\mathcal{R}_s = \vdash_{CL}, \mathcal{K}_p = \{q \supset \neg s, t\}$ and $\mathcal{R}_d = \{t \Rightarrow s\}$. Clearly B in Fig. 1i is in the grounded extension. But upon adding $\{k, r, \neg p\}$ and $\{k, r \Rightarrow p\}$ to \mathcal{K}_p and \mathcal{R}_d , which intuitively are irrelevant to deciding the status of B (by virtue of being syntactically disjoint from t and $t \Rightarrow s$), then A in (Fig. 1i) attacks B . Assuming $\prec = \emptyset$, A defeats B . One can easily verify that B is now no longer in the grounded extension. The ‘inconsistent’ A is said to be ‘contaminating’.

Secondly, suppose in Fig. 1iii that $C \prec F$ so that F defeats C but not vice versa. $F = [s] : s$ is in the grounded extension. Suppose one adds to the underlying theory the irrelevant (syntactically disjoint) g and $g \Rightarrow p$, additionally yielding D in Fig. 1iii. If $D \not\prec F$, then although F still defeats C , D and F defeat each other. F is no longer in the grounded extension. C has been contaminated by the redundant g and $g \Rightarrow p$, by virtue of the classical inference $p, e \vdash_{CL} e$.

3 Dialectical $ASPIC^+$

In providing a resource bounded fully rational account of $ASPIC^+$, we begin by observing that a common dialectical move is to commit to some arguments and their claims, arguing that if one *supposes for the sake of argument* the arguments and claims of an interlocutor, then α is necessarily (strictly) entailed. Hence, a $D-ASPIC^+$ argument differentially labels the ‘maximal fallible subarguments’ (*mfs*) of an $ASPIC^+$ argument A , according to whether they are committed or supposed for the sake of argument:

Definition 1. An $ASPIC^+$ argument A is a **D- $ASPIC^+$ argument** iff exactly one of L1...L4 hold:

L1: $A = [\phi] : \phi, \phi \in \mathcal{K}_p$ and A is exclusively labelled with either \odot (for ‘committed’) or \odot (for ‘supposed’)

L2: $A = [\phi] : \phi, \phi \in \mathcal{K}_n$

L3 $A = A_1, \dots, A_n \Rightarrow \varphi$ and A is exclusively labelled with either \odot or \odot

L4: $A = A_1, \dots, A_n \rightarrow \varphi$ and each A_i is a D- $ASPIC^+$ argument,¹ and:

C1 $\forall A' \in \text{Sub}(A), A'$ is not labelled with \odot and \odot

¹That is, traversing each path from the root (conclusion) to a leaf of an $ASPIC^+$ argument A , assign label \odot or \odot when first encountering either an ordinary premise or conclusion of a defeasible inference rule. Once a label is assigned, traversal is terminated.

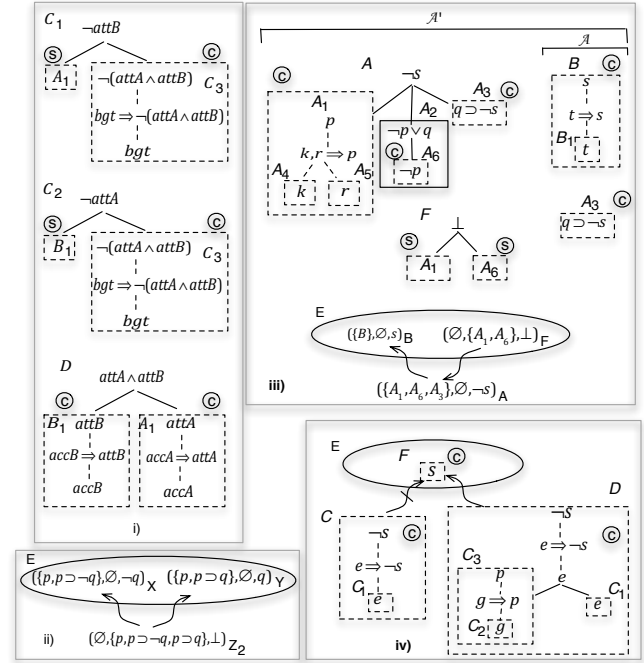


Figure 2: $D-ASPIC^+$ arguments.

C2 $\forall A' \in \text{Sub}(A)$, if A' is labelled \odot , then $\neg \exists A'' \in \text{Sub}(A)$ such that A'' is labelled \odot and $A' \in \text{Sub}(A'')$.

We may write $(\Delta, \Gamma, \phi)_A$ to denote a $D-ASPIC^+$ argument A , where $A' \in \Delta$ iff A' is labelled \odot , $A' \in \Gamma$ iff A' is labelled \odot in A . Then $\text{com}(A) = \Delta$ and $\text{sup}(A) = \Gamma$ are the committed, respectively supposed, *mfs* of A . C2 then states that no supposed $X \in \Gamma$ is a proper sub-argument of some committed $Y \in \Delta$, since commitment to Y implies commitment to all sub-arguments of Y . Also, for any set S :

$$\text{SubCom}(S) = \bigcup_{(\Delta, \Gamma, \phi)_X \in S} \text{Sub}(\Delta)$$

denotes the sub-arguments of committed arguments in S .

Example 1. Let $\mathcal{R}_s = \vdash_{CL}, \mathcal{K}_n = \{accA, accB\}, \mathcal{K}_p = \{bgt\}, \mathcal{R}_d = \{accA \Rightarrow attA; accB \Rightarrow attB; bgt \Rightarrow \neg(attA \wedge attB)\}$, where (letting X stand for A or B):

$accX$ and $attX$ respectively denote ‘paper accepted at conference X ’ and ‘will attend conference X ’, and bgt denotes ‘budget’.

$(\{A_1, B_1\}, \emptyset, attA \wedge attB)_D$ in Fig. 2-i claims the intention to attend A and B . $(\{C_3\}, \{A_1\}, \neg attB)_{C_1}$ and $(\{C_3\}, \{B_1\}, \neg attA)_{C_2}$, argue that the budget is insufficient to attend both conferences, and supposing attendance at A (respectively B), then one cannot attend B (respectively A).

Definition 2. Let $(\Delta, \Gamma, \alpha)_X$ and $(\Pi, \Sigma, \beta)_Y$ be $D-ASPIC^+$ arguments.

1) if $\alpha \neq \perp$, then X **attacks** Y if X undercuts, rebuts or undermines some $Y' \in \Pi$ on $Y'' \in \text{Sub}(Y')$ (one can only target committed arguments). X **defeats** Y iff X attacks Y and if the attack rebuts/undermines a contradictory then $X \not\prec Y''$.

2) if $\alpha = \perp$ (X is a ‘falsum argument’) then X **attacks** Y on any $Y' \in \Gamma \cap \text{Sub}(\Pi)$. If $\Delta = \emptyset$, then X **defeats** Y . If $\Delta \neq \emptyset$ then X **defeats** Y only if $\exists Y' \in \Gamma \cap \text{Sub}(\Pi)$, $X \not\prec Y'$.

Intuitively, attacks and defeats of type 1 are defined as for ASPIC^+ arguments, except that the attack must target a fallible committed sub-argument of the attackee. In Example 1, C_1 attacks D on B_1 , and defeats D on B_1 only if $C_1 \not\prec B_1$. Likewise, C_2 attacks D on A_1 , and defeats D on A_1 only if $C_2 \not\prec A_1$. Notice that C_1 (C_2) does not attack C_2 (C_1) on the supposed B_1 (A_1). For attacks from falsum arguments, if $\Delta = \emptyset$ then the attack amounts to a dialectical demonstration that the supposed Γ are inconsistent, and so succeeds as a defeat independently of preferences; else success is contingent on $X \not\prec Y'$ for some $Y' \in \Gamma \cap \text{Sub}(\Pi)$. Also, note that the *mfs* of any given constructed argument can be differentially labelled; this does not affect the strength of the argument:

Definition 3. Let \prec be a strict partial ordering over some set \mathcal{A} of D-ASPIC^+ arguments. Let $(\Delta, \Gamma, \alpha)_X \in \mathcal{A}$. Then:

- $[X] = \{(\Delta', \Gamma', \alpha)_{X'} \mid \Delta' \cup \Gamma' = \Delta \cup \Gamma, \text{ and } \text{Sub}(X') = \text{Sub}(X)\}$. $\forall Y, Z \in [X]$ we say that Y and Z are **logically equivalent**.
- Henceforth we assume that \prec satisfies the following: $\forall Y' \in [Y], \forall X' \in [X] : Y \prec X \text{ iff } Y' \prec X'$

In general, when establishing whether Y is defended by a set E of arguments (i.e., whether Y is *acceptable w.r.t. E*), one commits to fallible arguments in E and Y . Hence $(\Delta, \Gamma, \alpha)_X$ can then challenge (i.e., defeat) Y if X supposes/concedes a subset Γ of these fallible arguments, arguing that together with Δ , Γ necessarily entails a claim that conflicts with some fallible element in Y . Then, a defeat by $(\Delta, \Gamma, \perp)_X$ indicates that the fallible arguments Γ committed in $E \cup \{Y\}$, together with Δ , are inconsistent. Y should only then be targeted if at least one of the committed arguments in Y contributes to the inconsistency (i.e., is in Γ). Only if $\Delta = \emptyset$ does the attack by X dialectically demonstrate that $E \cup \{Y\}$ has necessarily committed to (strictly entails) an inconsistency, and so succeeds as a defeat independently of preferences. As for arguments (X) deployed in challenging Y , these are not required to define an admissible set; they are countered on an individual basis. Hence $Z \in E$ can defend Y if Z defeats X , while conceding only fallible arguments that X commits to (i.e., sub-arguments of Δ).

Definition 4. Let $(\mathcal{A}, \mathcal{D})$ be a dialectical AF (DAF) defined by a theory $\mathcal{T} = (\mathcal{AS}, \mathcal{K})$, where \mathcal{A} is a set of D-ASPIC^+ arguments, and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ is the set of defeats.

Then X **defeats** Y with respect to $S \subseteq \mathcal{A}$, denoted $X \rightsquigarrow_S Y$, if $(X, Y) \in \mathcal{D}$ and $\text{sup}(X) \subseteq \text{SubCom}(S \cup \{Y\})$.

- S is **conflict free** if $\forall Z, Y \in S, Z \not\rightsquigarrow_S Y$.
- Y is **acceptable w.r.t. S** if $\forall X$ s.t. $X \rightsquigarrow_S Y, \exists Z \in S$ s.t. $Z \Rightarrow_{\{X\}} X$.

The extensions of $(\mathcal{A}, \mathcal{D})$ are then defined as in Section 2, except in the case that S is stable, in which case we stipulate that $\forall X \notin S. \exists Y \in S$ s.t. $Y \rightsquigarrow_{\{X\}} X$.

Consider Example 1, and suppose a dialogue in which the acceptability of D w.r.t. some set E , is challenged by a defeat from either C_1 or C_2 . Then $C_1 \rightsquigarrow_E D$ and $C_2 \rightsquigarrow_E D$, where

C_1 (C_2) need not commit to A_1 (B_1), but rather supposes A_1 (B_1) only for the sake of argument, given A_1 (B_1) is committed in D . The corresponding ASPIC^+ arguments, which do not distinguish commitments from suppositions, illustrate the foreign commitment problem [Caminada *et al.*, 2014]: they unnaturally have to commit to attendance at one of the conferences (i.e., A_1 or B_1) in order to challenge the argument D which claims that both conferences cannot be attended.

Let $\mathcal{R}_s = \vdash_{CL}, \mathcal{K}_p = \{p, p \supset \neg q, p \supset q\}$ and \mathcal{A} include the D-ASPIC^+ arguments in Fig.2ii, as well as $(\{p\}, \emptyset, p)_F, (\{p \supset \neg q\}, \emptyset, p \supset \neg q)_G, (\{p \supset q\}, \emptyset, p \supset q)_H$. Now, E cannot be admissible. This is because the inconsistency of the premises $p, p \supset \neg q, p \supset q$ is recognised by virtue of constructing arguments with conflicting claims (X and Y). Then, intuitively, a minimal assumption on available resources would be that resources suffice to combine the *mfs* of these arguments to yield the falsum argument Z_2 in Fig.2ii. Then $Z_2 \rightsquigarrow_E X$ on F and G (since $\text{sup}(Z_2) \subseteq \text{SubCom}(E \cup \{X\})$), and $Z_2 \rightsquigarrow_E Y$ on F and H ; Z_2 is moved to dialectically demonstrate that E commits to *mfs* that strictly entail \perp . These defeats succeed independently of preferences. Z_2 cannot be defeated as it has not committed to any arguments. Hence neither X or Y can be defended and E cannot be admissible. Observe that unlike ASPIC^+ , satisfaction of consistency in D-ASPIC^+ does not: 1) assume resources suffice to construct A, B and C (in Fig.1iv); 2) require that \prec satisfy the property of being reasonable.

Definition 5. Let $(\mathcal{A}, \mathcal{D})$ be a DAF. We define properties that \mathcal{A} may satisfy:

- (P1) $\forall X \in \mathcal{A} : X' \in \text{Sub}(X)$ implies $X' \in \mathcal{A}$.
- (P2) $\forall X \in \mathcal{A} : X' \in [X]$ implies $X' \in \mathcal{A}$.
- (P3) If $(\Delta, \emptyset, \varphi)_X \in \mathcal{A}$ and $(\Gamma, \emptyset, \psi)_Y \in \mathcal{A}$, and $\varphi = \neg\psi$, then $(\Delta \cup \Gamma, \emptyset, \perp)_Z \in \mathcal{A}$.²

P1, P2 and P3 encode the assumption that resources suffice to: (P1) construct all sub-arguments of a given argument X ; (P2) differentially label the *mfs* of a given argument X ; (P3) combine the *mfs* of arguments with contradictory claims, to construct a falsum argument that strictly entails \perp .

We now state satisfaction of the consistency and closure postulates, having first defined the claims of arguments in an extension E as the conclusions of arguments in E that make no suppositions (intuitively, the argumentation defined inferences are the conclusions only of arguments one commits to):

$$\text{claims}(E) = \bigcup_{(\Delta, \emptyset, \phi)_X \in E} \phi$$

Theorem 1. Direct Consistency: Let E be an admissible extension³ of a DAF $(\mathcal{A}, \mathcal{D})$. If \mathcal{A} satisfies P1, P2 and P3, then $\forall \alpha, \beta \in \text{claims}(E), \alpha \neq \perp$ and $\alpha \not\prec \beta$.

Assuming only that \mathcal{A} satisfies P1, one can also show that for any complete extension E :

Closure under Strict Rules: If E contains arguments $A_1 \dots A_n$ with respective claims $\phi_1 \dots \phi_n$, and $r = \phi_1, \dots, \phi_n \rightarrow \phi \in \mathcal{R}_s$, and \mathcal{A} includes an argument A that extends $A_1 \dots A_n$ with r so as to strictly claim ϕ , then $A \in E$;

²Which given P2 implies that $(\emptyset, \Delta \cup \Gamma, \perp)_{Z'} \in \mathcal{A}$.

³Note: ASPIC^+ only shows consistency for *complete* extensions.

Sub-argument Closure: $\forall X \in \text{SubCom}(E), X \in E$ (every sub-argument of a committed argument in E , is also in E).

The following key results that hold for Dung AFs, also hold for DAFs: 1) Every admissible extension of a DAF is a subset of a preferred extension; 2) Letting $\mathcal{F}_{(\mathcal{A}, \mathcal{D})} : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ be the characteristic function of a DAF $(\mathcal{A}, \mathcal{D})$, defined as $\mathcal{F}_{(\mathcal{A}, \mathcal{D})}(E) = \{X \in \mathcal{A} \mid X \text{ is acceptable w.r.t. } E\}$, then there exists a unique least fixed point of $\mathcal{F}_{(\mathcal{A}, \mathcal{D})}$ (the grounded extension of $(\mathcal{A}, \mathcal{D})$).

4 Non-Contamination

Contamination may occur when the argumentation defined inferences (the claims of justified arguments) of a theory \mathcal{T} are invalidated when ‘expanding’ the theory with syntactically disjoint domain specific information. That is, when the defeasible rules \mathcal{R}_d and premises (ordinary or axiom) \mathcal{K} in \mathcal{T} are expanded to include premises \mathcal{R}'_d \mathcal{K}' and that are syntactically disjoint from \mathcal{R}_d and \mathcal{K} . We now specify the notions of syntactic disjointness and the expansion of a theory.

For any set of wff Δ in a language \mathcal{L} , $\text{sym}(\Delta)$ denotes the set of predicate, function and constant symbols in Δ .⁴ Henceforth, for any Δ_1, Δ_2 :

$\Delta_1 \parallel \Delta_2$ denotes $\text{sym}(\Delta_1) \cap \text{sym}(\Delta_2) = \emptyset$ (i.e., Δ_1 and Δ_2 are syntactically disjoint)

When specifying the symbols appearing in premises and defeasible rules, we need to account for the names assigned (by n) to these rules, as well as formulae declared to be in conflict (by \neg). To illustrate the latter, suppose an argument X constructed from \mathcal{T} , where X includes the ordinary premise *married*. Then, if *bachelor* = \neg *married*, *bachelor* should be considered to be in the symbols of \mathcal{T} , since an additional argument claiming *bachelor*, constructed using \mathcal{K}' and \mathcal{R}'_d , may attack X and so legitimately challenge the justified status of X . Henceforth, we write $\alpha \in r$ to denote that α appears in the antecedent, or is the consequent, of a defeasible rule r .

Definition 6. Let $\text{con_dr}(r) = \{n(r)\} \cup \overline{n(r)} \cup \bigcup_{\alpha \in r} \bar{\alpha}$, and:
 – $\text{conset}(\mathcal{R}_d \cup \mathcal{K}) = \bigcup_{r \in \mathcal{R}_d} \text{con_dr}(r) \cup \bigcup_{\alpha \in \mathcal{K}} \bar{\alpha}$;
 – $\text{base}(\mathcal{R}_d \cup \mathcal{K}) = \bigcup_{\alpha \in r, r \in \mathcal{R}_d} \alpha \cup \mathcal{K} \cup \text{conset}(\mathcal{R}_d \cup \mathcal{K})$

Let \mathcal{T}' be a D-ASPIC⁺ theory $(\mathcal{A}\mathcal{S}', \mathcal{K}')$, where

- $\mathcal{A}\mathcal{S}' = \langle \mathcal{L}, \neg, \mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_{d_1} \cup \mathcal{R}_{d_2}, n' \rangle$;
- $\mathcal{K}' = \mathcal{K}_1 \cup \mathcal{K}_2$

Then $\mathcal{T} = (\langle \mathcal{L}, \neg, \mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_{d_1}, n \rangle, \mathcal{K}_1)$ is said to be ‘expanded by the syntactically disjoint $(\mathcal{R}_{d_2}, \mathcal{K}_2)$ to obtain \mathcal{T}' ’, denoted $\mathcal{T}' = \mathcal{T} \oplus (\mathcal{R}_{d_2}, \mathcal{K}_2)$, iff $\forall r \in \mathcal{R}_{d_1}, n(r) = n'(r)$, and:

$$\text{base}(\mathcal{R}_{d_1} \cup \mathcal{K}_1) \parallel \text{base}(\mathcal{R}_{d_2} \cup \mathcal{K}_2)$$

Henceforth, we let \mathcal{A} be the arguments constructed from \mathcal{T} and the preference relation \prec over \mathcal{A} , and \mathcal{A}' the arguments constructed from \mathcal{T}' and the preference relation \prec' over \mathcal{A}' . It is straightforward to see that $\mathcal{A} \subseteq \mathcal{A}'$. Moreover, we assume that preferences over arguments in \mathcal{A} are unchanged when incorporating \mathcal{R}_{d_2} and \mathcal{K}_2 (i.e., $\prec = \prec' \cap \mathcal{A} \times \mathcal{A}$). Let us now revisit the example illustrating contamination in Fig. 1i):

⁴As usual, propositional atomic formulae are interpreted as predicate symbols of arity 0.

Example 2. Let \mathcal{T} be a theory s.t. \mathcal{L} is a propositional language, $\forall \phi \in \mathcal{L}, \phi$ is a contradictory of $\neg\phi$, $\mathcal{R}_s = \vdash_{CL}$ (\mathcal{T} is said to be a ‘classical propositional theory’), and:

- $\mathcal{R}_{d_1} = \{t \Rightarrow s\}, n(t \Rightarrow s) = r_1, \mathcal{K}_1 = \mathcal{K}_{p_1} = \{q \supset \neg s, t\}$, and $\prec = \emptyset$.

- Let $\mathcal{T}' = \mathcal{T} \oplus (\mathcal{R}_{d_2} = \{k, r \Rightarrow p\}, \mathcal{K}_2 = \mathcal{K}_{p_2} = \{k, r, \neg p\})$, where $n'(t \Rightarrow s) = r_1, n'(k, r \Rightarrow p) = r_2, \prec' = \emptyset$.

Note that $\text{base}(\mathcal{R}_{d_1}, \mathcal{K}_1) \parallel \text{base}(\mathcal{R}_{d_2}, \mathcal{K}_2)$. However, suppose \mathcal{K}_{p_2} had also included g , and $g \in \bar{r}_1$. Then the respective symbol bases would not be syntactically disjoint (the argument $[g] : g$ would undercut an argument that uses $t \Rightarrow s$).

Fig. 2iii shows arguments \mathcal{A} and \mathcal{A}' respectively constructed from \mathcal{T} and \mathcal{T}' (note, we do not show these arguments with all their differential labellings - i.e., their logical equivalents).

Observe that B is in the grounded extension of $(\mathcal{A}, \mathcal{D})$. Now notice that $A \in \mathcal{A}'$ defeats $B \in \mathcal{A}'$. However, unlike ASPIC⁺, B is still in the grounded extension E of $(\mathcal{A}', \mathcal{D}')$. Intuitively, given that resources suffice to strictly extend A_1 and A_6 so as to explosively entail q , then resources suffice to construct the corresponding falsum argument $F = (\emptyset, \{A_1, A_6\}, \perp)$. Since F has empty commitments, it cannot be defeated, and so $F \in E$. Moreover, F defends $B \in E$ since $F \Rightarrow_{\{A\}} A$ independently of preferences. Intuitively, F dialectically demonstrates that A is a contaminating argument by stint of committing to inconsistent fallible information (A_1 and A_6).

Example 3. Let \mathcal{T} be a classical propositional theory, and:

- $\mathcal{R}_{d_1} = \{e \Rightarrow \neg s\}, n(e \Rightarrow \neg s) = r_1, \mathcal{K}_1 = \mathcal{K}_{p_1} = \{e, s\}$. Fig. 2iv shows arguments C, C_1 and F . Let $C \prec F$.

- Let $\mathcal{T}' = \mathcal{T} \oplus (\mathcal{R}_{d_2} = \{g \Rightarrow p\}, \mathcal{K}_2 = \mathcal{K}_{p_2} = \{g\})$, where $n'(e \Rightarrow \neg s) = r_1, n'(g \Rightarrow p) = r_2$. Fig. 2iv) shows the additional arguments D, C_2 and C_3 . Let $C \prec' F, D \not\prec' F$.

Adding the syntactically disjoint $(\mathcal{R}_{d_2}, \mathcal{K}_2)$ results in contamination, since (as in ASPIC⁺) F is in the grounded extension of $(\mathcal{A}, \mathcal{D})$ but not $(\mathcal{A}', \mathcal{D}')$. However, non-contamination would be satisfied if adding syntactically disjoint information to an argument does not change the strength of the argument. Then one would have that $D \prec F$, hence $D \not\prec_E F$, and so F would be in the grounded extension.

Let us now more formally set out the conditions for satisfaction of non-contamination. Firstly, the following can be shown (in what follows we write $\text{DP}(A)$ to denote the defeasible rules and premises in an argument A):

Definition 7. \vdash is a regular logic if it satisfies the following:

If $\Gamma \vdash \alpha$ and $\exists \Delta \subseteq \Gamma$ such that $\Delta \parallel \Gamma \setminus \Delta \cup \{\alpha\}$, then: either $\Delta \vdash \perp$ or $\Gamma \setminus \Delta \vdash \alpha$.

Note, it is easy to show that classical logic is regular (see Proposition 30 in [D’Agostino and Modgil, 2018]).

Proposition 1. Let \mathcal{A} be the arguments defined by a theory $(\mathcal{A}\mathcal{S}, \mathcal{K})$ such that \mathcal{R}_s in $\mathcal{A}\mathcal{S}$ encodes a regular logic \vdash . Let $X \in \mathcal{A}$ such that $\text{conc}(X) = \alpha$ and $\Gamma = \text{DP}(X), \emptyset \neq \Delta \subseteq \Gamma$ and $\Delta \parallel \Gamma \setminus \Delta \cup \{\alpha\}$. Then

(P4) Either $\exists X' = (\Pi, \Sigma, \perp) \in \mathcal{A}$ such that $\text{DP}(X') = \Delta$, or $\exists X' = (\Pi, \Sigma, \alpha) \in \mathcal{A}$ such that $\text{DP}(X') = \Gamma \setminus \Delta$.

Henceforth we assume \mathcal{R}_s is regular and that P4 is satisfied by the arguments constructed from a theory. That is to say, given an argument of the form X in Proposition 1, resources suffice to:

- (i) construct the required X' with $\text{conc}(X') = \perp$, if Δ is inconsistent (in Example 3, A includes ‘inconsistent components A_1 and A_6 ’ and so $F \in \mathcal{A}'$); or else
- (ii) to construct X' with $\text{conc}(X') = \alpha$. In this case X' is said to be the ‘non-redundant counterpart’ of X (in Example 3, $D \in \mathcal{A}'$ and so resources suffice to construct its non-redundant counterpart, i.e., $C \in \mathcal{A}'$).

In general, non-contamination is then satisfied if a non-redundant counterpart X' is neither weakened or strengthened when adding syntactically disjoint information, to yield the argument X (see Definition 8). However, if \mathcal{T} contains no defeasible rules⁵ – i.e., \mathcal{T} is a *strict theory* – then it suffices that arguments are not strengthened. In general, whether or not P4 can be satisfied with bounded resources, depends on whether the proof-theoretical means used to obtain strict inferences is such that given a proof of a conclusion that is syntactically disjoint from the premises, we can feasibly (with virtually no computational cost) turn it into a proof of \perp from the same premises. A classical natural deduction system satisfying this requirement is presented in [D’Agostino and Modgil, 2018] and [D’Agostino et al., 2019].

Definition 8. $\forall X, Y, Y'$ such that:

$$\begin{aligned} \text{DP}(Y') &= \Gamma, \text{DP}(Y) = \Delta \cup \Gamma, \text{ and } \Delta \parallel \Gamma \cup \{\alpha\}; \\ \text{conc}(Y') &= \text{conc}(Y) = \alpha, \text{ and } \forall r, r' \in \Gamma, r' \text{ is a child} \\ &\text{of } r \text{ in } Y' \text{ iff } r' \text{ is a child of } r \text{ in } Y \end{aligned}$$

- \prec is said to be non-strengthening if:
 $Y' \prec X$ implies $Y \prec X$ and $X \not\prec Y'$ implies $X \not\prec Y$, and
- \prec is said to be non-weakening if:
 $X \prec Y'$ implies $X \prec Y$ and $Y' \not\prec X$ then $Y \not\prec X$;

Now, let the consequences of a theory $\mathcal{T} = (AS, \mathcal{K})$ be defined as $Cn(\mathcal{T}) = \{\text{claims}(E_1), \dots, \text{claims}(E_n)\}$, where E_1, \dots, E_n are the complete extensions of the *DAF* $(\mathcal{A}, \mathcal{D})$ defined by \mathcal{T} . Also, let $Cn(\mathcal{T})_{Sy}$ denote the formulae in $Cn(\mathcal{T})$ whose symbols are also in the set Sy . We now state satisfaction of the first of the non-contamination postulates – *non-interference* – studied in [Caminada et al., 2012].

Theorem 2. Non-Interference. Let $\mathcal{T}' = \mathcal{T} \oplus (\mathcal{R}_{d_2}, \mathcal{K}_2)$. Assume that: 1) preferences are non-strengthening in the case of strict theories, and both non-strengthening and non-weakening in the case that theories include defeasible rules; 2) the arguments \mathcal{A} and \mathcal{A}' respectively constructed from \mathcal{T} and \mathcal{T}' , satisfy P1-P4. Then:

$$Cn(\mathcal{T})_{\text{sym}(\mathcal{R}_{d_1}, \mathcal{K}_1)} = Cn(\mathcal{T}')_{\text{sym}(\mathcal{R}_{d_1}, \mathcal{K}_1)}$$

Intuitively, the above result follows from the result that whenever $(\mathcal{A}, \mathcal{D})$ is defined by \mathcal{T} and $(\mathcal{A}', \mathcal{D}')$ is defined by \mathcal{T}' , and $\mathcal{T}' = \mathcal{T} \oplus (\mathcal{R}_{d_2}, \mathcal{K}_2)$, then E is a complete extension of $(\mathcal{A}, \mathcal{D})$ iff E' is a complete extension of $(\mathcal{A}', \mathcal{D}')$, where

⁵E.g., when *D-ASPIC*⁺ formalises classical logic argumentation with preferences as in [D’Agostino and Modgil, 2018].

⁶ r' is a child of r in X iff r' is the top rule of some $X' \in \text{Sub}(X)$ that is a proper sub-argument of an $X'' \in \text{Sub}(X)$ with top rule r .

$E = E' \cap \mathcal{A}$. For brevity, we omit here the second non-contamination postulate – *crash resistance* – which follows straightforwardly from *non-interference*.

5 Conclusions

This paper is the first to study rationality under resource bounds, in the context of general frameworks for structured argumentation. A number of works show satisfaction of non-contamination for specific instances of *ASPIC*⁺. In [Caminada et al., 2012], propositional logic programming and Default Logic instantiations of Dung frameworks are shown to satisfy these postulates, but only under the semi-stable semantics. In [Wu and Podlaszewski, 2015], non-contamination is shown assuming syntactically restricted strict inference rules, no preferences, and frameworks ‘cleansed of’ (excluding) inconsistent arguments. [Heyninck and Straßer, 2017] prove all rationality postulates, only for the grounded semantics, and for a version of *ASPIC*⁺ restricted to axiom premises, without undercuts, and with preferences specifically defined by a weakest linking lifting of a total ordering over defeasible rules (generalised to preorders and a different lifting principle in [Heyninck and Straßer, 2019]). [Borg and Straßer, 2018] study approaches to structured argumentation that do not make use of preferences, and that assume a single set of premises. In particular, they identify strict rules encoding ‘pre-relevant’ logics as a sufficient condition for satisfying non-contamination (note that classical logic is not pre-relevant, so that non-contamination is only satisfied by cleansed frameworks). Finally, [D’Agostino and Modgil, 2018] distinguish between commitments and suppositions in the context of classical logic argumentation with preferences. Full rationality is shown under conditions analogous to P1-P4 (and without requiring that one check that the premises of arguments are consistent and subset minimal). Moreover, they consider the use of a recently proposed natural deduction proof theory for propositional classical logic [D’Agostino et al., 2019], which does not yield arguments that redundantly use syntactically disjoint information. The use of such a proof theory means that satisfaction of non-contamination is not contingent on preference relations for strict theories being non-strengthening. We believe it is straightforward to then show that in the context of *D-ASPIC*⁺, if \mathcal{R}_s is defined in terms of the inferences licensed by such ‘non-contaminating proof systems’, satisfaction of non-contamination can then be shown assuming *any* preference relation.

To conclude, the *D-ASPIC*⁺ framework for structured argumentation is the first to establish: 1) conditions under which non-contamination is satisfied for all complete semantics and without compromising the generality of *ASPIC*⁺; 2) minimally demanding requirements as to the resources available for constructing arguments, that suffice to guarantee satisfaction of *all* rationality postulates. This has significant implications for a wide range of existing and novel argumentation formalisms and non-monotonic logics that can be formalised in *ASPIC*⁺, rendering them more suitable for use by resource bounded real-world agents engaging in epistemic reasoning and decision making, either as individuals or via dialogue.

References

- [Borg and Straßer, 2018] Anne Marie Borg and Christian Straßer. Relevance in structured argumentation. In *Proc. Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 1753–1759, 2018.
- [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 Caminada, Walter Carnielli, and Paul Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22(5):1207–1254, 2012.
- [Caminada *et al.*, 2014] Martin Caminada, Sanjay Modgil, and Nir Oren. Preferences and unrestricted rebut. In *Computational Models of Argument: Proceedings of COMMA 2014*, pages 209–220. IOS Press, 2014.
- [D’Agostino and Modgil, 2018] Marcello D’Agostino and Sanjay Modgil. Classical logic, argument and dialectic. *Artificial Intelligence*, 262:15–51, 2018.
- [D’Agostino *et al.*, 2019] Marcello D’Agostino, Dov Gabbay, and Sanjay Modgil. Normality, non-contamination and logical depth in classical natural deduction. *Studia Logica*, pages 1–67, 2019.
- [Dong *et al.*, 2019] Huimin Dong, Beishui Liao, Reka Markovich, and Leon van der Torre. From classical to non-monotonic deontic logic using aspic+. In *International Workshop on Logic, Rationality and Interaction (LORI 2019)*, pages 71–85. Springer, 2019.
- [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.
- [Heyninck and Straßer, 2017] Jesse Heyninck and Christian Straßer. Revisiting unrestricted rebut and preferences in structured argumentation. In *Proc. Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 1088–1092, 2017.
- [Heyninck and Straßer, 2019] Jesse Heyninck and Christian Straßer. A fully rational argumentation system for pre-ordered defeasible rules. In *Proc. 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS’19)*, pages 1704–1712, 2019.
- [Modgil and Prakken, 2013] Sanjay Modgil and Henry Prakken. A general account of argumentation and preferences. *Artificial Intelligence*, 195(0):361 – 397, 2013.
- [Modgil and Prakken, 2018] Sanjay Modgil and Henry Prakken. Abstract rule-based argumentation. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation, Vol. 1*, pages 286–361. London: College Publications, 2018.
- [Wu and Podlaszewski, 2015] Yining Wu and Mikolaj Podlaszewski. Implementing crash-resistance and non-interference in logic-based argumentation. *Journal of Logic and Computation*, 25:303–333, 2015.