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) Δ in some non-monotonic logic (NML), the arguments and attacks defined by Δ 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 Δ. More formally, given an argumentation framework AF and a wff θ, the *argumentation-defined inference relation* $\mid\!\sim_{AF}$ over Δ is $\Delta \mid\!\sim_{AF} \theta$ iff θ is the conclusion of a sceptically justified argument in AF. Indeed, a correspondence has been shown between $\mid\!\sim_{AF}$ over Δ, and the instantiating logic's non-monotonic inference relation defined directly over Δ. For example, default logic (DL) [8], logic programming [8], defeasible logic [10] and preferred subtheories (PS) [13] have all been 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: ":=" means "is defined as". $\mathbb{N}$ 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 $\mathcal{P}_{\text{fin}}(X)$. $X \subseteq_{\text{fin}} Y$ iff X is a finite subset of Y, therefore $X \in \mathcal{P}_{\text{fin}}(Y) \Leftrightarrow X \subseteq_{\text{fin}} Y$. Undefined quantities are denoted by $*$, for example $1/0 = *$

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

in the real numbers. If $\langle P, \leq\rangle$ is a preordered set then the strict version of the preorder is $a < b \Leftrightarrow [a \leq b,\ b \not\leq a]$, which is also a strict partial order. For two sets A, B, $A \ominus B := (A - B) \cup (B - A)$ denotes their symmetric difference.

2.1 Dung's Abstract Argumentation Theory

We now recap the key definitions of [8]. An *argumentation framework* is a directed graph $\langle \mathcal{A}, \mathcal{C}\rangle$, where $\mathcal{A}$ is the *set of arguments* and $\mathcal{C} \subseteq \mathcal{A}^2$ is the *conflict relation* on $\mathcal{A}$. For arguments $A, B \in \mathcal{A}$ we write $\mathcal{C}(A, B) \Leftrightarrow (A, B) \in \mathcal{C} \Leftrightarrow A$ conflicts with B, i.e. A is a counterargument against B.

In what follows let $S \subseteq \mathcal{A}$ and $A, B \in \mathcal{A}$. S *conflicts with* B iff $(\exists A \in S)\, \mathcal{C}(A, B)$. S is *conflict-free* (cf) iff $\mathcal{C} \cap S^2 = \varnothing$. S *defends* A iff $(\forall B \in \mathcal{A})\, [\mathcal{C}(B, A) \Rightarrow S$ conflicts with $B]$. Let $Def(S) := \{A \in \mathcal{A} \mid S \text{ defends } A\} \subseteq \mathcal{A}$. 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* iff S is the $\subseteq$-least complete extension; a *stable extension* iff S is complete and conflicts with all arguments $B \in \mathcal{A} - S$.

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

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 and preferences, satisfy the rationality postulates of [7], and are hence *normatively rational*.

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

[1] ASPIC stands for *Argumentation Service Platform with Integrated Components*.

In ASPIC$^+$, a *knowledge base* is a set $\mathcal{K} := \mathcal{K}_n \cup \mathcal{K}_p \subseteq \mathcal{L}$ where $\mathcal{K}_n$ is the set of *axioms* and $\mathcal{K}_p$ 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) $[\theta]$ is a *singleton argument* with $\theta \in \mathcal{K}$, conclusion $Conc([\theta]) := \theta$, premise set $Prem([\theta]) := \{\theta\} \subseteq \mathcal{K}$, top rule $TopRule([\theta]) := *$ and set of subarguments is $Sub([\theta]) := \{[\theta]\}$.
2. (Inductive, strict) Let $A_1, \ldots, A_n$ be arguments with respective conclusions $Conc(A_1), \ldots, Conc(A_n)$ and premise sets $Prem(A_1), \ldots, Prem(A_n)$. If there is a strict rule $r := (Conc(A_1), \ldots, Conc(A_n) \to \phi) \in \mathcal{R}_s$, then $B := [A_1, \ldots, A_n \to \phi]$ is also an argument with $Conc(B) = \phi$, premises $Prem(B) := \bigcup_{i=1}^n Prem(A_i)$, $TopRule(B) = r \in \mathcal{R}_s$ and set of subarguments $Sub(B) := \{B\} \cup \bigcup_{i=1}^n Sub(A_i)$.
3. (Inductive, defeasible) Let $A_1, \ldots, A_n$ be arguments with respective conclusions $Conc(A_1), \ldots, Conc(A_n)$ and premise sets $Prem(A_1), \ldots, Prem(A_n)$. If there is a defeasible rule $r' := (Conc(A_1), \ldots, Conc(A_n) \Rightarrow \phi) \in \mathcal{R}_d$, then $C := [A_1, \ldots, A_n \Rightarrow \phi]$ is an argument with $Conc(C) = \phi$, premises $Prem(C) := \bigcup_{i=1}^n Prem(A_i)$, $TopRule(C) = r' \in \mathcal{R}_d$ and set of subarguments $Sub(C) := \{C\} \cup \bigcup_{i=1}^n Sub(A_i)$.

Let $\mathcal{A}$ 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 $A \in Sub(B)$ and we write $A \subseteq_{\text{arg}} B$. We say A is a *proper subargument* of B iff $A \in Sub(B) - \{B\}$ and we write $A \subset_{\text{arg}} B$. It can be shown that $\subseteq_{\text{arg}}$ is at least a preorder on $Sub(B)$.

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

An argument A *attacks* another argument B, denoted as $A \rightharpoonup B$, iff at least one of the following hold, where:

1. A is said to *undermine* attack B on the subargument $B' = [\phi]$ iff $[\exists \phi \in Prem(B) \cap \mathcal{K}_p]\ Conc(A) \in \overline{\phi}$, i.e. A conflicts with some ordinary premise of B.
2. There is some $B' \subseteq_{\text{arg}} B$ such that $r := TopRule(B') \in \mathcal{R}_d$, $\phi := Cons(r)$ and $Conc(A) \in \overline{\phi}$. A is then said to *rebut* attack B on the subargument B'.
3. There is some $B' \subseteq_{\text{arg}} B$ such that $r := TopRule(B') \in \mathcal{R}_d$ and $Conc(A) \in \overline{n(r)}$. A is then said to *undercut* attack B on the subargument B' (by arguing against the application of the defeasible rule r in B).

See [13, Section 2] for a further discussion of why attacks are distinguished in this way. We abuse notation to define the *attack relation* as $\rightharpoonup \subseteq \mathcal{A}^2$ such that $(A, B) \in \rightharpoonup \Leftrightarrow A \rightharpoonup B$.

A preference relation over arguments is then used to determine which attacks succeed as defeats. We denote the preference $\precsim \subseteq \mathcal{A}^2$ (not necessarily a preorder for now) such that $A \precsim B \Leftrightarrow A$ is not more preferred than B. The strict version is $A \prec B \Leftrightarrow [A \precsim B,\ B \not\precsim A]$, and equivalence is $A \approx B \Leftrightarrow [A \precsim B,\ B \precsim A]$. We define a *defeat* as

$$A \hookrightarrow B \Leftrightarrow \left(\exists B' \subseteq_{\text{arg}} B\right)\left[A \rightharpoonup B',\ A \not\prec B'\right] . \tag{2.1}$$

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 $\hookrightarrow \subseteq \mathcal{A}^2$ such that $(A, B) \in \hookrightarrow \Leftrightarrow A \hookrightarrow B$. A set of arguments $S \subseteq \mathcal{A}$ is *conflict-free* (cf) iff $\hookrightarrow \cap S^2 = \varnothing$.[2]

Preferences between arguments are calculated from the argument structure by endowing $\mathcal{K}_p$ and $\mathcal{R}_d$ with preorders $\leq_K$ and $\leq_D$ respectively, where (e.g.) $r_1 \leq_D r_2$ iff r_2 is *more preferred* than r_1 (and analogously for $\leq_K$). This preorder is then lifted to a set-comparison order $\trianglelefteq$ between the sets of premises or defeasible rules of the arguments, and then finally to $\precsim$, following the method in [13, Section 5]. We will summarise this in Section 4.

Given the preference relation $\precsim$ between arguments, we call the structure $\langle \mathcal{A}, \rightharpoonup, \precsim \rangle$ an *ASPIC*$^+$ *SAF* (structured argumentation framework), or *attack graph*. Its corresponding *defeat graph* is $\langle \mathcal{A}, \hookrightarrow \rangle$, where $\hookrightarrow$ is defined in terms of $\rightharpoonup$ and $\precsim$ as in Equation 2.1.

Given $\langle \mathcal{A}, \hookrightarrow \rangle$ one can then evaluate the extensions under Dung's semantics (Section 2.1) where $\mathcal{C}$ is $\hookrightarrow$, 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* $\mid\!\sim_{AS}$ is $\mathcal{K} \mid\!\sim_{AS} \theta$ iff $\theta = Conc(A)$ where $A \in \mathcal{A}$ is a sceptically justified argument.

2.2.2 *Normative Rationality*

Instantiations of ASPIC$^+$ should satisfy some properties to ensure it is rational [7]. Given an instantiation let $\langle \mathcal{A}, \rightharpoonup, \precsim \rangle$ be its ASPIC$^+$ attack graph with corresponding defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$. For $S \subseteq \mathcal{A}$ let $Conc(S) := \bigcup_{A \in S} Conc(A)$. Let $\mathcal{E}$ be any of its admissible extensions. The *Caminada-Amgoud rationality postulates* state:

1. If $\mathcal{E}$ is a complete extension then $\mathcal{E}$ is subargument closed.
2. If $\mathcal{E}$ is a complete extension then
$$Cl_{\mathcal{R}_s}\left[Conc(\mathcal{E})\right] = Conc(\mathcal{E}) . \tag{2.2}$$
3. The sets $Conc(\mathcal{E})$, $Cl_{\mathcal{R}_s}\left[Conc(\mathcal{E})\right] \subseteq \mathcal{L}$ 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 5.4. ASPIC$^+$ also identifies sufficient conditions for an instantiation to satisfy these postulates [13, Section 4], which we will discuss in Section 6.

[2] 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 $\mathcal{FL}$ and the set of closed first order formulae[3] is $\mathcal{SL} \subseteq \mathcal{FL}$, with the usual quantifiers and connectives. Given $S \subseteq \mathcal{FL}$, the *deductive closure of* S is $Th(S)$, and given $\theta \in \mathcal{FL}$, the *addition operator* $+ : \mathcal{P}(\mathcal{FL}) \times \mathcal{FL} \to \mathcal{P}(\mathcal{FL})$ is defined as $S + \theta := Th(S \cup \{\theta\})$.

A *normal default* is an expression $\frac{\theta : \phi}{\phi}$ where $\theta, \phi \in \mathcal{FL}$ and read "if θ is the case and ϕ is consistent with what we know, then ϕ is the case". In this case we call θ the *antecedent* and ϕ the *consequent*. A normal default $\frac{\theta : \phi}{\phi}$ is *closed* iff $\theta, \phi \in \mathcal{SL}$. We will assume all defaults are closed and normal unless stated otherwise. Given $S \subseteq \mathcal{SL}$, a default is *active (in S)* iff $[\theta \in S,\ \phi \notin S,\ \neg\phi \notin S]$. 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 $\langle D, W, \prec \rangle$, where $W \subseteq \mathcal{SL}$ is not necessarily a finite set and $\langle D, \prec \rangle$ is a *finite* strict poset (partially ordered set) of defaults, where $d' \prec d \Leftrightarrow d$ is *more*[4] *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 $\prec^+ \supseteq \prec$ be a linearisation[5] of $\prec$. A *prioritised default extension (with respect to* $\prec^+$*)* (PDE) is a set $E := \bigcup_{i \in \mathbb{N}} E_i \subseteq \mathcal{SL}$ built inductively as:

$$E_0 := Th(W) \text{ and} \tag{2.3}$$

$$E_{i+1} := \begin{cases} E_i + \phi\ , & \text{if property 1} \\ E_i\ , & \text{else} \end{cases} \tag{2.4}$$

where "property 1" iff "ϕ 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 $E_0 = E = \mathcal{FL}$.

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 $\prec$. We say the PDT $\langle D, W, \prec \rangle$ *sceptically infers* $\theta \in \mathcal{SL}$ iff $\theta \in E$ for *all* extensions E.

Henceforth, we will refer to a PDT $\langle D, W, \prec \rangle$ where $\prec$ is a strict total order as a *linearised PDT* (LPDT). If $\prec$ 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

[3] i.e. first order formulae without free variables

[4] We have defined the order dually to [4] so as to comply with orderings over the ASPIC$^+$ defeasible inference rules.

[5] i.e. $\prec^+$ is a strict total order and hence $\langle D, \prec^+ \rangle$ 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 $\frac{R(a):A(a)}{A(a)}$. She does not want to teach and would rather focus on her research, so her set of desires Des include $\frac{R(a):\neg T(a)}{\neg T(a)}$. However, she is obliged to teach, so her set of obligations Obl include $\frac{A(a):T(a)}{T(a)}$. The set of defaults is $D = Bel \cup Des \cup Obl$, and we assume no other defaults are relevant for this example.

In [6], 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.[6]

1. Our arguments are expressed in FOL, so our set of wffs is $\mathcal{FL}$.

2. The contrary function $-$ syntactically defines conflict in terms of classical negation.[7] Let θ, ϕ be wffs, then $\overline{\theta} = \phi$ if θ is of the form $\neg\phi$; else $\phi = \neg\theta$.

3. The set of strict rules $\mathcal{R}_s$ characterises inference in first order classical logic. We leave the proof theory implicit. $Cl_{\mathcal{R}_s}$ instantiates to deductive closure.

4. The set of defeasible rules $\mathcal{R}_d$ is defined as:
$$\mathcal{R}_d := \left\{(\theta \Rightarrow \phi)\middle|\frac{\theta : \phi}{\phi} \in D\right\}\ , \tag{3.1}$$
with $n \equiv *$. Clearly, there is a bijection[8] f where
$$f : D \to \mathcal{R}_d : \frac{\theta : \phi}{\phi} \mapsto f\left(\frac{\theta : \phi}{\phi}\right) := (\theta \Rightarrow \phi) \tag{3.2}$$
and we will define the *strict version of the* preorder $\leq_D$ over $\mathcal{R}_d$ as[9]
$$(\theta \Rightarrow \phi) <_D (\theta' \Rightarrow \phi') \Leftrightarrow \frac{\theta : \phi}{\phi} \prec^+ \frac{\theta' : \phi'}{\phi'}\ . \tag{3.3}$$
We can see that the strict toset $\langle \mathcal{R}_d,\ <_D\rangle$ is order isomorphic to $\langle D,\ \prec^+\rangle$, where the non-strict version of the order $\leq_D$ is [$<_D$ or equality].

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

The set $\mathcal{A}$ of ASPIC$^+$ arguments are defined as in Section 2.2. It is easy to see that all arguments are firm because $\mathcal{K}_p = \varnothing$, and so there are no undermining attacks. As n is undefined, no attack can be an undercut. Therefore, we only have rebut attacks, where $A \rightharpoonup B$ iff
$$\left(\exists B',\ B'' \subseteq_{\text{arg}} B\right)\ B' = \left[B'' \Rightarrow \overline{Conc(A)}\right]\ . \tag{3.4}$$
Defeats are defined as in Equation 2.1. This leaves the question as to how the argument preference $\precsim$ should be defined based on the strict total order $<_D$ over $\mathcal{R}_d$.

4. PREFERENCES

We want to define $\precsim$ 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 $\langle \mathcal{A},\ \hookrightarrow\rangle$, 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 $\precsim$.

More formally, ASPIC$^+$ defines the *elitist order* [13, Section 5], where for $A,\ B \in \mathcal{A}$ and $DR(A),\ DR(B) \subseteq \mathcal{R}_d$, the argument preference $\precsim$ is
$$A \precsim B \Leftrightarrow DR(A) \trianglelefteq_{Eli} DR(B)\ , \tag{4.1}$$
such that for[10] $\Gamma,\ \Gamma' \subseteq_{\text{fin}} \mathcal{R}_d$,
$$\Gamma \trianglelefteq_{Eli} \Gamma' \Leftrightarrow \left[\Gamma = \Gamma' \text{ or } \Gamma \triangleleft_{Eli} \Gamma'\right] \text{ and} \tag{4.2}$$
$$\Gamma \triangleleft_{Eli} \Gamma' \Leftrightarrow (\exists x \in \Gamma)\left(\forall y \in \Gamma'\right)\ x <_D y\ , \tag{4.3}$$
where in Equation 4.3 the order $<_D$ is defined by[11] Equation 3.3. It is easy to show that $\precsim$ is a preorder on $\mathcal{A}$, and $A \approx B \Leftrightarrow DR(A) = DR(B)$. 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 $\trianglelefteq_{Eli}$ to calculate $\precsim$ [13]. Therefore, one might consider using $\trianglelefteq_{Eli}$ (Equations 4.1, 4.2 and 4.3) for calculating $\precsim$. But $\trianglelefteq_{Eli}$ does not yield a correspondence with PDL as the following example illustrates.

Example 2. Consider $\langle D,\ W,\ \prec^+\rangle$ where $W = \{a\}$,
$$D = \left\{d_1 := \frac{a : b}{b},\ d_2 := \frac{b : c}{c},\ d_3 := \frac{b : \neg c}{\neg c}\right\} \tag{4.4}$$
and $d_1 \prec^+ d_2 \prec^+ d_3$. The extension is $Th(\{a, b, \neg c\})$.

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

Under the elitist ordering (Equation 4.2), it is neither the case that $\{r_1,\ r_2\} \triangleleft_{Eli} \{r_1,\ r_3\}$ nor $\{r_1,\ r_3\} \triangleleft_{Eli} \{r_1,\ r_2\}$. As the sets are not equal, we have $A \not\precsim B$, $B \not\precsim A$ and $A \not\approx B$. This means $A \hookrightarrow B$ and $B \hookrightarrow A$, which means there are two possible stable extensions $\{A\}$ and $\{B\}$ so that neither argument is sceptically justified, and so $\neg c$ is not an argumentation-defined inference. However $\neg c$ is a PDL

[6]We will discuss why we only consider LPDTs in Section 6.
[7]For example, $\neg(\theta \wedge \neg\phi)$ is the contrary of $(\theta \wedge \neg\phi)$, but $(\theta \to \phi)$, where $\to$ in this case denotes material implication, is not the contrary of $(\theta \wedge \neg\phi)$.
[8]Two defeasible rules are *equal* iff they have the same antecedents and consequent *syntactically*.
[9]From Footnote 4, we do not need to define $<_D$ as the order-theoretic dual to $\prec^+$, avoiding potential confusion as to which item is more preferred.
[10]It suffices to consider finite sets as arguments are finite.
[11]It can be shown that this definition avoids counterexamples like [9, Example 5.1], as explained in [17].

inference. Therefore the elitist ordering cannot be used to calculate $\precsim$.

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

$$\Gamma \lhd_{DEli} \Gamma' \Leftrightarrow \left(\exists x \in \Gamma - \Gamma'\right) \left(\forall y \in \Gamma' - \Gamma\right) \ x <_D y \ , \quad (4.5)$$

with $\lhd_{DEli}$ replacing $\lhd_{Eli}$ in Equations 4.2 and 4.3. Given $r_2 <_D r_3$, it is easy to see that $A \prec B$, $B \not\prec A$, and so $A \not\rightharpoonup B$ and $B \hookrightarrow A$. Hence there is only a single stable extension containing the now sceptically justified argument B with conclusion $\neg c$.

It seems very intuitive for the disjoint elitist order to ignore shared rules, because when deciding whether $A \precsim B$ or $B \precsim A$, we should only focus on the fallible information on which the arguments differ.[12] 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 $D = \{d_k\}_{k=1}^5$, $W = \varnothing$,

$$d_1 := \frac{\top : c_1}{c_1},\ d_4 := \frac{c_3 : c_4}{c_4},\ d_3 := \frac{\top : c_3}{c_3},$$
$$d_2 := \frac{c_1 : c_2}{c_2},\ d_5 := \frac{c_1 : \neg(c_2 \wedge c_4)}{\neg(c_2 \wedge c_4)} \ ,$$

such that $d_1 \prec^+ d_4 \prec^+ d_3 \prec^+ d_2 \prec^+ d_5$. Our PDE is constructed in the usual manner starting from $E_0 = Th(\varnothing)$. Equation 2.4 gives the order of application of the defaults:

$$E_1 = E_0 + c_3,\ E_2 = E_1 + c_4,$$
$$E_3 = E_2 + c_1,\ E_4 = E_3 + \neg(c_2 \wedge c_4) \ , \quad (4.6)$$

with $E_k = E_4$ for all $k \geq 5$. The default d_2 is blocked because $\neg(c_2 \wedge c_4) \equiv (\neg c_2 \vee \neg c_4)$, and with c_4 (from d_4), we have $\neg c_2$, which blocks d_2. The unique PDE from this LPDT is

$$Th(\{c_1, \neg c_2, c_3, c_4\}) \ . \quad (4.7)$$

Now consider the corresponding arguments following our instantiation. We have the defeasible rules[13]

$$r_1 <_D r_4 <_D r_3 <_D r_2 <_D r_5 \ . \quad (4.8)$$

The relevant arguments and sets of defeasible rules are

$$A := [[\top \Rightarrow c_1] \Rightarrow c_2] \, , DR(A) = \{r_1, r_2\} \quad (4.9)$$
$$B := [[\top \Rightarrow c_3] \Rightarrow c_4] \, , DR(B) = \{r_3, r_4\} \quad (4.10)$$
$$C := [[\top \Rightarrow c_1] \Rightarrow \neg(c_2 \wedge c_4)] \, , DR(C) = \{r_1, r_5\} \ , \quad (4.11)$$
$$D := [B, C \to \neg c_2] \, , DR(D) = \{r_1, r_3, r_4, r_5\} \ . \quad (4.12)$$

For the correspondence to hold, the desired stable extension is $\{D,\ B,\ C,\ [\top \Rightarrow c_3],\ [\top \Rightarrow c_1]\}$ and all strict extensions thereof[14], which does have a conclusion set corresponding to Equation 4.7. However, this would require $D \hookrightarrow A$, which means, by Equation 2.1, $D \rightharpoonup A$ and $D \not\prec A$. Clearly, $D \rightharpoonup A$ on A. However, $D \not\prec A$ 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 $D \prec A$. 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.[15] 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:

$$\left\{ r \in \mathcal{R}_d \middle| Ante(r) \subseteq Conc\left[Args\left(\bigcup_{k=1}^{i-1} \{a_k\}\right)\right]\right\} - \bigcup_{j=1}^{i-1} \{a_j\} \, . \quad (4.13)$$

The intuition is: a_1 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 4.13 ensures that once a rule is applied it cannot be applied again. We then define $<_{SP}$ as (notice the dual order)

$$a_i <_{SP} a_j \Leftrightarrow j < i \ , \quad (4.14)$$

where $1 \leq i,\ j \leq |\mathcal{R}_d|$. We define the non-strict order to be $a_i \leq_{SP} a_i \Leftrightarrow [a_i = a_j \text{ or } a_i <_{SP} a_j]$. This makes sense because $i \mapsto a_i$ 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, $\lhd_{SP}$, as, for $\Gamma,\ \Gamma' \subseteq_{\text{fin}} \mathcal{R}_d$,

$$\Gamma \lhd_{SP} \Gamma' \Leftrightarrow \left(\exists x \in \Gamma - \Gamma'\right) \left(\forall y \in \Gamma' - \Gamma\right) x <_{SP} y \ . \quad (4.15)$$

The corresponding strict argument preference is

$$A \prec_{SP} B \Leftrightarrow DR(A) \lhd_{SP} DR(B) \ . \quad (4.16)$$

We define the corresponding non-strict preference as $A \precsim_{SP} B \Leftrightarrow [DR(A) \lhd_{SP} DR(B) \text{ or } DR(A) = DR(B)]$. This is the disjoint elitist order[16] (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* $\precsim_{SP}$ *satisfies*

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

[12] This has been considered in a different context in [5].

[13] Where, similar to Example 2, r_i corresponds to d_i via Equation 3.2, for $1 \leq i \leq 5$.

[14] Informally, in ASPIC$^+$, for $S \subseteq \mathcal{A}$ the strict extension of S is the smallest set containing S extended with all strict and firm arguments, and all possible applications of strict rules to those arguments. This becomes the deductive closure when ASPIC$^+$ is instantiated into classical logic. See [13, page 370, Definition 17] for more details.

[15] Informally, the "structure" of an argument A is given by the preordered set $\langle Sub(A), \subseteq_{\text{arg}} \rangle$.

[16] We use the disjoint elitist order instead of the usual elitist order because Example 2 shows that the usual elitist order does not give the correspondence in general.

Proof. If $DR(B) = DR(A)$ then $B \approx A$, so $B \precsim_{SP} A$. If $DR(A) \subset DR(B)$, then $DR(A) - DR(B) = \varnothing$, which means $B \prec_{SP} A$ is vacuously true from Equation 4.5 so $B \precsim_{SP} A$. □

This is intuitive because if $A \subseteq_{\text{arg}} B$, 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 $B \precsim_{SP} A$ 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 $\mathcal{P}_{\text{fin}}(\mathcal{R}_d)$.

Lemma 4.2. *If $\langle \mathcal{R}_d, <_D \rangle$ is a strict toset, then $\triangleleft_{DEli}$ is a strict total order on $\mathcal{P}_{fin}(\mathcal{R}_d)$.*

Proof. Let Γ, Γ', $\Gamma'' \in \mathcal{P}_{\text{fin}}(\mathcal{R}_d)$ be arbitrary. *Irreflexivity:* From Equation 4.5, as $\exists$ precedes $\forall$, $\Gamma \triangleleft_{DEli} \Gamma$ is false.

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

$$\begin{aligned}\Gamma &= \{a_1, \ldots, a_{n_1}\} \cup \{d_1, \ldots. d_{n_4}\} \cup \\ &\quad \{f_1, \ldots, f_{n_6}\} \cup \{g_1, \ldots, g_{n_7}\} \\ \Gamma' &= \{b_1, \ldots, b_{n_2}\} \cup \{d_1, \ldots. d_{n_4}\} \cup \\ &\quad \{e_1, \ldots, e_{n_5}\} \cup \{g_1, \ldots, g_{n_7}\} \\ \Gamma'' &= \{c_1, \ldots, c_{n_3}\} \cup \{e_1, \ldots. e_{n_5}\} \cup \\ &\quad \{f_1, \ldots, f_{n_6}\} \cup \{g_1, \ldots, g_{n_7}\} ,\end{aligned}$$

where the a's to g's in $\mathcal{R}_d$ denote *distinct* defeasible rules. The seven disjoint finite sets $\{a_1, \ldots, a_{n_1}\}, \ldots, \{g_1, \ldots, g_{n_7}\}$ partition $\Gamma \cup \Gamma' \cup \Gamma''$ and can be represented by the subregions of three overlapping circles of the corresponding Venn diagram. If $n_i = 0$ then the corresponding set is empty, e.g. if $n_1 = 0$ then $\{a_1, \ldots, a_{n_1}\} = \varnothing$.

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 $1 \le k \le n_1$ or $1 \le l \le n_4$,

$$\begin{aligned}&\left[\left(\bigwedge_{i=1}^{n_3} a_k < c_i\right) \wedge \left(\bigwedge_{j=1}^{n_5} a_k < e_j\right)\right] \\ \text{or } &\left[\left(\bigwedge_{i=1}^{n_3} d_l < c_i\right) \wedge \left(\bigwedge_{j=1}^{n_5} d_l < e_j\right)\right] . \end{aligned} \tag{4.17}$$

By writing out $\Gamma \triangleleft_{DEli} \Gamma'$ and $\Gamma' \triangleleft_{DEli} \Gamma''$ in terms of the elements $a_1, \ldots, g_{n_7}$ similar to how $\Gamma \triangleleft_{DEli} \Gamma''$ is written in Equation 4.17, we get four cases (as "and" and "or" bidistribute). One case gives a contradiction (due to irreflexivity of $<_D$), while the other three cases imply $\Gamma \triangleleft_{DEli} \Gamma''$ from the totality of $<_D$. Therefore, $\Gamma \triangleleft_{DEli} \Gamma''$.

Trichotomy: Assume $\Gamma \neq \Gamma'$ and consider $\Gamma \ominus \Gamma' \in \mathcal{P}_{\text{fin}}(\mathcal{R}_d)$. The structure $\langle \Gamma \ominus \Gamma', <_D \rangle$ is a strict finite toset, and thus has a $<_D$-least element m. Either $m \in \Gamma - \Gamma'$ exclusive-or $m \in \Gamma' - \Gamma$, where the former implies $\Gamma \triangleleft_{DEli} \Gamma'$ and the latter implies $\Gamma' \triangleleft_{DEli} \Gamma$. □

It follows that $\prec_{SP}$ is a strict total preorder on $\mathcal{A}$.

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

$$r_2 <_{SP} r_5 <_{SP} r_1 <_{SP} r_4 <_{SP} r_3 . \tag{4.18}$$

Notice that this is precisely the order in which the corresponding normal defaults are added in PDL, as Equation 4.6 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 $<_{SP}$-least, so $D \not\prec_{SP} A$, therefore $A \prec_{SP} D$.

However, $<_{SP}$ does not necessarily follow the PDL order of the application of defaults as the following example illustrates.

Example 5. Consider $\left\langle \{d_1, d_2\}, \{a\}, \prec^+ \right\rangle$ with $d_1 := \frac{a:\neg a}{\neg a}$ and $d_2 := \frac{\top:b}{b}$ such that $d_2 \prec^+ d_1$. We have $E = Th\left(\{a, b\}\right)$, where d_1 is blocked by W, so d_2 is the only default added. In argumentation, we have $\mathcal{K}_n = \{a\}$, $r_1 := (a \Rightarrow \neg a)$ and $r_2 := (\top \Rightarrow b)$ where for $i = 1, 2$, $r_i := f(d_i)$, such that $r_2 <_D r_1$. The arguments are $A_0 := [a]$, $A_1 := [A_0 \Rightarrow \neg a]$ and $B := [\top \Rightarrow b]$. Applying Equation 4.13, we have $r_2 <_{SP} r_1$, which clearly is not the order of how the corresponding defaults are added in PDL. Yet the correspondence still holds, since $A_0 \hookrightarrow A_1$ 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 5 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 5.3), which guarantees that the inferences with argumentation semantics under the preference $\precsim_{SP}$ 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 $E = \bigcup_{i \in \mathbb{N}} E_i$ one of its extensions generated from the linearisation $\prec^+ \supseteq \prec$. The *set of generating defaults (with respect to $\prec^+$), $GD(\prec^+)$,* is defined as

$$\begin{aligned}GD_i(\prec^+) &:= \left\{ d \in D \,\middle|\, d \text{ is } \prec^+\text{-greatest active in } E_i \right\} , \\ GD(\prec^+) &:= \bigcup_{i \in \mathbb{N}} GD_i(\prec^+) \subseteq D .\end{aligned} \tag{5.1}$$

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 $\left\langle \left\{ \frac{a:c}{c}, \frac{b:c}{c} \right\}, \{a, b\}, \varnothing \right\rangle$. We have two possible linearisations $\frac{a:c}{c} \prec_1^+ \frac{b:c}{c}$ and $\frac{b:c}{c} \prec_2^+ \frac{a:c}{c}$. By Footnote 4 we have $GD(\prec_1^+) = \left\{ \frac{b:c}{c} \right\}$ and $GD(\prec_2^+) = \left\{ \frac{a:c}{c} \right\}$, which are not equal, even though both linearisations give the same extension $E = Th\left(\{a, b, c\}\right)$. But in the case of $\prec_1^+$, $\frac{b:c}{c}$ is not active not because it is blocked by $\neg c$, 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 $\frac{\theta:\phi}{\phi}$ *semi-active (in $S \subseteq \mathcal{SL}$)* iff $[\theta \in S,\ \neg\phi \notin S,\ \phi \in S]$. The *set of semi-active defaults with respect to the linearisation* $\prec^+$ is

$$SAD(\prec^+) := \left\{d \in D \middle|\ d \text{ is semi-active w.r.t. } \prec^+\right\} . \quad (5.2)$$

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

$$NBD(\prec^+) := GD(\prec^+) \cup SAD(\prec^+) \subseteq D . \quad (5.3)$$

NBD has a more elegant characterisation:

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

$$NBD(\prec^+) := \left\{ \frac{\theta : \phi}{\phi} \in D \middle| \theta \in E,\ \neg\phi \notin E \right\} . \quad (5.4)$$

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 $d := \frac{\theta:\phi}{\phi} \in GD \Rightarrow [\theta \in E,\ (\exists i \in \mathbb{N})\ \neg\phi \notin E_i]$, and assuming $\neg\phi \in E$ gives a contradiction (by considering the E_i's in E), so $\neg\phi \notin E$. Trivially, $d \in SAD \Rightarrow [\theta \in E,\ \neg\phi \notin E]$, therefore Equation 5.3 is a subset of Equation 5.4. Assuming that $d \in$ the right hand side of Equation 5.4 gives $d \in GD \cup SAD$ through simple quantifier manipulations. The result follows. □

Given E, $NBD(\prec^+)$ is uniquely determined, so we will write $NBD(E)$ instead. Equation 5.4 adapts Reiter's idea of a *generating default* [16, page 92 Definition 2] to PDL.

5.2 Uniqueness of Stable Extensions

In this section we show that the defeat graph $\langle \mathcal{A},\ \hookrightarrow \rangle$ associated with any ASPIC$^+$ attack graph $\langle \mathcal{A},\ \rightharpoonup,\ \precsim_{SP} \rangle$ constructed from a LPDT $\left\langle D,\ W,\ \prec^+ \right\rangle$ with the SP-order has a unique stable extension.

Theorem 5.2. *Let $\langle \mathcal{A},\ \rightharpoonup,\ \precsim_{SP} \rangle$ be an ASPIC$^+$ attack graph constructed from $\mathcal{L} = \mathcal{FL}$, $-$ is $\neg$, $\mathcal{R}_s$ the rules of proof of FOL, $\langle \mathcal{R}_d,\ <_{SP} \rangle$ the toset of defeasible rules under the SP-order, the argument preference $\precsim_{SP}$, n is undefined on $\mathcal{R}_d$, $\mathcal{K}_p = \varnothing$ and $\mathcal{K}_n \subseteq \mathcal{FL}$ is a consistent set of formulae. The defeat graph $\langle \mathcal{A},\ \hookrightarrow \rangle$ from this attack graph 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 $S \subseteq \mathcal{A}$ we define, for $r \in \mathcal{R}_d$, $S \oplus r := Args(DR(S) \cup \{r\})$, 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 $\langle \mathcal{A},\ \rightharpoonup,\ \precsim_{SP} \rangle$ obeying the conditions of the theorem, and outputs a set of arguments.

Algorithm 1 Generating a Stable Extension

```
1: function GenerateStableExtension(⟨A, ⇀, ≾_SP⟩)
2:     S ← {all strict arguments in A}
3:     for r ∈ R_d from <_SP-greatest to <_SP-smallest do
4:         if S ⊕ r has no attacks, (S ⊕ r)^2 ∩ ⇀ = ∅, then
5:             S ← S ⊕ r
       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 5) and never attack each other because $\mathcal{K}_n$ 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 4–5). As $<_{SP}$ 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 $<_{SP}$ while ensuring conflict freeness mimics the condition of Equation 2.4.

It is clear from the algorithm that S exists and is unique, as it is of the form $Args(R)$ for some $R \subseteq \mathcal{R}_d$. We show S is a stable extension [2, page 26 Definition 2.2.7]: cf is guaranteed by the consistency of $\mathcal{K}_n$ and that defeasible rules $r \in \mathcal{R}_d$ 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 $A \hookrightarrow B$. 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 $B' \subseteq_{\text{arg}} B$ such that $TopRule(B') = r$. Let $A \in S$ be the attacker of B' at r. If r is $<_{SP}$-greatest, then $Args(\varnothing) \oplus r$ contains attacking arguments, so A must be strict and hence $A \hookrightarrow B$. If r is not $<_{SP}$-greatest, then consider the strict up-set of r in $\langle \mathcal{R}_d,\ <_{SP} \rangle$, $T := \{r' \in \mathcal{R}_d \,|\, r <_{SP} r'\} \neq \varnothing$. If $T \cap R = \varnothing$, then adding r to S means there is an attack from $A \in Args(\varnothing)$ under $\precsim_{SP}$, and hence B is defeated. If $T \cap R \neq \varnothing$, then $A \in Args(T \cap R)$. Assume A is not strict, then $\varnothing \neq DR(A) \subseteq T \cap R$, so $(\forall s \in DR(A))\, r <_{SP} s$, hence A defeats B. Therefore, in all cases, $A \hookrightarrow B$, and hence B is defeated by some argument in S. Therefore, the defeat graphs of such 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 $\langle \mathcal{A},\ \hookrightarrow \rangle$ with the extension of the corresponding LPDT $\left\langle D,\ W,\ \prec^+ \right\rangle$.

Theorem 5.3. *Let $\langle \mathcal{A},\ \rightharpoonup,\ \precsim_{SP} \rangle$ be the attack graph corresponding to an LPDT $\left\langle D, W, \prec^+ \right\rangle$ with defeat graph $\langle \mathcal{A},\ \hookrightarrow \rangle$ under $\precsim_{SP}$.*

1. *Let E be the extension of $\left\langle D, W, \prec^+ \right\rangle$. Then there exists a unique stable extension $\mathcal{E} \subseteq \mathcal{A}$ of $\langle \mathcal{A}, \hookrightarrow \rangle$ such that $Conc\,(\mathcal{E}) = E$.*
2. *Let $\mathcal{E} \subseteq \mathcal{A}$ be the unique stable extension of $\langle \mathcal{A}, \hookrightarrow \rangle$ by Theorem 5.2, then $Conc(\mathcal{E})$ is the extension of $\left\langle D, W, \prec^+ \right\rangle$.*

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

$DR(A) \subseteq f(NBD(E))$ and hence $a \in E$. Now let $B' \subseteq_{\text{arg}} B$ be the argument such that $TopRule(B') = (b \Rightarrow \neg a)$ for some appropriate formula b in B. As $B \in \mathcal{E}$, this means $(b \Rightarrow \neg a) \in f(NBD(E))$ and hence $a \notin E$ by Equation 5.4 and that E is deductively closed – contradiction. Therefore, $\mathcal{E}$ is cf. To show $Args(R)$ defeats all other arguments, let $B \notin Args(R)$ so there is some rule $r \in DR(B) - R$. Let $B' \subseteq_{\text{arg}} B$ be such that $TopRule(B') = r$. r corresponds to a default $f^{-1}(r) = \frac{\theta:\phi}{\phi} \notin NBD(E)$. Either $\theta \notin E$ or $\neg\phi \in E$ by Equation 5.4. If $\neg\phi \in E$, then one can prove there exists an argument $A \in Args(R)$ such that $A \hookrightarrow B'$ and hence $A \hookrightarrow B$ under $\precsim_{SP}$. Assume $\theta \notin E$. There is some $B'' \subset_{\text{arg}} B'$ such that $Conc(B'') = \theta$. If $\theta \notin E$ then B'' is neither strict nor in $Args(f(NBD(E)))$. Thus there is some other $s \in DR(B'') - f(NBD(E))$. 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 $Args(f(NBD(E)))$. Therefore, $\theta \in E$. Therefore, the only reason for $r \notin R$ is because $\neg\phi \in E$, and hence there is an argument A that defeats any argument containing the rule r, which means $Args(R)$ defeats all other arguments and hence it is a stable extension. To show that $E = Conc(\mathcal{E})$, we show $E \subseteq Conc(\mathcal{E})$ and $Conc(\mathcal{E}) \subseteq E$. In the first case, let $\theta \in E$, then if $\theta \in E_0$, we have $W \models \theta$ and by the compactness theorem in FOL, we have $\Delta \subseteq_{\text{fin}} W$ such that $\Delta \models \theta$. From this we build a strict argument A such that $Prem_n(A) = \Delta$ and $Conc(A) = \theta$, and necessarily $A \in \mathcal{E}$ so $\theta \in Conc(\mathcal{E})$. Similarly, if $\theta \in E_k$ for some $k \in \mathbb{N}^+$, we can construct a defeasible argument A concluding θ such that $DR(A) \subseteq f(NBD(E))$ and hence $A \in \mathcal{E}$, so $\theta \in Conc(\mathcal{E})$. Conversely, if $\theta \in Conc(\mathcal{E})$ there is an argument in $\mathcal{E}$ concluding θ. If this argument is strict then $\theta \in E_0 \subseteq E$, else, as the defeasible rules are in $f(NBD(E))$ then $\theta \in E_k \subseteq E$ for some $k \in \mathbb{N}^+$ 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 $Conc(\mathcal{E}) \subseteq E$ and $E \subseteq Conc(\mathcal{E})$. For the former, if $\theta \in Conc(\mathcal{E})$ then there is some $A \in \mathcal{E}$ concluding θ. If A is strict then $\theta \in E_0 \subseteq E$. If A is defeasible, then say $DR(A) = \{d_i\}_{i=1}^k$ for some $k \in \mathbb{N}^+$. All of these defaults do not introduce any inconsistency because $\mathcal{E}$ is stable and hence cf. Take the smallest $i \in \mathbb{N}$ such that sufficiently many corresponding defeasible rules are applied from $DR(A)$ to conclude θ in E_{i+1} from W. Therefore, $\theta \in E_{i+1} \subseteq E$. Conversely, let $\theta \in E$, so there is some $i \in \mathbb{N}$ such that $\theta \in E_i$. If $i = 0$, then there is a strict argument A, necessarily in $\mathcal{E}$, that concludes θ so $\theta \in Conc(\mathcal{E})$. If $i > 0$, then we can use the appropriate defeasible rules to construct a defeasible argument A such that $Prem(A) \subseteq W$, $Conc(A) = \theta$ and $DR(A) \neq \varnothing$. To show $A \in \mathcal{E}$, we assume for contradiction that $A \notin \mathcal{E}$ so there is some $B \in \mathcal{E}$ such that $B \hookrightarrow A$. But this would result in at least one of the defaults of A being blocked. This is a contradiction because given that $\theta \in E$, the defeasible argument A concluding θ cannot have its defeasible rules correspond to blocked PDL defaults. Therefore, such a B cannot exist and $A \in \mathcal{E}$. This proves that $Conc(\mathcal{E}) = E$. ☐

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, $Cl_{\mathcal{R}_s}$ becomes deductive closure.

Theorem 5.4. *Let $\langle \mathcal{A}, \rightharpoonup, \precsim_{SP} \rangle$ be the ASPIC$^+$ attack graph of PDL and let $\mathcal{E}$ be any of the admissible extensions of the corresponding defeat graph $\langle \mathcal{A}, \hookrightarrow \rangle$. Our instantiation satisfies the Caminada-Amgoud rationality postulates.*

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

6. DISCUSSION AND CONCLUSION

We have endowed PDL [4] with argumentation semantics using ASPIC$^+$ [13]. 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 5.3), 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 $\prec^+$ 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 semantics[17] 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 $\precsim_{SP}$ 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

[1] K. Atkinson, T. Bench-Capon, and P. McBurney. A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. *Autonomous Agents and Multi-Agent Systems*, 11(2):153–171, 2005.

[2] P. Besnard and A. Hunter. *Elements of Argumentation.* The MIT Press, 2008.

[3] G. Brewka. Preferred Subtheories: An Extended Logical Framework for Default Reasoning. In *IJCAI*, volume 89, pages 1043–1048, 1989.

[4] G. Brewka. Adding Priorities and Specificity to Default Logic. In *Logics in Artificial Intelligence*, pages 247–260. Springer, 1994.

[5] G. Brewka, M. Truszczynski, and S. Woltran. Representing Preferences Among Sets. In *AAAI*, 2010.

[6] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal Generation in the BOID Architecture. *Cognitive Science Quarterly Journal*, 2(3-4):428–447, 2002.

[7] M. Caminada and L. Amgoud. On the Evaluation of Argumentation Formalisms. *Artificial Intelligence*, 171(5):286–310, 2007.

[8] P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77:321–357, 1995.

[9] P. M. Dung. An Axiomatic Analysis of Structured Argumentation for Prioritised Default Reasoning. In *ECAI2014*, pages 267–272. IOS Press, 2014.

[10] G. Governatori, M. J. Maher, G. Antoniou, and D. Billington. Argumentation Semantics for Defeasible Logic. *Journal of Logic and Computation*, 14(5):675–702, 2004.

[11] S. Modgil. An Argumentation Based Semantics for Agent Reasoning. In M. Dastani, A. El Fallah Seghrouchni, J. Leite, and P. Torroni, editors, *Languages, Methodologies and Development Tools for Multi-Agent Systems*, volume 5118 of *Lecture Notes in Computer Science*, pages 37–53. Springer Berlin Heidelberg, 2008.

[12] S. Modgil and M. Caminada. Proof Theories and Algorithms for Abstract Argumentation Frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.

[13] S. Modgil and H. Prakken. A General Account of Argumentation with Preferences. *Artificial Intelligence*, 195:361–397, February 2013.

[14] S. Modgil, F. Toni, et al. The Added Value of Argumentation. In S. Ossowski, editor, *Agreement Technologies*, volume 8, pages 357–403. Springer Netherlands, 2013.

[15] H. Prakken. An Abstract Framework for Argumentation with Structured Arguments. *Argument and Computation*, 1(2):93–124, 2010.

[16] R. Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[17] A. P. Young, S. Modgil, and H. Prakken. A General Account of Argumentation with Preferences - Erratum. Available from `http://www.dcs.kcl.ac.uk/staff/smodgil/2015-08-30_ASPIC+_fix_revised.pdf`, last accessed 7/11/2015.