Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Prioritised Default Logic as Rational Argumentation

Anthony P. Young  
Department of Informatics  
King’s College London  
Strand, London, U.K.  
peter.young@kcl.ac.uk

Sanjay Modgil  
Department of Informatics  
King’s College London  
Strand, London, U.K.  
sanjay.modgil@kcl.ac.uk

Odinaldo Rodrigues  
Department of Informatics  
King’s College London  
Strand, London, U.K.  
odinaldo.rodrigues@kcl.ac.uk

ABSTRACT

We endow Brewka’s prioritised default logic (PDL) with argumentation semantics using the ASPIC+ framework for structured argumentation. We prove that the conclusions of the justified arguments correspond to the prioritised default extensions in a normatively rational manner. Argumentation semantics for PDL will allow for the application of argument game proof theories to the process of inference in PDL, making the reasons for accepting a conclusion transparent and the inference process more intuitive. This also opens up the possibility for argumentation-based distributed reasoning and communication amongst agents with PDL representations of mental attitudes.

General Terms

Theory

Keywords

Abstract argumentation; ASPIC+; Prioritised Default Logic; agent reasoning and communication

1. INTRODUCTION

Dung’s argumentation theory [8] has become established as a general framework for non-monotonic reasoning (NMR). Given a set of well-formed formulae (wffs) $\Delta$ in some non-monotonic logic (NML), the arguments and attacks defined by $\Delta$ instantiate a Dung argumentation framework. Additionally, a preference relation over the defined arguments can be used to determine which attacks succeed as defeats. The justified arguments are then evaluated under various Dung semantics, and the claims of the sceptically justified arguments (i.e., arguments contained in all extensions under some semantics) identify the inferences from the underlying $\Delta$. More formally, given an argumentation framework $AF$ and a wff $\theta$, the argumentation-defined inference relation $\models_{AF}$ over $\Delta$ is $\Delta \models_{AF} \theta$ iff $\theta$ is the conclusion of a sceptically justified argument in $AF$. Indeed, a correspondence has been shown between $\models_{AF}$ over $\Delta$, and the instantiating logic’s non-monotonic inference relation defined directly over $\Delta$. For example, default logic (DL) [8], logic programming [8], defeasible logic [10] and preferred subtheories (PS) [13] have all been endowed with argumentation semantics. This in turn allows the application of argument game proof theories [12] to the process of inference, and the generalisation of these dialectical proof theories to distributed NMR amongst computational agents, whereby agents can engage in argumentation-based dialogues, submitting arguments and counter-arguments from their own non-monotonic knowledge bases [1, 11, 14]. Furthermore, argumentative characterisations of NMR make use of principles familiar in everyday reasoning and debate, thus rendering transparent the reasons for accepting a conclusion and allowing for human participation and inspection of the inference process.

One well-studied NML that has not yet been endowed with argumentation semantics is Brewka’s prioritised default logic (PDL) [4]. PDL is important because it upgrades DL [16] with priorities over defaults, so that, for example, one can account for recent information taking priority over information in the distant past, or that more specific information should take priority over more general information. PDL has also been used to represent the (possibly conflicting) beliefs, obligations, intentions and desires (BOID) of agents, and model how these different categories of mental attitudes override each other in order to generate goals and actions that attain those goals [6].

This paper contributes to research in argumentative characterisations of NMR, by endowing PDL with argumentation semantics, proving a correspondence between PDL inference and the inference relation defined by the argumentation semantics, and proving that the result is normatively rational. We realise these contributions by appropriately instantiating the ASPIC+ framework for structured argumentation [13]. ASPIC+ identifies conditions under which logics and preference relations instantiating Dung’s frameworks satisfy the Caminada-Amgoud rationality postulates [7].

In Section 2 we review ASPIC+ and PDL. In Section 3 we define a PDL instantiation of ASPIC+. In Section 5 we present a representation theorem proving that inferences defined by the argumentation semantics correspond exactly to inferences in PDL under an appropriate preference relation devised in Section 4. We will also prove that this instantiation is normatively rational in the sense of [7]. In Section 6 we discuss possible generalisations of PDL via its argumentation semantics.

2. BACKGROUND

In the remainder of this paper we make use of the following notation: ‘$\models$’ means “is defined as”. $N$ is the set of natural numbers. For a set $X$ its power set is $\mathcal{P}(X)$ and its finite power set (set of all finite subsets) is $\mathcal{P}_{\text{fin}}(X)$. If $X \subseteq Y$ iff $X$ is a finite subset of $Y$, therefore $X \in \mathcal{P}_{\text{fin}}(Y) \iff X \subseteq \mathcal{P}_{\text{fin}}(Y)$. Undefined quantities are denoted by $\ast$, for example $1/0 = \ast$. 


Copyright © 2016, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.
We now recap the key definitions of \[S\]. An argumentation framework is a directed graph \((A, C)\), where \(A\) is the set of arguments and \(C \subseteq A^2\) is the conflict relation on \(A\). For arguments \(A, B \in A\) we write \(C(A, B) \Leftrightarrow (A, B) \in C \Leftrightarrow A\) conflicts with \(B\), i.e. \(A\) is a counterargument against \(B\).

In what follows let \(S \subseteq A\) and \(A, B \in A\). \(S\) conflicts with \(B\) iff \((\exists A \in S)\)(\(C(A, B)\)). \(S\) is conflict-free (cf) iff \(C \cap S^2 = \emptyset\). \(S\) defends \(A\) iff \((\forall B \in A)\)[\(C(B, A) \Rightarrow S\) conflicts with \(B\)]. Let \(Def(S) := \{ A \in A | S \text{ defends } A \} \subseteq A\). \(S\) is an admissible extension iff \(S\) is cf and \(S \subseteq Def(S)\). An admissible extension \(S\) is: a complete extension iff \(S = Def(S)\); a preferred extension iff \(S\) is a \(\subseteq\)-maximal complete extension; the grounded extension \(S\) is \(S\) is \(\subseteq\)-least complete extension; a stable extension \(S\) is complete and conflicts with all arguments \(A \in A - S\).

Let \(S := \{ \text{complete, preferred, grounded, stable} \} \) be the set of Dung semantics. An argument \(A \in A\) is sceptically justified under the semantics \(s \in S\) iff \(A\) belongs to all \(s\)-extensions of \((A, C)\).

### 2.2 The ASPIC\(^+\) Framework

#### 2.2.1 Arguments, Attacks, Preferences and Defeats

Dung’s framework provides an intuitive calculus of opposition for determining the justified arguments based on conflict alone; it abstracts from the internal logical structure of arguments, the nature of defeats and how they are determined by preferences, and consideration of the conclusions of the arguments. However, these features are referenced when studying whether any given logical instantiation of a framework yields complete extensions that satisfy the rationality postulates of \([7]\). ASPIC\(^+\) \([13]\) provides a structured account of abstract argumentation, allowing one to reference the above features, while at the same time accommodating a wide range of instantiating logics and preference relations. \(^1\) ASPIC\(^+\) then identifies conditions under which complete extensions defined by the arguments, attacks, preferences, satisfy the rationality postulates of \([7]\), and are hence normatively rational.

In ASPIC\(^+\), the tuple \((L, -; R_d, R_d, n)\) is an argumentation system, where \(L\) is a logical language, \(\sim \) : \(L \to \mathcal{P}(L)\) is the contrary function \(\theta \to \bar{\theta}\) that identifies when one wff in \(L\) conflicts with another. Let \(\theta_1, \ldots, \theta_m, \phi \in L\) be wffs for \(m \in \mathbb{N}^+\), \(R_d\) is the set of strict inference rules of the form \((\theta_1, \ldots, \theta_m \rightarrow \phi)\), denoting that if \(\theta_1, \ldots, \theta_m\) are then \(\phi\) is also true, and \(R_d\) is the set of defeasible inference rules of the form \((\theta_1, \ldots, \theta_m \rightarrow \phi)\), denoting that if \(\theta_1, \ldots, \theta_m\) are true then \(\phi\) is tentatively true. For a strict or defeasible rule \(r = (\theta_1, \ldots, \theta_m \rightarrow \phi)\), we define \(\text{Ante}(r) := (\theta_1, \ldots, \theta_m) \subseteq \text{m}\) \(L\), and \(\text{Cons}(r) := \phi \in L\). Finally \(n : R_d \to L\) is a partial function that assigns a name to some of the defeasible rules. For any \(S \subseteq L\) we define the set \(CL_{R_d}(S) \subseteq L\) to be the smallest superset of \(S\) that also contains \(\text{Cons}(r)\) for all \(r \in R_d\) such that \(\text{Ante}(r) \subseteq CL_{R_d}(S)\). We call \(CL_{R_d}\) the closure under strict rules operator.

\(^1\)ASPIC stands for Argumentation Service Platform with Integrated Components.

In ASPIC\(^+\), a knowledge base is a set \(K := K_a \cup K_p \subseteq L\) where \(K_a\) is the set of axioms and \(K_p\) is the set of ordinary premises. Intuitively, the knowledge base consists of premises used in the construction of arguments. Given an argumentation system and knowledge base, an argument is defined inductively as:

1. (Base) \([\emptyset]\) is a singleton argument with \(\emptyset \in K\), conclusion \(\text{Cons}(\emptyset) := \emptyset\), premise set \(\text{Prem}(\emptyset) := \emptyset \subseteq K\), top rule \(\text{TopRule}(\emptyset) := *\) and set of subarguments is \(\text{Sub}(\emptyset) := \{ \emptyset \}\).

2. (Inductive, strict) Let \(A_1, \ldots, A_n\) be arguments with respective conclusions \(\text{Cons}(A_1), \ldots, \text{Cons}(A_n)\) and premise sets \(\text{Prem}(A_1), \ldots, \text{Prem}(A_n)\). If there is a strict rule \(r = (\text{Cons}(A_1), \ldots, \text{Cons}(A_n) \rightarrow \phi) \in R_a\), then \(B := [A_1, \ldots, A_n \rightarrow \phi] \in R_a\) is also an argument with \(\text{Cons}(B) = \phi\), premises \(\text{Prem}(B) := \bigcup_{i=1}^n \text{Prem}(A_i)\), \(\text{TopRule}(B) = r \in R_d\) and set of subarguments \(\text{Sub}(B) := \{ B \} \cup \bigcup_{i=1}^n \text{Sub}(A_i)\).

3. (Inductive, defeasible) Let \(A_1, \ldots, A_n\) be arguments with respective conclusions \(\text{Cons}(A_1), \ldots, \text{Cons}(A_n)\) and premise sets \(\text{Prem}(A_1), \ldots, \text{Prem}(A_n)\). If there is a defeasible rule \(r' := (\text{Cons}(A_1), \ldots, \text{Cons}(A_n) \rightarrow \phi) \in R_d\), then \(C := [A_1, \ldots, A_n \Rightarrow \phi] \in R_d\) with \(\text{Cons}(C) = \phi\), premises \(\text{Prem}(C) := \bigcup_{i=1}^n \text{Prem}(A_i)\), \(\text{TopRule}(C) = r' \in R_d\) and set of subarguments \(\text{Sub}(C) := \{ C \} \cup \bigcup_{i=1}^n \text{Sub}(A_i)\).

Let \(A\) be the (unique) set of all arguments constructed in this way. It is clear that arguments are finite objects.

Two arguments are equal iff they are constructed identically in the above manner. We say \(A\) is a subargument of \(B\) iff \(A \subseteq B\), and we write \(A \triangleleft B\). We say \(A\) is a proper subargument of \(B\) iff \(A \subseteq B\) \(- \{ B \}\), and we write \(A \triangleleft B\). It can be shown that \(\triangleleft\) is at least a preorder on \(\text{Sub}(B)\).

An argument \(A \in A\) is firm iff \(\text{Prem}(A) \subseteq K_a\). Further, \(DR(A) \subseteq R_d\) is the set of defeasible rules applied in constructing \(A\). An argument \(A\) is strict iff \(DR(A) = \emptyset\), else \(A\) is defeasible. Given \(R \subseteq R_d\), we introduce the set of all arguments freely constructed with defeasible rules restricted to those in \(R\) as the set \(\text{Args}(R) \subseteq A\), which are all arguments with premises in \(K\), strict rules in \(R_a\) and defeasible rules in \(R\). Formally, \(\text{Args}(R)\) is defined inductively just as arguments are constructed. It is easy to show that \(A \in \text{Args}(R) \Rightarrow DR(A) \subseteq R\). Clearly, \(\text{Args}(R_a) = A\). Given \(R\), \(\text{Args}(R)\) is unique. Further, \(\text{Args}(R)\) is closed under subarguments, i.e. \(A \in \text{Args}(R)\) and \(B \triangleleft A\) implies \(B \in \text{Args}(R)\).

An argument \(A\) attacks another argument \(B\), denoted as \(A \rightarrow B\), iff at least one of the following hold, where:

1. \(A\) is said to undermine attack \(B\) on the subargument \(B' = [\emptyset]\) iff \(\exists \phi \in \text{Prem}(B) \cap K_a\) \(\text{Cons}(A) \in \phi\), i.e. \(A\) attacks with some ordinary premise of \(B\).

2. There is some \(B' \triangleleft B\) such that \(r := \text{TopRule}(B') \in R_a, \phi := \text{Cons}(r)\) and \(\text{Cons}(A) \in \phi\). \(A\) is then said to rebut attack \(B\) on the subargument \(B'\).

3. There is some \(B' \triangleleft B\) such that \(r := \text{TopRule}(B') \in R_d\) and \(\text{Cons}(A) \in \text{Sub}(r)\). \(A\) is then said to undercut attack \(B\) on the subargument \(B'\) (by arguing against the application of the defeasible rule \(r\) in \(B\)).
A preference relation over arguments is then used to determine which attacks succeed as defeats. We denote the preference \(\preceq \subseteq A^2\) (not necessarily a preorder for now) such that \(A \preceq B \Leftrightarrow A \rightarrow B\). The strict version is \(A < B \Leftrightarrow [A \preceq B, B \not\preceq A]\), and equivalence is \(A \approx B \Leftrightarrow [A \preceq B, B \preceq A]\). We define a defeat as

\[ A \rightarrow B \equiv (\exists B' \subset_{arg} B) [A \rightarrow B', A \not\approx B'] \quad (2.1)\]

That is to say, \(A\) defeats \(B\) (on \(B'\)) iff \(A\) attacks \(B\) on the subargument \(B'\), and \(B'\) is not strictly preferred to \(A\). Notice the comparison is made at the subargument \(B'\) instead of the whole argument \(B\). We then abuse notation to define the defeat relation as \(\preceq \subseteq A^2\) such that \((A, B) \in \preceq \Leftrightarrow A \rightarrow B\). A set of arguments \(S \subseteq A\) is conflict-free (cf) iff \(\cap S = \emptyset\).

Preferences between arguments are calculated from the argument structure by endowing \(K_\theta\) and \(R_\theta\) with preorders \(\preceq_\theta\) and \(\leq_\theta\) respectively, where (e.g.) \(r_1 \preceq_\theta r_2\) iff \(r_2\) is more preferred than \(r_1\) (and analogously for \(\leq_\theta\)). This preorder is then lifted to a set-comparison order \(\preceq\) between the sets of premises or defeasible rules of the arguments, and then finally to \(\preceq\), following the method in [13] Section 5. We will summarise this in Section 4.

Given the preference relation \(\preceq\) between arguments, we can calculate the structure \((A, \rightarrow, \preceq)\) an ASPIC\(^+\) SAF (structured argumentation framework), or attack graph. Its corresponding defeat graph is \((A, \rightarrow, \preceq)\), where \(\rightarrow\) is defined in terms of \(\preceq\) as in Equation (2.1).

Given \((A, \rightarrow)\) one can then evaluate the extensions under Dung’s semantics (Section 2.1) where \(\preceq\) is \(\rightarrow\), and thus identify the argumentation defined inferences as the conclusions of the sceptically justified arguments as follows. Let \(AS\) be an argumentation system. The argumentation-defined inference relation \(\models_{AS}\) is \(\models_{\K,\theta} \iff \theta = \text{Conc}(A)\) where \(A \in A\) is a sceptically justified argument.

### 2.2.2 Normative Rationality

Instantiations of ASPIC\(^+\) should satisfy some properties to ensure it is rational [2]. Given an instantiation let \((A, \rightarrow, \preceq)\) be its ASPIC\(^+\) attack graph with corresponding defeat graph \((A, \rightarrow)\). For \(S \subseteq A\) let \(\text{Conc}(S) := \bigcup_{A \in S} \text{Conc}(A)\). Let \(E\) be any of its admissible extensions. The Caminada-Angoud rationality postulates state:

1. If \(E\) is a complete extension then \(E\) is subargument closed.
2. If \(E\) is a complete extension then
\[
\text{CL}_{\rightarrow} [\text{Conc}(E)] = \text{Conc}(E) \quad (2.2)
\]
3. The sets \(\text{Conc}(E), \text{CL}_{\rightarrow} [\text{Conc}(E)] \subseteq L\) are consistent. An ASPIC\(^+\) instantiation is normatively rational iff it satisfies these rationality postulates. These postulates may be proved directly given an instantiation, as we will show for our instantiation to PDL in Theorem 5.3. ASPIC\(^+\) also identifies sufficient conditions for an instantiation to satisfy these postulates [13] Section 4, which we will discuss in Section 6.

Note that [13] studies two different notions of cf sets: one where no two arguments attack each other, and the other where no two arguments defeat each other. We choose the latter notion of cf as this is more commonplace in argumentation formalisms that distinguish between attacks and defeats, e.g. in [15].

### 2.3 Brewka’s Prioritised Default Logic

In this section we recap Brewka’s PDL [4]. We work in full first order logic (FOL) of arbitrary signature where the set of first-order formulae is \(FL\) and the set of closed first order formulae \(\mathcal{L}\) is \(\mathcal{L} \subseteq FL\), with the usual quantifiers and connectives. Given \(S \subseteq \mathcal{L}\), the deductive closure of \(S\) is \(T(S)\), and given \(\theta \in \mathcal{L}\), the addition operator \(+ P(\mathcal{L}) \times P(\mathcal{L}) \rightarrow P(\mathcal{L})\) is defined as \(S + \phi := T(S\cup\{\phi\})\).

A normal default is an expression \(\frac{\theta \vdash \phi}{\theta, \phi}\) where \(\theta, \phi \in \mathcal{L}\) and read “if \(\theta\) is the case and \(\phi\) is consistent with what we know, then \(\phi\) is the case”. In this case we call \(\theta\) the antecedent and \(\phi\) the consequent. A normal default \(\frac{\theta \vdash \phi}{\theta, \phi}\) is closed iff \(\theta, \phi \in \mathcal{L}\). We will assume all defaults are closed and normal unless stated otherwise. Given \(S \subseteq \mathcal{L}\), a default is active (in \(S\)) iff \(\theta \in S, \phi \not\in S\). Intuitively, the first requirement says we need to know the antecedent before applying the default, the second requirement is that the consequent must add new information, and the third requirement ensures that what we infer is consistent with what we know.

A finite prioritised default theory (PDT) is a structure \((D, W, <)\), where \(W \subseteq \mathcal{L}\) is not necessarily a finite set and \((D, <)\) is a finite strict total poset (partially ordered set) of defaults, where \(d' < d \Longleftrightarrow d' \text{ is more prioritised than } d\). Intuitively, \(W\) are the known facts and \(D\) the defaults that nonmonotonically extend \(W\).

The inferences of a PDT are defined by its extensions. Formally, let \(\prec < <\prec\) be a linearisation of \(<\). A prioritised default extension (with respect to \(\prec\)) (PDE) is a set \(E := \bigcup_{i \in E} E_i \subseteq \mathcal{L}\) built inductively as:

\[ E_0 := T(S) \quad (2.3) \]
\[ E_{i+1} := \begin{cases} E_i + \phi, & \text{if property 1} \\ E_i, & \text{else} \end{cases} \quad (2.4) \]

where “property 1” iff “\(\phi\) is the consequent of the \(\prec\)-greatest default \(d\) active in \(E_i\)”. Intuitively, one first generates all classical consequences from the facts \(W\), and then iteratively adds the nonmonotonic consequences from the most prioritised default to the least. Notice if \(W\) is inconsistent then \(E_0 = E = \bigcup_{i \in E} E_i\).

It can be shown that the ascending chain \(E_i \subseteq E_{i+1}\) stabilises at some finite \(i \in \mathbb{N}\) and that \(E\) is consistent provided that \(W\) is consistent. \(E\) does not have to be unique because there are many distinct linearisations of \(<\). We say the PDT \((D, W, <)\) sceptically infers \(\theta \in \mathcal{L}\) iff \(\theta \in E\) for all extensions \(E\).

Henceforth, we will refer to a PDT \((D, W, <)\) where \(<\) is a strict total order as a linearised PDT (LPDT). If \(<\) is total then there is only one way to apply the defaults in \(D\) by Equation 2.4; hence the extension is unique and all inferences are sceptical. In what follows, we will use \(\prec\) to emphasise that the order is total.

**Example 1.** PDL can be used to model the mental attitudes of agents when deliberating over which goals to pursue. Suppose a research assistant Alice (a) is considering whether she should teach undergraduates. We can model her mental attitudes as a BOID agent’s PDT [6] as follows. Define the

- i.e. first order formulae without free variables
- We have defined the order dually to [4] so as to comply with orderings over the ASPIC\(^+\) defeasible inference rules.
- i.e. \(\prec\) is a strict total order and hence \((D, \prec)\) is a strict toset (totally ordered set).
predicates $R(x) \Leftrightarrow "x\ is\ a\ research\ assistant"$, $A(x) \Leftrightarrow "x\ is\ an\ academic"$, and $T(x) \Leftrightarrow "x\ is\ teaching\ (undergraduates)"$. Alice is a research assistant, so $W = \{ R(a) \}$. She believes that research assistants are academics, so her set of beliefs $Bel$ has the default $R(a) \land A(a) \rightarrow T(a)$. She does not want to teach and would rather focus on her research, so her set of desires $Des$ include $R(a) \rightarrow \neg T(a)$. However, she is obliged to teach, so her set of obligations $Obl$ include $A(a) \rightarrow T(a)$. The set of defaults is $D = Bel \cup Des \cup Obl$, and we assume no other defaults are relevant for this example.

In [1], the relative prioritisations of categories of mental attitudes define different agent types. For example, if Alice is a realistic selfish agent, the priority (abuse of notation) is $Obl \prec^+ Des \prec^+ Bel$, and therefore the extension is $Th(\{R(a), A(a), \neg T(a)\})$. She thus generates the goal $\neg T(a)$, i.e. she does not teach. However, if she is a realistic social agent, the priority (abuse of notation) is $Des \prec^+ Obl \prec^+ Bel$, and therefore she teaches, as $T(a)$ is in the extension.

3. INSTANTIATING ASPIC+ TO PDL

We now instantiate ASPIC+ to PDL. Let $\langle D, W, \prec^+ \rangle$ be a LPDT.

1. Our arguments are expressed in FOL, so our set of wffs is $\mathcal{F}$.
2. The contrary function — syntactically defines conflict in terms of classical negation [1]. Let $\theta, \phi$ be wffs, then $\bar{\theta} \equiv \phi$ if $\theta$ is of the form $\neg \phi$; else $\phi \equiv \neg \theta$.
3. The set of strict rules $\mathcal{R}_a$ characterises inference in first order classical logic. We leave the proof theory implicit. $Cl_{\mathcal{R}_a}$ instantiates to deductive closure.
4. The set of defeasible rules $\mathcal{R}_d$ is defined as:

$$\mathcal{R}_d := \left\{ (\theta \Rightarrow \phi \mid \bar{\theta} : \phi \in D) \right\}, \quad (3.1)$$

with $n \equiv \ast$. Clearly, there is a bijection $f$ where $f : D \rightarrow \mathcal{R}_d : \bar{\theta} : \phi \mapsto \bar{\theta} : \phi \Rightarrow \phi$.

and we will define the strict version of the preorder $\preceq_D$ over $\mathcal{R}_a$ and $\mathcal{R}_d$:

$$\theta \Rightarrow \phi \prec_D \theta' \Rightarrow \phi' \iff \bar{\theta} : \phi \prec^+ \bar{\theta'} : \phi' \quad (3.3)$$

We can see that the strict toset $\langle \mathcal{R}_d, \prec_D \rangle$ is order isomorphic to $\langle D, \prec^+ \rangle$, where the non-strict version of the preorder is $\preceq_D$ is $\preceq_D$.

5. The set of axiom premises is $\mathcal{K}_a = W$, because we take $W$ to be the set of facts. Furthermore, $\mathcal{K}_p = \emptyset$.

\footnote{We will discuss why we only consider LPDTs in Section 6.}
\footnote{For example, $\bar{\theta} \land \phi$ is the contrary of $(\theta \land \neg \phi)$, but $(\theta \rightarrow \phi)$, where $\rightarrow$ in this case denotes material implication, is not the contrary of $(\theta \land \neg \phi)$.}
\footnote{Two defeasible rules are equal iff they have the same antecedents and consequent syntactically.}
\footnote{From Footnote 4, we do not need to define $\preceq_D$ as the order-theoretic dual to $\prec^+$, avoiding potential confusion as to which item is more preferred.}

The set $A$ of ASPIC+ arguments are defined as in Section 2.2. It is easy to see that all arguments are firm because $\mathcal{K}_a = \emptyset$, and so there are no undermining attacks. As $n$ is undefined, no attack can be an undercut. Therefore, we only have rebut attacks, where $A \rightarrow B$ if

$$\exists B', B'' \subseteq arg B \quad (\exists B'' \Rightarrow \text{Conc}(A)) \quad (3.4)$$

Defeats are defined as in Equation 2.1. This leaves the question as to how the argument preference $\succsim$ should be defined based on the strict total order $\prec_D$ over $\mathcal{R}_d$.

4. PREFERENCES

We want to define $\succsim$ in such a way that the extension of the LPDT $\langle D, W, \prec^+ \rangle$ corresponds to the conclusions of the justified arguments of the defeat graph $\langle A, \rightarrow \rangle$, instantiated by the corresponding ASPIC+ instantiation, and the result is rational. In ASPIC+, preferences over arguments are calculated from the argument structure, and by comparing the fallible information (ordinary premises and defeasible rules) they contain. In our instantiation, we only compare defeasible rules as there are no ordinary premises. Such a comparison is then lifted to a comparison between the sets of defeasible rules of two arguments, which is then lifted to $\succsim$.

More formally, ASPIC+ defines the elitist order [12] Section 5), where for $A, B \in A$ and $DR(A), DR(B) \subseteq \mathcal{R}_d$, the argument preference $\succsim$ is

$$A \succsim B \iff DR(A) \subseteq \\mathcal{E}_\mathcal{E}_i \mathcal{E}_i DR(B), \quad (4.1)$$

such that for $\Gamma^0 \subseteq \mathcal{F}$, $\Gamma' \subseteq \mathcal{E}_\mathcal{E}_i \mathcal{E}_i \mathcal{R}_d$,

$$\Gamma \subseteq \mathcal{E}_\mathcal{E}_i \quad \Gamma' \subseteq \mathcal{E}_i \Gamma^{= \ast} \quad \text{and} \quad (4.2)$$

$$\Gamma \subseteq \mathcal{E}_\mathcal{E}_i \mathcal{E}_i \Gamma' \subseteq \exists (\exists x \in \Gamma) \left( \forall y \in \Gamma^0 \right) x < \prec_D y, \quad (4.3)$$

where in Equation 4.3 the order $\prec_D$ is defined by $\Gamma^0 \mathcal{E}_\mathcal{E}_i \mathcal{E}_i \mathcal{E}_i \mathcal{F} = \mathcal{F}$, $\Gamma^{= \ast}$ in Equation 3.3. It is easy to show that $\succsim$ is a preorder on $A$, and $A \approx B \Rightarrow DR(A) = DR(B)$. It is known that Brewka’s preferred subtheories (PS) [3] Section 6) is a special case of PDL, and the argumentation semantics for PS uses $\subseteq \mathcal{E}_\mathcal{E}_i \mathcal{E}_i$ to calculate $\succsim$ [12]. Therefore, one might consider using $\subseteq \mathcal{E}_\mathcal{E}_i \mathcal{E}_i$ (Equations 4.2 and 4.3) for calculating $\succsim$. But $\subseteq \mathcal{E}_\mathcal{E}_i \mathcal{E}_i$ does not yield a correspondence with PDL as the following example illustrates.

Example 2. Consider $\langle D, W, \prec^+ \rangle$ where $W = \{ a \}$,

$$D = \left\{ \begin{array}{lcl} & d_1 := a : b \rightarrow c & d_2 := b : c \rightarrow c \rightarrow c \end{array} \right\} \quad (4.4)$$

and $d_1 \prec^+ d_2 \prec^+ d_3$. The extension is $Th(\{a, b, \sim c\})$.

In the ASPIC+ instantiation: $r_1 <_D r_2 <_D r_3 \quad (where \ for \ i = 1, 2, 3, r_i := f(d_i) \ and \ f \ is \ Equation \ 3.2 \). The arguments are $A := \{ [a \rightarrow b] \sim c \}$ and $B := \{ [[a] \rightarrow b] \sim c \}$, which rebut each other at their conclusions.

Under the elitist ordering (Equation 4.2), it is neither the case that $(r_1, r_2) \mathcal{E}_\mathcal{E}_i \mathcal{E}_i \{ r_1, r_2 \}$ nor $(r_1, r_3) \mathcal{E}_\mathcal{E}_i \mathcal{E}_i \{ r_1, r_2 \}$. As the sets are not equal, we have $A \nleq B$, but $A \nleq B$ and $A \nleq B$. This means $A \nleq B$, that means there are two possible stable extensions $\{ A \}$ and $\{ B \}$ so that neither argument is sceptically justified, and so $\sim c$ is not an argumentation-defined inference. However $\sim c$ is a PDL.

\footnote{It suffices to consider finite sets as arguments are finite.}
\footnote{It can be shown that this definition avoids counterexamples like Example 5.1, as explained in [17].}
inference. Therefore the elitist ordering cannot be used to calculate \( \preceq \).

We now investigate a modified elitist order. Suppose that in Example 2 we use the disjoint elitist order,

\[
\Gamma \upharpoonright_{DEF}, \Gamma' \iff (\exists x \in \Gamma - \Gamma') (\forall y \in \Gamma' - \Gamma) x <_{D} y ,
\]

with \( \upharpoonright_{DEF} \) replacing \( \upharpoonright_{SP} \) in Equations 4.2 and 4.3. Given \( r_2 <_{D} r_3 \), it is easy to see that \( A \nless B, B \nless A \), and so \( A \nangled R B \). Hence there is only a single stable extension containing the now sceptically justified argument \( B \) with conclusion \( \neg a \).

It seems very intuitive for the disjoint elitist order to ignore shared rules, because when deciding whether \( A \nless B \) or \( B \nless A \), we should only focus on the fallible information on which the arguments differ\(^{19}\). However, despite this intuitive motivation, the conclusions of the justified arguments given by the disjoint elitist order do not correspond to those obtained in PDL.

**Example 3.** Let \( \langle D, W, \prec + \rangle \) have \( D = \{ d_k \}_{k=1}^{14} \), \( W = \emptyset \),

\[
d_1 := \Gamma : c_1, \; d_2 := \Gamma : c_2, \; d_3 := \Gamma : c_3, \; d_5 := \Gamma : c_4 ,
\]

\[
d_6 := \Gamma : c_4 \land \neg (c_2 \land c_1), \; d_7 := \Gamma : c_4 \land \neg (c_2 \land c_1)
\]

such that \( d_1 \prec + d_2 \prec + d_3 \prec + d_4 \prec + d_5 \). Our PDE is constructed in the usual manner starting from \( E_0 = Th(\emptyset) \).

Equation 4.7 gives the order of application of the defaults:

\[
E_1 = E_0 + c_3, \; E_2 = E_1 + c_4,
\]

\[
E_3 = E_2 + c_1, \; E_4 = E_3 + \neg (c_2 \land c_4)
\]

with \( E_k = E_k \) for all \( k \geq 5 \). The default \( d_2 \) is blocked because \( \neg (c_2 \land c_4) \equiv (\neg c_2 \lor \neg c_4) \), and with \( c_4 \) (from \( d_4 \)), we have \( \neg c_2 \), which blocks \( d_2 \). The unique PDE from this LPDT is

\[
Th(\{c_1, \neg c_2, c_3, c_4\}) .
\]

Now consider the corresponding arguments following our instantiation. We have the defeasible rules\(^{20}\)

\[
r_1 <_{D} r_4 <_{D} r_3 <_{D} r_2 <_{D} r_5 .
\]

The relevant arguments and sets of defeasible rules are

\[
A := \{ \Gamma \Rightarrow c_1 \} \cup \{ c_2 \} , \; DR(A) = \{ r_1, r_2 \}
\]

\[
B := \{ \Gamma \Rightarrow c_1 \} \cup \{ c_4 \} , \; DR(B) = \{ r_3, r_4 \}
\]

\[
C := \{ \Gamma \Rightarrow c_1 \} \cup \{ c_4 \} , \; DR(C) = \{ r_1, r_3 \}
\]

\[
D := [B, C \Rightarrow \neg c_4] , \; DR(D) = \{ r_1, r_3, r_4, r_5 \} .
\]

For the correspondence to hold, the desired stable extension is \( \{ D, B, C, [\exists \Rightarrow c_4], [\Gamma \Rightarrow c_1] \} \) and all strict extensions thereof\(^{21}\) which does have a conclusion set corresponding to Equation 4.7. However, this would require \( D \leftarrow A \), which means, by Equation 4.7, \( D \rightarrow A \) and \( D \nless A \). Clearly, \( D \less A \) on \( A \). However, \( D \nless A \) is equivalent to, under the disjoint elitist order, that \( r_2 \) is \( <_{D} \)-least in \( \mathcal{R}_d \). From Equation 4.8 it is not the case that \( r_2 <_{D} r_1, r_3, r_4 \), so we conclude \( D <_{D} A \). Therefore, argumentation does not generate the corresponding stable extension to Equation 4.7.

Example \( ^{3} \) shows that we cannot use the disjoint elitist order to compare sets of defeasible rules because it ignores the structure of how arguments are constructed\(^{22}\). The way that PDL arranges the defaults in order of preference suggests a new way of defining an argument preference relation compatible with how the arguments are constructed. We now propose the structure-preference (SP) order. The idea is to transform \( <_{D} \) into a new strict total order, \( <_{SP} \), on \( \mathcal{R}_d \), such that it captures the original preference \( <_{D} \) and when the defeasible rules become applicable during the construction of the arguments.

Since we assumed \( \mathcal{R}_d \) is finite, let \( 1 \leq i \leq |\mathcal{R}_d| \). We define \( a_i \in \mathcal{R}_d \) to be the \( <_{D} \)-greatest element of the following set:

\[
\{ r \in \mathcal{R}_d | \text{Ante}(r) \subseteq \text{Conc} [ \bigcup_{j=1}^{i-1} \{ a_j \}] - \bigcup_{j=1}^{i-1} \{ a_j \} \} .
\]

(14.13)

The intuition is: \( a_i \) is the most preferred rule whose antecedent is amongst the conclusions of all strict arguments, \( a_2 \) is the next most preferred rule, whose antecedent is amongst the conclusions of all arguments having at most \( a_1 \) as a defeasible rule. Similarly, \( a_3 \) is the next most preferred rule, whose antecedent is amongst the conclusions of all arguments having at most \( a_1 \) and \( a_2 \) as defeasible rules, and so on until all of the rules of \( \mathcal{R}_d \) are exhausted. This process orders the rules by how preferred they are under \( <_{D} \) and by when they are applicable when constructing the arguments. Notice that the second union after the set difference in Equation 14.13 ensures that once a rule is applied it cannot be applied again. We then define \( <_{SP} \) as (notice the dual order)

\[
a_i <_{SP} a_j \iff j < i ,
\]

(14.14)

where \( 1 \leq i, j \leq |\mathcal{R}_d| \). We define the non-strict order to be \( a_i \leq_{SP} a_i \iff a_i = a_j \) or \( a_i <_{SP} a_j \). This makes sense because \( i \rightarrow a_i \) is bijective between \( \mathcal{R}_d \) and \( \{ 1, 2, 3, \ldots, |\mathcal{R}_d| \} \). Clearly \( <_{SP} \) is a strict total order on \( \mathcal{R}_d \). We call this the structure preference order on \( \mathcal{R}_d \), which exists and is unique given \( <_{D} \). We define the corresponding strict set comparison relation, \( <_{SP} \), as, for \( \Gamma', \Gamma' \subseteq \mathcal{R}_d \),

\[
\Gamma <_{SP} \Gamma' \iff (\exists x \in \Gamma - \Gamma') (\forall y \in \Gamma' - \Gamma) x <_{SP} y .
\]

(14.15)

The corresponding strict argument preference is

\[
A <_{SP} B \iff DR(A) <_{SP} DR(B) .
\]

(14.16)

We define the corresponding non-strict preference as \( A <_{SP} B \iff [DR(A) \forall x \in \mathcal{R}_d ] DR(B) \) or \( DR(A) = DR(B) \). This is the disjoint elitist order\(^{11}\) (Equation 4.5) with \( <_{D} \) specialised to \( <_{SP} \). The SP-order thus allows us to mimic how PDL applies defaults when calculating extensions.

**Lemma 4.1.** The preference \( <_{SP} \) satisfies

\[
(\forall A, B \in \mathcal{A}) [DR(A) \subseteq DR(B) \Rightarrow B <_{SP} A] .
\]

(14.17)

Informally, the “structure” of an argument \( A \) is given by the preordered set \( (\text{Sub}(A), <_{SP}) \).

\(^{11}\)We use the disjoint elitist order instead of the usual elitist order because Example \( ^{3} \) shows that the usual elitist order does not give the correspondence in general.
Proof. If $DR(B) = DR(A)$ then $B \approx A$, so $B \not\preceq_{SP} A$.
If $DR(A) \subset DR(B)$, then $DR(A) - DR(B) = \emptyset$, which means $B \prec_{SP} A$ is vacuously true from Equation 4.5, so $B \not\preceq_{SP} A$.

This is intuitive because if $A \sqsubseteq_{ag} B$, then $A$ may contain less fallible information (in our case defeasible rules) than $B$, so $A$ can be said to be more certain than $B$. It makes sense to have $B \not\preceq_{SP} A$ because rational agents should prefer more certainty to less certainty. It is easy to see that strict arguments are most preferred.

We now show that the disjoint elitist order $\triangleleft_{DEli}$ in general, and hence $\triangleleft_{SP}$ in particular, is a strict total order on $P_{fin}(R_d)$.

**Lemma 4.2.** If $(R_d, \prec_D)$ is a strict toset, then $\triangleleft_{DEli}$ is a strict total order on $P_{fin}(R_d)$.

**Proof.** Let $\Gamma$, $\Gamma'$, $\Gamma''$ be arbitrary. **Irreflexivity:** From Equation 4.5 as $\exists$ precedes $\forall$, $\triangleleft_{DEli} \Gamma$ is false.

**Transitivity:** (Sketch) Assume $\Gamma \triangleleft_{DEli} \Gamma'$ and $\Gamma' \triangleleft_{DEli} \Gamma''$. Let $n_1$, $n_2$, ..., $n_7 \in N$ be such that

$\Gamma = \{a_1, ..., a_{n_1}\} \cup \{d_1, ..., d_{n_2}\} \cup
\{f_1, ..., f_{n_3}\} \cup \{g_1, ..., g_{n_4}\}$

$\Gamma' = \{b_1, ..., b_{n_5}\} \cup \{d_1, ..., d_{n_6}\} \cup
\{e_1, ..., e_{n_7}\} \cup \{g_1, ..., g_{n_8}\}$

$\Gamma'' = \{c_1, ..., c_{n_9}\} \cup \{e_1, ..., e_{n_7}\} \cup
\{f_1, ..., f_{n_3}\} \cup \{g_1, ..., g_{n_4}\}$

where the a's to g's in $R_d$ denote distinct defeasible rules. The seven disjoint finite sets $\{a_1, ..., a_{n_1}\}$, $\{b_1, ..., b_{n_5}\}$ partition $\Gamma \cup \Gamma' \cup \Gamma''$ and can be represented by the subregions of three overlapping circles of the corresponding Venn diagram. If $n_1 = 0$ then the corresponding set is empty, e.g. if $n_1 = 0$ then $\{a_1, ..., a_{n_1}\} = \emptyset$.

Assuming $\Gamma \triangleleft_{DEli} \Gamma'$ and $\Gamma' \triangleleft_{DEli} \Gamma''$, we use Equation 4.5 to prove $\Gamma \triangleleft_{DEli} \Gamma''$, which is the equivalent of proving, for at least one of $1 \leq k \leq n_1$ or $1 \leq l \leq n_4$,

$$\left[ \bigwedge_{i=1}^{n_k} a_k \prec c_i \land \bigwedge_{j=1}^{n_l} a_k \prec e_j \right]$$

or

$$\left[ \bigwedge_{i=1}^{n_k} d_i \prec c_i \land \bigwedge_{j=1}^{n_l} d_i \prec e_j \right].$$

(4.17)

By writing out $\Gamma \triangleleft_{DEli} \Gamma'$ and $\Gamma' \triangleleft_{DEli} \Gamma''$ in terms of the elements $a_1, ..., a_{n_1}$ similar to how $\Gamma \triangleleft_{DEli} \Gamma''$ is written in Equation 4.17 we get four cases (as "and" and "or" biddistributive). One case gives a contradiction (due to irreflexivity of $\prec_D$), while the other three cases imply $\Gamma \triangleleft_{DEli} \Gamma''$ from the totality of $\prec_D$. Therefore, $\Gamma \triangleleft_{DEli} \Gamma''$.

**Trichotomy:** Assume $\Gamma \not\preceq_{DEli} \Gamma'$ and consider $\Gamma \triangleright_{DEli} \Gamma'$ (as $\triangleright_{DEli} \Gamma'$ is a strict toset, and thus has a $\prec_{DEli}$-least element $m$. Either $m \in \Gamma - \Gamma'$, exclusive-or $m \in \Gamma - \Gamma'$, where the former implies $\Gamma \not\triangleright_{DEli} \Gamma'$ and the latter implies $\Gamma\not\prec_{DEli} \Gamma'$. It follows that $\triangleleft_{SP}$ is a strict total preorder on $A$.

**Example 4.** (Example 3 continued) By applying Equations 4.13 and 4.14 we can show that $a_1 = r_3$, $a_2 = r_4$, $a_3 = r_1$, $a_4 = r_5$ and $a_5 = r_2$. The structure preference order is

$r_2 \prec_{SP} r_5 \prec_{SP} r_1 \prec_{SP} r_4 \prec_{SP} r_3$.  

(4.18)

Notice that this is precisely the order in which the corresponding normal defaults are added in PDL, as Equation 4.16 shows. It is easy to show that the corresponding stable extension under the argument preference $\prec_{SP}$ corresponds to the PDL inference, because $r_2$ is now $\prec_{SP}$-least, so $D \not\prec_{SP} A$.

Therefore $\prec_{SP}$ does not necessarily follow the PDL order of the application of defaults as the following example illustrates.

**Example 5.** Consider $\{\{d_1, d_2\}, \{a\}, \prec^+\}$ with $d_1 := \frac{a}{a}$ and $d_2 := \frac{b}{c}$ such that $d_2 \prec^+ d_1$. We have $E = Th\{(a, b)\}$, where $d_1$ is blocked by $W$, so $d_2$ is the only default added. In argumentation, we have $K_A = \{a\}$, $r_1 := (a \Rightarrow \neg a)$ and $r_2 := (\top \Rightarrow b)$ where for $i = 1, 2$, $r_i := f(d_i)$, such that $r_2 \not\prec_D r_1$. The arguments are $A_0 := \{a\}$, $A_1 := \{a \Rightarrow \neg a\}$ and $B := \{\top \Rightarrow b\}$. Applying Equation 4.13 we have $r_2 \prec_{SP} r_1$, which clearly is not the order of how the corresponding defaults are added in PDL. Yet the correspondence still holds, since $A_0 \not\prec A_1$ because $A_0$ is strict and strict arguments always defeat any non-strict argument they attack, so the stable extension is the strict extension of $(A_0, B)$, the conclusion set of which (after deductive closure) is the extension of the underlying LPDT.

Example 4 highlights how blocked defaults and defeated arguments are related. Where PDL blocks the application of a given default, hence preventing its conclusion from featuring in the extension, ASPIC* allows for the construction of the argument with the corresponding defeasible rule, but that argument is always defeated by another strictly stronger argument and therefore cannot be in any extension.

5. **THE REPRESENTATION THEOREM**

In this section we state and prove the representation theorem (Theorem 4.3), which guarantees that the inferences with argumentation semantics under the preference $\not\prec_{SP}$ correspond exactly to the inferences in PDL.

5.1 **Non-Blocked Defaults**

We first introduce some concepts to help prove the representation theorem. Let $\langle D, W, \prec \rangle$ be a PDT and $E = \bigcup_{i \in E} E_i$ one of its extensions generated from the linearisation $\prec^+ \preceq$. The set of generating defaults (with respect to $\prec^+$), $GD(\prec^+)$, is defined as

$$GD(\prec^+) := \{d \in D | d \text{ is } \prec^+-\text{greatest active in } E_i\},$$

$$GD(\prec^+) := \bigcup_{i \in N} GD_i(\prec^+) \subseteq D.$$  

(5.1)

Intuitively, this is the set of defaults applied to calculate $E$ following the order $\prec^+$. However, the same $E$ can be generated by distinct total orders.

**Example 6.** Consider the PDT $\langle \{\frac{a}{b}, \frac{b}{c}\}, \{a, b\}, \emptyset \rangle$. We have two possible linearisations $\frac{a}{c} \prec^+ \frac{b}{c}$ and $\frac{b}{c} \prec^+ \frac{a}{c}$. By Footnote 4 we have $GD(\frac{a}{c}) = \{ \frac{a}{c} \}$ and $GD(\frac{b}{c}) = \{ \frac{b}{c} \}$, which are not equal, even though both linearisations give the same extension $E = Th\{(a, b, c)\}$. But in the case of $\frac{a}{c}, \frac{b}{c}$ is not active not because it is blocked by $\not\prec$, but rather it adds no new information.

We wish to distinguish between inactive defaults that conflict with something we already know, and inactive defaults
that do not add any new information. We call a default \( \frac{\phi}{\psi} \) semi-active (in \( S \subseteq SL \)) iff \( \{ \theta \in S, \neg \phi \notin S, \phi \in S \} \). The set of semi-active defaults with respect to the linearisation \( \prec^+ \) is

\[
SAD(\prec^+) := \{ d \in D \mid d \text{ is semi-active w.r.t. } \prec^+ \}.
\]  

(5.2)

Intuitively, semi-active defaults add no new information. We then define the set of non-blocked defaults to be

\[
NBD(\prec^+) := GD(\prec^+) \cup SAD(\prec^+) \subseteq D.
\]

(5.3)

\( NBD \) has a more elegant characterisation:

**Lemma 5.1.** If \( \prec^+ \) generates the PDE \( E \), then we have

\[
NBD(\prec^+) := \left\{ \left( \frac{\phi}{\psi} \right) \in D \mid \phi \in E, \neg \psi \notin E \right\}.
\]

(5.4)

**Proof.** (Sketch) It is sufficient to show Equation 5.3 (with Equations 5.1 and 5.2) is the same as the right hand side of Equation 5.4. Let \( E \) be the extension from \( \prec^+ \). For readability we will omit writing “\( \prec^+ \)” from “\( GD(\prec^+) \)”. It can be shown that \( d := \left( \frac{\phi}{\psi} \right) \in GD(\prec^+) \Rightarrow \phi \in E, \neg \psi \notin E \), and assuming \( \neg \phi \notin E \) gives a contradiction (by considering the \( E_i \)’s in \( E \)), so \( \neg \phi \notin E \). Trivially, \( d \in SD(\prec^+) \Rightarrow \phi \in E, \neg \psi \notin E \), therefore Equation 5.3 is a subset of Equation 5.4. Assuming that \( d \) is in the right hand side of Equation 5.4 gives \( d \in GD(\prec^+) \cup SAD(\prec^+) \) through simple manipulations. The result follows.

Given \( E \), \( NBD(E) \) is uniquely determined, so we will write \( NBD(E) \) instead. Equation 5.4 adapts Reiter’s idea of a generating default [10] page 92 Definition 2] to PDL.

### 5.2 Uniqueness of Stable Extensions

In this section we show that the defeat graph \( (A, \rightarrow) \) associated with any ASPIC\(^+\) attack graph \( (A, \rightarrow, \lesssim_{SP}) \) constructed from a LPDT \( (D, W, \prec^+) \) with the SP-order has a unique stable extension.

**Theorem 5.2.** Let \( (A, \rightarrow, \lesssim_{SP}) \) be an ASPIC\(^+\) attack graph constructed from a LPDT \( (D, W, \prec^+) \) with the SP-order. The defeat graph \( (A, \rightarrow) \) has a unique stable extension.

**Proof.** (Sketch) The construction of the unique stable extension imitates how extensions are constructed over an LPDT (Equation 2.4). Given a set of arguments \( S \subseteq A \) we define, for \( r \in R_d \), \( S \oplus r := \text{Args}(DR(S) \cup \{ r \}) \), i.e. we close \( S \) under all arguments with the addition of a new defeasible rule \( r \). Now consider Algorithm 1 which takes an attack graph \( (A, \rightarrow, \lesssim_{SP}) \) obeying the conditions of the theorem, and outputs a set of arguments.

**Algorithm 1 Generating a Stable Extension**

1. function GENERATESTABLEEXTENSION((A, \( \rightarrow, \lesssim_{SP} \))
2. \( S := \{ \text{all strict arguments in } A \} \)
3. for \( r \in R_d \) from \( \lesssim_{SP}\text{-greatest to } \lesssim_{SP}\text{-smallest} \) do
4. \( \text{if } S \oplus r \text{ has no attacks, } (S \oplus r)^\cap \rightarrow = \emptyset, \text{ then} \)
5. \( S := S \oplus r \)
6. return \( S \)

The intuition of Algorithm 1 is to first create the largest possible set of undefeated arguments that do not attack each other, first by including all strict arguments because strict arguments are never defeated (Line 2 recall also Example 2) and never attack each other because \( K_n \) is consistent. Then, the algorithm includes the defeasible rules from most to least preferred and tests whether the resulting arguments that are constructed by the inclusion of such a defeasible rule attack each other (Lines 1-3). As \( \lesssim_{SP} \) is total, all defeasible rules are considered, and the result includes as many defeasible rules as possible such that the result is consistent. Adding the rules in the order of \( \lesssim_{SP} \) while ensuring conflict freeness mimics the condition of Equation 2.3.

It is clear from the algorithm that \( S \) exists and is unique, as it is of the form \( \text{Args}(R) \) for some \( R \subseteq R_d \). We show \( S \) is a stable extension [2 page 26 Definition 2.2.7]: cf is guaranteed by the consistency of \( K_n \) and that defeasible rules \( r \in R_d \) are only added if the resulting arguments do not attack each other. Therefore, \( S \) contains no defeats and must be cf. To show that the arguments of \( S \) defeat all arguments not in \( S \), let \( R := DR(S) \), i.e. the set of all defeasible rules added to \( S \). Let \( B \notin S \) be arbitrary. We find an \( A \in S \) such that \( A \rightarrow B \). Given that \( B \notin S \), there is some rule \( r \in DR(B) \) - \( R \) that causes \( S \) to attack the subargument of \( B \) with top rule \( r \) if \( r \) is included, according to Algorithm 1 Line 4. Let \( B' \subseteq_{sp} B \) such that TopRule(\( B' \)) = \( r \). Let \( A \in S \) be the attacker of \( B' \) at \( r \). If \( r \) is \( \lesssim_{SP}\text{-greatest}, \) then \( \text{Args}(\( B \)) \oplus r \) contains attacking arguments, so \( A \) must be strict and hence \( A \rightarrow B \). If \( r \) is not \( \lesssim_{SP}\text{-greatest}, \) then consider the strict setup of \( r \) in \( (R_d, \lesssim_{SP}) \), \( T := \{ r' \in R_d \mid r \lesssim_{SP} r' \} \neq \emptyset \). If \( T \cap R = \emptyset \), then adding \( r \) to \( S \) means there is an attack from \( A \in \text{Args}(\( B \)) \) under \( \lesssim_{SP} \), and hence \( B \) is defeated. If \( T \cap R \neq \emptyset \), then \( A \in \text{Args}(\( T \cap R \)) \). Assume \( A \) is not strict, then \( B \notin DR(A) \subseteq T \cap R \), so \( \forall s \in DR(A) \) \( r \lesssim_{SP} s \), hence \( A \) defeats \( B \). Therefore, in all cases, \( A \rightarrow B \), and hence \( B \) is defeated by some argument in \( S \). Therefore, the defeat graphs of each ASPIC\(^+\) attack graphs have a unique stable extension.

### 5.3 The Representation Theorem

In this section we state and prove the representation theorem. This shows that inferences made in PDL correspond exactly to the conclusions of the justified arguments in the argumentation semantics of PDL by relating the stable extension of \( (A, \rightarrow) \) with the extension of the corresponding LPDT \( (D, W, \prec^+) \).

**Theorem 5.3.** Let \( (A, \rightarrow, \lesssim_{SP}) \) be the attack graph corresponding to an LPDT \( (D, W, \prec^+) \) with defeat graph \( (A, \rightarrow) \) under \( \lesssim_{SP} \).

1. Let \( E \) be the extension of \( (D, W, \prec^+) \). Then there exists a unique stable extension \( E \subseteq A \) of \( (A, \rightarrow) \) such that Conc\((E) = E \).

2. Let \( E \subseteq A \) be the unique stable extension of \( (A, \rightarrow) \) by Theorem 5.2 then Conc\((E) \) is the extension of the corresponding LPDT \( (D, W, \prec^+) \).

**Proof.** (Sketch) To prove the first statement we construct \( E \) in terms of \( E \) and show \( E \) is a stable extension of \( (A, \rightarrow) \) (which is unique given Theorem 5.2), and show \( \text{Conc}(E) = E \). Given \( E \), we construct \( E := \text{Args}(\{ NBD(E) \}) \). This set is unique from the properties of \( \text{Args} \). Then we show this \( E \) is a stable extension, which means \( E \) is cf and defeats all arguments not belonging to it. Assume for contradiction that \( E \) is not cf, which means there are arguments \( A, B \in E \) such that \( A \rightarrow B \), which means \( A \rightarrow B \). Let \( a := \text{Conc}(A) \), and as \( A \in \),

\[ ]
DR(A) ⊆ f (NBDE(E)) and hence a ∈ E. Now let B′ ⊆ arg B be the argument such that TopRule(B′) = (b ⇒ ¬a) for some appropriate formula b in B. As B ∈ E, this means (b ⇒ ¬a) ∈ f (NBDE(E)) and hence a ∉ E by Equation 5.4 and that E is deductively closed – contradiction. Therefore, E is cf. To show Arg(R) defeats all other arguments, let B ∉ Arg(R) so there is some rule r ∈ DR(B) – R. Let B′ ⊆ arg B be such that TopRule(B′) = r, r corresponds to a default f⁻¹(r) = 2Φ ∉ NBDE(E). Either θ ∉ E or ¬φ ∈ E by Equation 5.4. If ¬φ ∈ E, then one can prove there exists an argument A ∈ Arg(R) such that A → B′ and hence A → B under ≤sp. Assume θ ∉ E. There is some B″ ⊆ arg B′ such that Conc(B′) = θ. If θ ∉ E then B″ is neither strict nor in Arg(f (NBDE(E))). Thus there is some other s ∈ DR(B″) – f (NBDE(E)). We can repeat the above reasoning for s but not indefinitely as arguments are well-founded. We will end up with either a strict subargument of B″ or an argument in Arg(f (NBDE(E))). Therefore, θ ∈ E. Therefore, the only reason for r ∉ R is because ¬φ ∈ E, and hence there is an argument A that defeats any argument containing the rule r, which means Arg(R) defeats all other arguments and hence it is a stable extension. To show that E = Conc(E), we show E ⊆ Conc(E) and Conc(E) ⊆ E. In the first case, let θ ∈ E, then if θ ∈ E₀, we have W |= θ and by the compactness theorem in FOL, we have ∆ ≤f₃ W such that ∆ |= θ. From this we build a strict argument A such that Prem₃(A) = ∆ and Conc₃(A) = θ, and necessarily A ∈ E so θ ∈ Conc(E). Similarly, if θ ∈ E₁ for some k ∈ N⁺, we can construct a defeasible argument A concluding θ such that DR(A) ⊆ f (NBDE(E)) and hence A ∈ E, so θ ∈ Conc(E). Conversely, if θ ∈ Conc(E) there is an argument in E concluding θ. If this argument is strict then θ ∈ E₀ ⊆ E, else, as the defeasible rules are in f (NBDE(E)), then θ ∈ E₂ ⊆ E for some k ∈ N⁺ that indicates when all of the appropriate defaults needed to conclude θ are included. This establishes the first statement.

For the second statement, we show Conc(E) ⊆ E and E ⊆ Conc(E). For the former, if θ ∈ Conc(E) then there is some A ∈ E concluding θ. If A is strict then θ ∈ E₀ ⊆ E. If A is defeasible, then say DR(A) = \{d_i\} for some k ∈ N⁺. All of these defaults do not introduce any inconsistency because E is stable and hence cf. Take the smallest i ∈ N such that sufficiently many corresponding defeasible rules are applied from DR(A) to conclude θ in E₀⁺ from W. Therefore, θ ∈ E₀⁺ ⊆ E. Conversely, let θ ∈ E, so there is some i ∈ N such that θ ∈ Eᵢ. If i = 0, then there is a strict argument A, necessarily in E, that concludes θ so θ ∈ Conc(E). If i > 0, then we can use the appropriate defeasible rules to construct a defeasible argument A such that Prem₃(A) ⊆ W, Conc₃(A) = θ and DR(A) ≠ ∅. To show A ∈ E, we assume for contradiction that A ∉ E so there is some B ∈ E such that B ↩ A. But this would result in at least one of the defaults of A being blocked. This is a contradiction because given that θ ∈ E, the defeasible argument A concluding θ cannot have its defeasible rules correspond to blocked PDL defaults. Therefore, such a B cannot exist and A ∈ E. This proves that Conc(E) = E.

The representation theorem means that PDL is sound and complete with respect to its argumentation semantics. We now show that this ASPIC⁺ instantiation to PDL satisfies the Caminada-Amgould rationality postulates (Section 2.2.2) as a corollary to the representation theorem. Recall that when instantiated to FOL, Clₑ₂ becomes deductive closure.

**Theorem 5.4.** Let \( \langle A, \rightarrow, \leq_{SP} \rangle \) be the ASPIC⁺ attack graph of PDL and let E be any of the admissible extensions of the corresponding defeat graph \( \langle A, \rightarrow \rangle \). Our instantiation satisfies the Caminada-Amgould rationality postulates.

**Proof.** By Theorem 5.2 \( \langle A, \rightarrow \rangle \) has a unique stable extension \( E \), which is a complete and an admissible extension. It is sufficient to prove the postulates for \( E \) because \( \langle A, \rightarrow \rangle \) only has \( E \) as its sole complete extension. (1) To show that \( E \) is subargument closed, recall that Algorithm 1 gives an explicit construction of \( E \), which is of the form \( \text{Arg}(R) \) for some \( R \subseteq \mathcal{R}_d \) and is clearly subargument closed. (2) The representation theorem states that \( \text{Conc}(E) = E \) and as \( E \) is deductively closed, \( \text{Conc}(E) \) is closed under strict rules. (3) As \( W \) is consistent and \( \text{Conc}(E) \) is the extension, \( \text{Conc}(E) \) must also be consistent and its deductive closure is consistent.

### 6. DISCUSSION AND CONCLUSION

We have endowed PDL with argumentation semantics using ASPIC⁺. This is achieved by specialising ASPIC⁺ to PDL (Section 3), discussing which preferences can be suitable for the correspondence of inferences (Section 4), proving that the inferences do correspond (Theorem 5.2) and that this instantiation is normatively rational (Theorem 5.4). As explained in Section 1, this allows us to interpret the inferences of PDL as conclusions of justified arguments, clarifying the reasons for accepting or rejecting a conclusion. Further, this makes the inference process more intuitive, and amenable to human participation and inspection. The argumentative characterisation of PDL provides for distributed reasoning in the course of deliberation and persuasion dialogues. For example, BOID agents with PDL representations of mental attitudes can now exchange arguments and counterarguments when deliberating about which goals to select, and consequently which actions to pursue (Example 1).

However, it seems that we have restricted our attention to LPDTs. This does not lose generality because calculating extensions in PDL always presupposes a linearisation of \( \prec \) (see 4 or recall Section 2.3), and Theorem 5.4 shows that for any linearisation the correspondence between PDL and its argumentation semantics holds. However, ASPIC⁺ can identify argumentation-based inferences assuming only a partial ordering, unlike in PDL. This suggests that our argumentative characterisation can be used to generalise PDL; for example, under a partial ordering one might not only generate multiple stable extensions, but extensions under other Dung semantic may become relevant. Future work will look at to what extent we can lift the requirement of linearity, as well as the significance of other Dung semantics.

Further, one aspect of ASPIC⁺ that has not been discussed in much detail here is how normative rationality (in the sense of 7) follows for any well-defined instantiation with a reasonable preference (13 Definitions 12 and 18). We have shown that our instantiation is normatively rational via a direct proof, but it is worthwhile strengthening this result by asking whether the instantiation we have presented here is well-defined and whether \( \leq_{SP} \) is reasonable. This would also allow us to investigate generalisations and variations of PDL via its argumentation semantics.

---

17 When the extension is unique, the distinction between the different Dung semantic types is lost.
REFERENCES


