Relating Carneades with Abstract Argumentation

Bas van Gijzel
Utrecht University
B.M.vanGijzel@students.uu.nl

Henry Prakken
Utrecht University & University of Groningen
henry@cs.uu.nl

Abstract

Carneades is a recently proposed formalism for structured argumentation with varying proof standards. An open question is its relation with Dung's seminal abstract approach to argumentation. In this paper the two formalisms are formally related by translating Carneades into *ASPIC*$^+$, another recently proposed formalism for structured argumentation. Since *ASPIC*$^+$ is defined to generate Dung-style abstract argumentation frameworks, this in effect translates Carneades graphs into abstract argumentation frameworks. It is proven that Carneades always induces a unique Dung extension, which is the same in all of Dung's semantics.

1 Introduction

Carneades [Gordon *et al.*, 2007; Gordon and Walton, 2009] is a recently proposed formalism for structured argumentation with varying proof standards, inspired by legal reasoning but more generally applicable. Its distinctive feature is that each statement can be given its own proof standard, which is claimed to allow a more natural account of reasoning under burden of proof than existing formalisms for structured argumentation, in which proof standards are defined globally. This makes the Carneades formalism very useful, as signified by the large number of citations due to its proof standards. However, to date its relation with [Dung, 1995]'s seminal abstract approach to argumentation is unknown, which obscures its relation with mainstream work on argumentation in AI. Recently, [Brewka and Gordon, 2010] translated Carneades into [Brewka and Woltran, 2010]'s abstract dialectical frameworks. Such frameworks generalise Dung's approach in that abstract argumentation frameworks are a special case of abstract dialectical frameworks. However, this translation relies on the full expressiveness of abstract dialectical frameworks, so that it does not clarify the relation of Carneades with Dung's abstract argumentation frameworks, nor with formalisms that generate such frameworks[1]

In this paper we provide such a formal relation between Carneades and abstract argumentation, by translating Carneades into *ASPIC*$^+$ [Prakken, 2010]. Since *ASPIC*$^+$ is defined to generate Dung-style abstract argumentation frameworks, we in effect translate Carneades graphs into abstract argumentation frameworks. Thus, contrary to what was suggested in [Gordon *et al.*, 2007], we show that varying proof standards can be modelled in Dung-style semantics. Also, contrary to what was claimed by [Brewka and Gordon, 2010], we prove that Carneades can be modelled cycle-free, thus always inducing a unique Dung extension, which is the same in all of Dung's semantics. This allows us to generalise Carneades' argument evaluation structures to cycle-containing structures, addressing an important issue left for future research by [Gordon and Walton, 2009].

The paper is structured as follows: Section 2 reviews abstract argumentation, the *ASPIC*$^+$ framework, and relevant parts of the Carneades framework. Section 3 then translates Carneades into Dung's argumentation frameworks through *ASPIC*$^+$ and proves the correspondence result. Finally, Section 4 concludes and discusses future work.

2 Background

In this section we review Dung's abstract argumentation frameworks and the *ASPIC*$^+$ framework.

2.1 Abstract Argumentation Frameworks

Dung's abstract argumentation frameworks consist of a set of arguments ordered by a binary relation of defeat.[2]

Definition 2.1 (Abstract argumentation framework). An *abstract argumentation framework* is a tuple $\langle Args, Def \rangle$, such that $Args$ is a set of arguments and $Def \subseteq Args \times Args$ is a defeat relation on the arguments in $Args$.

Definition 2.2. Let $AF = \langle Args, Def \rangle$ and $S \subseteq Args$.

1. S is called *conflict-free* iff $\neg\exists A, B \in S$ such that $(A, B) \in Def$.
2. An argument $A \in Args$ is *acceptable* w.r.t. S iff $\forall B \in Args$, if $(B, A) \in Def$ then $\exists C \in S$ such that $(C, B) \in Def$.

[1] When writing the final version of this paper, we were informed that [Brewka *et al.*, 2011] have meanwhile proved a formal correspondence between ADFs and Dung AFs.

[2] Dung calls it 'attack' but to unify terminology we rename it to 'defeat'.

3. The *characteristic function* of an AF, F_{AF} is a function such that:
   - $F_{AF} : 2^{Args} \mapsto 2^{Args}$,
   - $F_{AF}(S) = \{A \mid A$ is acceptable w.r.t. to $S\}$.
4. A conflict-free set of arguments S is *admissible* iff every argument $A \in S$ is acceptable w.r.t. S, i.e. $S \subseteq F_{AF}(S)$.

Definition 2.3 (Extensions). Given a conflict-free set of arguments S and an argumentation framework AF, then if F is monotonic:

- S is a *complete extension* iff $S = F_{AF}(S)$.
- S is a *grounded extension* iff it is the least fixed point of F_{AF}.
- S is a *preferred extension* iff it is a greatest fixed point of F_{AF}.
- S is a *stable extension* iff it is a preferred extension defeating all arguments in $Args \backslash S$.

Definition 2.4 (Well-founded argumentation framework). An argumentation framework is *well-founded* iff there does not exist an infinite sequence of arguments: $A_0, A_1, \ldots, A_n, \ldots$ such that for each i, $(A_{i+1}, A_i) \in Def$.

The differences between the semantics collapse in an argumentation framework in which there are no cycles.

Theorem 2.5 (Theorem 30 of Dung [Dung, 1995]). *Every well-founded argumentation framework has exactly one complete extension which is grounded, preferred and stable.*

2.2 Structured Argumentation Frameworks

[Prakken, 2010]'s *ASPIC*$^+$ framework further develops [Amgoud *et al.*, 2006]'s way to give structure to Dung's arguments and defeat relation. It assumes an unspecified logical language $\mathcal{L}$, and defines arguments as inference trees formed by applying strict or defeasible inference rules of the form $\varphi_1, \ldots, \varphi_n \to \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, interpreted as 'if the *antecedents* $\varphi_1, \ldots, \varphi_n$ hold, then *without exception*, respectively *presumably*, the *consequent* φ holds'. In order to define attacks, some minimal assumptions on $\mathcal{L}$ are made; namely that certain wff are a contrary or contradictory of certain other wff. Apart from this the framework applies to any set of strict and defeasible inference rules, and to any logical language with a defined contrary relation.

The basic notion of *ASPIC*$^+$ is that of an argumentation system. Arguments are then constructed w.r.t a knowledge base that is assumed to contain four kinds of formulas.

Definition 2.6 (Argumentation system). An *argumentation system* is a tuple $AS = \langle \mathcal{L}, ^-, \mathcal{R}, \leq \rangle$ where:

- $\mathcal{L}$ is a logical language.
- $^-$ is a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$, such that if $\varphi \in \overline{\psi}$ then:
  - if $\psi \notin \overline{\varphi}$ then φ is called a *contrary* of ψ,
  - otherwise, $\psi \in \overline{\varphi}$ and φ and ψ are called *contradictory*, denoted by $\varphi = -\psi$ (i.e., $\varphi \in \overline{\psi}$ and $\psi \in \overline{\varphi}$).
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $\leq$ is a partial preorder on $\mathcal{R}_d$.

Definition 2.7 (Knowledge base). A *knowledge base* in an argumentation system $\langle \mathcal{L}, ^-, \mathcal{R}, \leq \rangle$ is a pair $\langle \mathcal{K}, \leq' \rangle$ where $\mathcal{K} \subseteq \mathcal{L}$ and $\leq'$ is a preorder on $\mathcal{K} \setminus \mathcal{K}_n$. Here, $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a \cup \mathcal{K}_i$ where these subsets of $\mathcal{K}$ are disjoint, being the (necessary) *axioms* (which cannot be attacked), the *ordinary premises* (on which attacks succeed contingent upon preferences), the *assumptions* (on which attacks are always successful) and the *issues* (which must always be backed with a further argument).

An argument that contains issue premises should not be acceptable. Accordingly, Definition 2.2 is changed to: An argument $A \in Args$ is acceptable w.r.t. S iff *A contains no issue premises and* ....

Definition 2.8 (Argument). An *argument* A on the basis of a knowledge base $\langle \mathcal{K}, \leq' \rangle$ in an argumentation system $\langle \mathcal{L}, ^-, \mathcal{R}, \leq \rangle$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$.
2. $A_1, \ldots A_n \quad \to/\Rightarrow \quad \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \to/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
   $\texttt{Conc}(A) = \psi$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.

Where Prem, Conc and Sub respectively are the premises, conclusions and subarguments of an argument.

Definition 2.9 (Argumentation theories). An *argumentation theory* is a triple $AT = \langle AS, KB, \preceq \rangle$ where AS is an argumentation system, KB is a knowledge base in AS and $\preceq$ is an argument ordering on the set of all arguments that can be constructed from KB in AS.

Arguments can be attacked in three ways: by attacking a premise (undermining), a conclusion (rebutting), or an inference (undercutting). To model undercutting attacks, defeasible inference rules are given names, e.g. $P \Rightarrow_{app} c$, allowing arguments to attack an inference rule by using their shorthand name, *app*. Apart from undercut attacks and attacks on contraries, the success of attacks as defeats depends on the preference relation between the attacker and its target.

Definition 2.10 (Types of attack).

- Argument A *undermines* argument B (on φ) iff $Conc(A) \in \overline{\varphi}$ for some $\varphi \in Prem(B) \backslash \mathcal{K}_n$. In such a case A *contrary-undermines* B iff $Conc(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$.
- Argument A *undercuts* argument B (on B') iff $Conc(A) \in \overline{B'}$ for some $B' \in Sub(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$.
- Argument A *rebuts* argument B (on B') iff $Conc(A) \in \overline{\varphi}$ for some $B' \in Sub(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$ In such a case A *contrary-rebuts* B iff $Conc(A)$ is a contrary of φ.

Definition 2.11 (Types of defeat).

- Argument A *successfully rebuts* argument B if A rebuts B on B' and either A contrary-rebuts B' or $A \not\prec B'$.
- Argument A *successfully undermines* argument B if A undermines B on φ and either A contrary-undermines B or $A \not\prec \varphi$.

The previous notions can be combined in an overall definition of defeat:

Definition 2.12 (Defeat). Argument A *defeats* argument B iff no premise of A is an issue and A undercuts or successfully rebuts or successfully undermines B. Argument A *strictly defeats* argument B iff A defeats B and B does not defeat A.

ASPIC$^+$'s argumentation theories are then linked to Dung's abstract argumentation frameworks as follows:

Definition 2.13 (Argumentation framework). An *abstract argumentation framework (AF) corresponding to an argumentation theory* $\langle AS, KB, \preceq\rangle$ is a pair $\langle Args, Def\rangle$ such that:

- $Args$ is the set of arguments on the basis of KB in AS as defined by Definition 2.8,
- Def is the relation on $Args$ given by Definition 2.12.

2.3 Carneades

As in *ASPIC*$^+$, arguments in Carneades are not left abstract but given structure. Arguments are constructed by linking premises and exceptions to a conclusion. Unlike *ASPIC*$^+$, Carneades does not assume that arguments are constructed by applying inference rules. Also, Carneades' notion of an argument is not inductive; subarguments are modelled implicitly by the inductive definition of applicability of arguments.

Definition 2.14 (Arguments). Let $\mathcal{L}$ be a propositional language. An *argument* is a tuple $\langle P, E, c\rangle$ where $P \subset \mathcal{L}$ are its *premises*, $E \subset \mathcal{L}$ with $P \cap E = \emptyset$ are its *exceptions* and $c \in \mathcal{L}$ is its *conclusion*. Both c and all members of P and E are propositional literals. Let p be a literal. If p is c, then the argument is an argument *pro* p. If p is the complement of c, then the argument is an argument *con* p.

In Carneades a dialogue is a sequence of stages but for evaluating arguments in a specific stage the other stages are irrelevant. As in [Brewka and Gordon, 2010] we therefore only consider *stage specific Carneades argument evaluation structures*. To define them, the concepts of an *audience* and an *acyclic set* of arguments must be introduced.

Definition 2.15 (Audience). Let $\mathcal{L}$ be a propositional language. An *audience* is a tuple $\langle assumptions, weight\rangle$, where $assumptions \subset \mathcal{L}$ is a **consistent** set of literals assumed to be acceptable by the audience and *weight* is a function mapping arguments to real numbers in the range $0.0 \ldots 1.0$, representing the relative weights assigned by the audience to the arguments.

Definition 2.16 (Acyclic set of arguments). A set of *arguments* is *acyclic* iff its corresponding dependency graph is acyclic. The corresponding dependency graph has nodes for every literal appearing in the set of arguments. A node p has a directed link to node q whenever p depends on q in that there is an argument pro or con p that has q or $\overline{q}$ in its set of premises or exceptions.[3]

The previous definitions can now be combined to define Carneades' concept of an evaluation structure:

Definition 2.17 (Stage specific Carneades argument evaluation structure). A *(stage specific) Carneades argument evaluation structure* (CAES) is a tuple $\langle arguments, audience, standard\rangle$, where $arguments$ is an aycyclic set of arguments, $audience$ is an audience and $standard$ is a total function mapping literals in $\mathcal{L}$ to their applicable proof standards.

In a CAES each statement is assigned a standard of proof. The current Carneades model includes five proof standards, *scintilla of evidence*, *preponderance of the evidence*, *clear and convincing evidence*, *beyond reasonable doubt* and *dialectical validity*. A proof standard is a function that given a literal p, aggregates applicable arguments pro and con p and evaluates to $true$ or $false$ depending on a specific audience.

Definition 2.18 (Proof standard). A *proof standard* is a function mapping tuples $\langle issue, arguments, audience\rangle$ to $\{true, false\}$, where $issue$ is a literal in $\mathcal{L}$, *arguments* is an acyclic set of arguments and *audience* is an audience.

Given a CAES and the concept of a proof standard the acceptability of a literal is defined.

Definition 2.19 (Acceptability of literals). Let $C = \langle arguments, audience, standard\rangle$ be a CAES, p a literal in $\mathcal{L}$ and $s = standard(p)$ the proof standard corresponding to P. Then the literal p is *acceptable* in C iff $s(p, arguments, audience)$ is $true$.

All proof standards defined depend on the concept of argument applicability and thus this needs to be defined first.

Definition 2.20 (Applicability of arguments). Let $C = \langle arguments, audience, standard\rangle$ be a CAES. An argument $\langle P, E, c\rangle \in arguments$ is *applicable* in C iff

- $p \in P$ implies p is an assumption of the $audience$ or [$\overline{p}$ is not an assumption and p is acceptable in C] and
- $p \in E$ implies p is not an assumption of the $audience$ and [$\overline{p}$ is an assumption or p is not acceptable in C].

Now Carneades' proof standards can be defined.

Definition 2.21 (Proof standards). Given a CAES $C = \langle arguments, audience, standard\rangle$ and a literal p in $\mathcal{L}$.

- *scintilla*$(p, arguments, audience) = true$ iff there exists at least one applicable argument pro p in $arguments$.
- *preponderance*$(p, arguments, audience) = true$ iff there exists at least one applicable argument pro p in $arguments$ for which the weight assigned by the $audience$ is greater than the weight of the applicable arguments con p.
- *clear-and-convincing*$(p, arguments, audience) = true$ iff there is an applicable argument A, pro p for which:
  - *preponderance*$(p, arguments, audience)$ holds and
  - the weight for A exceeds the threshold α, and
  - the difference between the weight of A and the maximum weight of the applicable con arguments exceeds the treshold β.

[3] As usual $\overline{p}$ is a complement of p, e.g. $\neg p$.

- *beyond-reasonable-doubt*$(p, arguments, audience) = true$ iff *clear-and-convincing*$(p, arguments, audience)$ holds and the maximum weight of the applicable con arguments is less than the threshold γ.
- *dialectical-validity*$(p, arguments, audience) = true$ iff there exists at least one applicable argument pro p in $arguments$ and no argument con p in $arguments$ is applicable.

3 Translation of Carneades

We now provide our translation of Carneades into *ASPIC*$^+$. The ideas are as follows. Assumptions in Carneades correspond to axiom premises in *ASPIC*$^+$, while non-assumption premises of Carneades arguments for which there is no further argument are issue premises in *ASPIC*$^+$. For each CAES argument $a = \langle P, E, c\rangle$ two defeasible rules are added to $\mathcal{R}_d$: a rule $P \Rightarrow_{app_a} arg_a$, saying that if P then a is applicable[4], and $arg_a \Rightarrow_{acc_a} c$, saying that if a is applicable, its conclusion is acceptable (app_a and acc_a are the rules' names). To this end, the language $\mathcal{L}$ of a CAES must be enriched with literals composed of arg_a, app_a and acc_a for each argument a in CAES. Next, for each exception $e \in E$ an undercutter $e \Rightarrow \neg app_a$ is added to $\mathcal{R}_d$. Finally, the contrariness relation on $\mathcal{L}$ is extended to let applicability conclusions for one argument defeat the acceptability of conflicting arguments, depending on the proof standards of their conclusions. This is essentially where the proof standards are encoded.

Definition 3.1 (Argumentation system corresponding to a CAES). Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$ and propositional language $\mathcal{L}_{CAES}$ the corresponding argumentation system, *AS*, is a tuple $\langle \mathcal{L}, ^-, \mathcal{R}, \leq\rangle$ where:

- $\mathcal{L} = \mathcal{L}_{CAES} \cup$ argument nodes $\cup$ rule names,
- $^-$ consists of all tuples specified below,
- $\mathcal{R}_d = \bigcup_{a \in arguments} \mathcal{R}_{d_a}$,
- $\mathcal{R}_s = \bigcup_{a \in arguments} \mathcal{R}_{s_a}$,
- $\leq = \{(r, r) \mid r \in \mathcal{R}_d\}$.

For every argument $a = \langle P, E, c\rangle$ in $arguments$:

$$\mathcal{R}_{d_a} = \{P \Rightarrow_{app_a} arg_a;\ arg_a \Rightarrow_{acc_a} c\} \cup \{e_i \Rightarrow \neg app_a \mid e_i \in E\}$$

For every argument $a = \langle P, E, c\rangle$ in $arguments$ with $standard(c) =$ *scintilla*:

$$\mathcal{R}_{s_a} = \emptyset$$

For every argument $a = \langle P, E, c\rangle$ in $arguments$ with $standard(c) =$ *preponderance*:

$$\mathcal{R}_{s_a} = \emptyset$$
$$^-(acc_a) = \{arg_b \mid b = \langle P', E', \overline{c}\rangle \in arguments, weight(a) \leq weight(b)\}$$

For every argument $a = \langle P, E, c\rangle$ in $arguments$ with $standard(c) =$ *clear-and-convincing*:

$$\mathcal{R}_{s_a} = \{\rightarrow \neg acc_a \mid weight(a) \leq \alpha\}$$
$$^-(acc_a) = \{arg_b \mid b = \langle P', E', \overline{c}\rangle \in arguments, weight(a) \leq weight(b) + \beta\} \cup \{\neg acc_a\}$$

For every argument $a = \langle P, E, c\rangle$ in $arguments$ with $standard(c) =$ *beyond-reasonable-doubt*:

$$\mathcal{R}_{s_a} = \{\rightarrow \neg acc_a \mid weight(a) \leq \alpha\}$$
$$^-(acc_a) = \{arg_b \mid b = \langle P', E', \overline{c}\rangle \in arguments, weight(a) \leq weight(b) + \beta\} \vee weight(b) \geq \gamma\} \cup \{\neg acc_a\}$$

For every argument $a = \langle P, E, c\rangle$ in $arguments$ with $standard(c) =$ *dialectical-validity*:

$$\mathcal{R}_{s_a} = \emptyset$$
$$^-(acc_a) = \{arg_b \mid b = \langle P', E', \overline{c}\rangle \in arguments\}$$

Definition 3.2 (Knowledge base corresponding to a CAES). Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$ and propositional language $\mathcal{L}_{CAES}$. Then the corresponding knowledge base, in an argumentation system corresponding to C defined in Definition 3.1, is a pair $\langle \mathcal{K}, \leq'\rangle$ where:

- $\mathcal{K}_n = assumptions$,
- $\mathcal{K}_p = \mathcal{K}_a = \emptyset$,
- $\mathcal{K}_i = \mathcal{L}_{CAES} \backslash (assumptions \cup \{c \mid \langle P, E, c\rangle \in arguments\})$,
- $\leq' = \{(k, k) \mid k \in (\mathcal{K} \backslash \mathcal{K}_n)\}$.

We can now relate an argumentation theory and consequently an argumentation framework to a CAES:

Definition 3.3 (Argumentation theory corresponding to a CAES). Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$ and propositional language $\mathcal{L}_{CAES}$ the argumentation theory AT corresponding to C is a tuple $\langle AS, KB, \preceq\rangle$ where:

- AS is the argumentation system corresponding to C according to Definition 3.1,
- KB is the knowledge base in the argumentation system AS corresponding to C according to Definition 3.2,
- $\preceq = \emptyset$.

Definition 3.4 (Argumentation framework corresponding to a CAES). Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$, propositional language $\mathcal{L}_{CAES}$ and argumentation theory AT corresponding to C as given by Definition 3.3, the AF corresponding to C is the argumentation framework corresponding to AT as given by Definition 2.13.

[4] This idea is adapted from [Brewka and Woltran, 2010]).

The main difficulty in finding a translation is dealing with the ambiguity-blocking nature of Carneades, while *ASPIC*$^+$ is ambiguity-propagating. Let p, r and t be in $assumptions$ and let there be CAES arguments for q given p, for $\neg q$ given r, for s given $\neg q$ and for $\neg s$ given t. There are no exceptions and all arguments are equally strong. Then with, say, *preponderance* neither q nor $\neg q$ is acceptable so the argument for s is not applicable: hence $\neg s$ is acceptable. However, a naive, direct translation of the arguments into defeasible inference rules in *ASPIC*$^+$ would instead make no corresponding arguments acceptable. The above translation solves this by using an explicit argument node, yielding undefeated undercutters for the acceptability of q and $\neg q$, thereby yielding an undefeated undercutter for the argument for s constructed by using the argument for q, so that $\neg s$ is acceptable in *ASPIC*$^+$.

Argumentation frameworks corresponding to a CAES have the following properties.

Proposition 3.5. *Every argumentation framework corresponding to a CAES is well-founded.*

Proof (sketch)**.** By construction of the rules and contrariness relation in Definition 3.1 the only possible attack is undercutting. There are three cases: a conclusion of the form arg_b that attacks another argument on the inference rule representing acceptability acc_a, a direct undercut on the acceptability of the form: $\rightarrow \neg acc_a$ or finally an exception to an argument expressed in the form of $e_i \Rightarrow \neg app_a$. The second case cannot create cycles by construction and the other two cases will not occur due to the acyclicity of *arguments*. □

The next result follows directly from Proposition 3.5 and Theorem 2.5:

Corollary 3.6. *Every argumentation framework corresponding to a CAES according to Definition 3.4 has exactly one complete extension which is grounded, preferred and stable.*

Theorem 3.7. *Let C be a CAES, $\langle arguments, audience, standard\rangle$, $\mathcal{L}_{CAES}$ the propositional language used and let the argumentation framework corresponding to C be AF. Then the following holds:*

1. *An argument $a \in arguments$ is applicable in C iff there is an argument contained in the complete extension of AF with the corresponding conclusion arg_a.*
2. *A propositional literal $c \in \mathcal{L}_{CAES}$ is acceptable in C or $c \in assumptions$ iff there is an argument contained in the complete extension of AF with the corresponding conclusion c.*

Proof. We prove 1. and 2. by induction on the number of arguments, n, in the CAES C.

For $n = 0$, there is neither an (applicable) argument nor an acceptable proposition in C. The knowledge base KB corresponding to C will only contain axioms in $\mathcal{K}_n$ for each assumption in C and issue premises in $\mathcal{K}_i$ for other propositional literals in $\mathcal{L}_{CAES}$. The defeasible and strict rules $\mathcal{R}_d$ and $\mathcal{R}_d$ will be empty. Therefore all arguments on the basis of KB will either be an argument using an issue premise and thus not in the complete extension of the argumentation framework (CE_{AF}), or an argument containing only an axiom and therefore in CE_{AF}. So CE_{AF} contains an argument with corresponding conclusion for every assumption in C and no argument with a conclusion of the form arg_a, therefore every conclusion of an argument in CE_{AF} is an assumption, making 1. and 2. hold.

Assuming 1. and 2. hold for n arguments we consider a CAES, C, with $n + 1$ arguments. Due to acyclicity of $arguments$ there is at least one argument $a = \langle P, E, c\rangle \in arguments$ for which the conclusion c is not contained in the premises or exceptions of another argument in $arguments$. Now consider the CAES C' constructed from C by taking $arguments' = arguments\backslash\{a\}$ and let AF' be the corresponding argumentation framework. We then obtain a CAES with n arguments for which the induction hypothesis holds.

(1. $\Leftrightarrow$) We must prove that for all (not) applicable arguments b in C there is (not) an argument in CE_{AF} with conclusion arg_b. For all arguments in C' this follows from the induction hypothesis. By our selection of a, the applicability of a does not influence applicability of the arguments that were in C'. In the translation of a to *ASPIC*$^+$, corresponding arguments for arg_a will not defeat arguments in AF'. Then by the satisfaction of the directionality criterion of complete semantics [Baroni and Giacomin, 2007] it follows that all arguments acceptable in $CE_{AF'}$ are also in CE_{AF}, thus leaving correspondence of the applicability of a in C to prove. Acceptability of the premises and exceptions of a is not influenced by the applicability of a, and thus by the induction hypothesis on C' and the directionality criterion, premises and exceptions of a are acceptable in C or part of the $assumptions$ iff there is an argument contained in CE_{AF} with the corresponding conclusion. By our translation, we know that $P \Rightarrow_{app_a} arg_a$ and the set $\{e_i \Rightarrow \neg app_a \mid e_i \in E\}$ are in $\mathcal{R}_d$.

Now suppose first that a is applicable in C. Then by the induction hypothesis for all premises $p_i \in P$ there exists an argument A_i in CE_{AF}. We prove that if for $P = \{p_1, \ldots, p_n\}$ the argument $A_1, \ldots, A_n \Rightarrow_{app_a} arg_a$ also is in CE_{AF}. By conflict-freeness of CE_{AF}, no defeater of any A_i is in CE_{AF} so it suffices to prove that no argument for $\neg app_a$ is in CE_{AF}. By applicability of a and the induction hypothesis, for no $e \in E$ there exists an argument in CE_{AF} with conclusion e and thus this follows directly.

Suppose next that a is not applicable in C. Then by the induction hypothesis either not all A_i are in CE_{AF} or for some $e \in E$ an argument A_e with conclusion e is in CE_{AF}. In the first case $A = A_1, \ldots, A_n \Rightarrow_{app_a} arg_a \notin CE_{AF}$ by closure of CE_{AF} under subarguments (Proposition 6.1 of [Prakken, 2010]). In the second case A for arg_a is defeated by A_e so $A \notin CE_{AF}$ by conflict-freeness of CE_{AF}.

(2. $\Rightarrow$) If d is an assumption, then by translation $d \in \mathcal{K}_n$ and thus there is an argument A with corresponding conclusion d in CE_{AF}.

Otherwise, we must prove that if a propositional literal $d \in \mathcal{L}_{CAES}$ is acceptable in C then there is an argument contained in CE_{AF} with the corresponding conclusion d. For the CAES C' defined before, the induction hypothesis holds and therefore acceptable literals (or literals in assumptions) of C' have an argument with corresponding conclusion in $CE_{AF'}$. By our selection of a and acyclicity of $arguments$ we know that a only influences the acceptability of its conclusion and negation, c and $\overline{c}$. Then, again by the directionality criterion, we have $(2. \Rightarrow)$ left to prove for c and $\overline{c}$ in C. Moreover, if $d \neq c$ then $(2. \Rightarrow)$ trivially holds, so in the following we assume $d = c$.

Suppose a is not applicable, then by $(1.)$, no argument for arg_a will be in CE_{AF} and therefore neither an argument for c in CE_{AF}. This also prevents a from influencing acceptability of $\overline{c}$, letting $(2. \Rightarrow)$ hold.

If a is applicable, then by $(1.)$ there exists an argument A_1 with conclusion arg_a in CE_{AF}. By translation $arg_a \Rightarrow_{acc_a} c \in \mathcal{R}_d$, allowing A_1 to be extended to an argument A_2 for c. If c is acceptable in C, then its proof standard is satisfied. Then by translation there will be neither a contrary of acc_a in $^-$ nor a strict rule of the form $\rightarrow \neg acc_a \in \mathcal{R}_s$ and therefore there will be no undercutter of A_2 in CE_{AF} on the final inference. Furthermore since A_1 is in CE_{AF}, by conflict-freeness no defeater of A_1 is in CE_{AF}. Thus $A_2 \in CE_{AF}$. Similarly, if a makes the proof standard for $\overline{c}$ unsatisfiable in C, by construction of AF, A_1 will defeat any argument b with conclusion $\overline{c}$ on its inference rule arg_b. So by conflict-freeness no such argument will be in CE_{AF}, correctly preserving acceptability of $\overline{c}$.

$(2. \Leftarrow)$ Proof by contraposition. First, $d \notin assumptions$ and therefore $d \notin \mathcal{K}_n$. Similar to the proof of $(2. \Rightarrow)$, $(2. \Leftarrow)$ holds if $d \neq c$ or a is not applicable.

So assume a is applicable and $d = c$. Since c is not acceptable, the proof standard of c is not satisfied in C. Consider for example $standard(c) = \textit{clear-and-convincing}$. Then either $weight(a) \leq \alpha$ or $weight(a) \leq weight(b) + \beta$ for another applicable argument b with conclusion $\overline{c}$. Therefore the argumentation system either has $\rightarrow \neg acc_a \in \mathcal{R}_s$ or otherwise $arg_b \in {}^-(acc_a)$. Finally the AF on the basis of this argumentation system will either have an argument of the form $\rightarrow \neg acc_a$, or by applicability of b and the induction hypothesis, arg_b will be in CE_{AF} and defeats any argument using the defeasible inference acc_a. Concluding any argument constructed for the acceptability of c will be defeated and thus by conflict-freeness not in CE_{AF}.

Acceptability of $\overline{c}$ is analogous to $(2. \Rightarrow)$.

□

Finally, as in [Brewka and Gordon, 2010] we can generalise Carneades to cycle-containing structures.

Definition 3.8. Given a CAES $C = \langle arguments, audience, standard \rangle$ without the acyclicity restriction, $\mathcal{L}_{CAES}$ the propositional language used and let the argumentation framework corresponding to C be AF. Then for $s \in \{complete, preferred, grounded, stable\}$:

- An argument $a \in arguments$ is applicable in C under sceptical (credulous) s semantics iff all (some) s extensions of AF contain an argument with conclusion arg_a.
- A propositional literal $c \in \mathcal{L}_{CAES}$ is acceptable in C or $c \in assumptions$ under sceptical (credulous) s semantics iff all (some) s extensions of AF contain an argument with conclusion c.

4 Conclusion

This paper has shown that Carneades can be reconstructed through *ASPIC*$^+$ as Dung's abstract argumentation frameworks. Thus we have shown that the idea of varying proof standards for statements can be modelled within a Dungean approach, while retaining a correspondence between both through Theorem 3.7. Furthermore, addressing issues from Carneades [Gordon and Walton, 2009], the translation allows the semantics of Carneades to be generalised to argument evaluation structures that containing cycles, in a way similar to [Brewka and Gordon, 2010].

By translating Carneades into *ASPIC*$^+$, the consistency and closure results of [Prakken, 2010] can be directly applied to Carneades. it is easy to verify that argumentation frameworks generated by our translation satisfy the assumptions under which [Prakken, 2010] proves consistency and strict closure of extensions. Finally, we note that our translation enables a standard Dung semantics for an 'ambiguity blocking' non-monotonic logic ([Gordon *et al.*, 2007], section 7.1); to our knowledge, we are the first to have achieved such a result.

Acknowledgements

We thank the anonymous reviewers for their very useful comments on the submitted version of this paper.

References

[Amgoud *et al.*, 2006] L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, H. Prakken, J. van Veenen, and G.A.W. Vreeswijk. Final review and report on formal argumentation system. Deliverable D2.6, ASPIC IST-FP6-002307, 2006.

[Baroni and Giacomin, 2007] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artif. Intell.*, 171:675–700, July 2007.

[Brewka and Gordon, 2010] G. Brewka and T.F. Gordon. Carneades and abstract dialectical frameworks: A reconstruction. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 3–12. IOS Press, Amsterdam etc, 2010.

[Brewka and Woltran, 2010] G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning*, pages 102–111. AAAI Press, 2010.

[Brewka *et al.*, 2011] G. Brewka, P.E. Dunne, and S. Woltran. Relating the semantics of abstract dialectical frameworks and standard AFs. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, 2011.

[Dung, 1995] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–357, 1995.

[Gordon and Walton, 2009] T.F. Gordon and D.N. Walton. Proof burdens and standards. In G. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer US, 2009.

[Gordon *et al.*, 2007] T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artif. Intell.*, 171(10-15):875–896, 2007.

[Prakken, 2010] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument & Computation*, 1:93–124, 2010.