Argument and Computation
Vol. 3, No. 1, March 2012, 21–47

Taylor & Francis
Taylor & Francis Group

Relating Carneades with abstract argumentation via the ASPIC$^+$ framework for structured argumentation

Bas van Gijzel[a]* and Henry Prakken[b,c]

[a]*School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham, UK;* [b]*Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands;* [c]*Faculty of Law, University of Groningen, Groningen, The Netherlands*

(*Received 12 October 2011; final version received 25 January 2012*)

Carneades 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. In this article, the two formalisms are formally related by translating Carneades into the ASPIC$^+$ framework for structured argumentation. Since ASPIC$^+$ is defined to generate Dung-style abstract argumentation frameworks, this in effect translates Carneades graphs into abstract argumentation frameworks. For this translation, we prove a formal correspondence and show that certain rationality postulates hold. It is furthermore proved that Carneades always induces a unique Dung extension, which is the same in all of Dung's semantics, allowing us to generalise Carneades to cycle-containing structures.

Keywords: abstract argumentation; structured argumentation; non-monotonic reasoning; proof standards; burden of proof

1. Introduction

Argumentation involves the construction of arguments in favour of and against statements, selecting the acceptable arguments, and in the end determining which statements hold. How arguments support their conclusion depends on the knowledge they use and the inference rules they apply, so any full theory of argument evaluation should take the structure and content of arguments into account. One way to do so is to define a defeat relation between arguments that takes into account the structure and content of arguments and (if available) information on their relative strength. This approach thus results in an abstract argumentation framework in the sense of Dung (1995), so that the full theory of abstract argumentation can be applied. Two frameworks for structured argumentation that are designed following this approach are assumption-based argumentation (Bondarenko, Dung, Kowalski, and Toni 1997, Dung, Kowalski, and Toni 2009) and ASPIC$^+$ (Prakken 2010). In fact, Prakken (2010) shows that assumption-based argumentation can be translated into ASPIC$^+$ as a special case.

However, there have also been advances in structured argumentation that diverge from this approach. A recent application in legal reasoning is the Carneades argumentation system, both a logical model (Gordon, Prakken, and Walton 2007, Gordon and Walton 2009b) and a software toolbox for structured argument evaluation, construction and visualisation (Gordon 2010). Carneades innovates models of structured argumentation by allowing varying proof standards to be assigned to individual propositions. It is claimed that this allows for a more natural account of

*Corresponding author. Email: psxbv@nottingham.ac.uk

ISSN 1946-2166 print/ISSN 1946-2174 online

http://dx.doi.org/10.1080/19462166.2012.661766
http://www.tandfonline.com

reasoning under burden of proof than existing formalisms for structured argumentation, in which proof standards are defined globally (Atkinson and Bench-Capon 2007, Bench-Capon, Doutre, and Dunne 2007). This makes the Carneades formalism potentially very attractive, as signified by the large number of citations due to its proof standards.

Recently, Brewka and Gordon (2010a) translated Carneades into Brewka and Woltran's (2010b) abstract dialectical frameworks. Moreover, Brewka, Dunne, and Woltran (2011) have proved a formal correspondence between abstract dialectical frameworks and Dung's abstract argumentation frameworks. By combining these results, a formal relation between Carneades and Dung's semantics can be obtained. However, this relation is rather indirect. In this paper, we therefore take a different approach, by translating Carneades into the ASPIC$^+$ framework. Since ASPIC$^+$ is defined to generate abstract argumentation frameworks, which are the input of Dung's approach, a translation of Carneades into ASPIC$^+$ provides a more direct way to translate Carneades' graphs into Dung's frameworks. Futhermore, we will 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 (CAES) to cycle-containing structures, addressing an important issue left for future research by Gordon and Walton (2009b). An additional advantage of translating Carneades to ASPIC$^+$ is that the results of Prakken (2010) on the rationality postulates of Caminada and Amgoud (2007) can be shown to hold for the translation.

This article is structured as follows: first we will review the necessary background in Section 2, namely, Dung's abstract argumentation frameworks and Prakken's ASPIC$^+$ framework. In Section 3, we introduce Carneades in a different formulation than that of Gordon and Walton (2009b), keeping only the relevant stage-specific parts. In Section 4, we then establish a formal relation between Carneades and Dung's frameworks by developing a translation and proving formal results.[1] Finally, in Section 5, we conclude and discuss related and future work.

2. Background

In this section, we briefly review Dung's abstract argumentation frameworks (Dung 1995), followed by an introduction to the structured argumentation frameworks of Prakken (2010). For a recent in-depth introduction to Dung's frameworks, see Baroni and Giacomin (2009).

2.1. *Abstract argumentation frameworks*

In 1995, Dung introduced his seminal theory of argumentation, with which he, due to its abstract nature, was able to model logic programming and several of the contemporary approaches to non-monotonic reasoning. By showing that these forms of reasoning can be represented as a form of argumentation, the relationship between these approaches was clarified.

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 $(C, B) \in Def$.

(3) The *characteristic function* of an AF, F_{AF} is a function such that
   - $F_{\mathrm{AF}} : 2^{Args} \rightarrow 2^{Args}$,
   - $F_{\mathrm{AF}}(S) = \{A | A \text{ is acceptable w.r.t. to } S\}$.

(4) A conflict-free set of arguments S is *admissible* iff every argument $A \in S$ is acceptable w.r.t. S, that is, $S \subseteq F_{\mathrm{AF}}(S)$.

DEFINITION 2.3 (Extensions) Given a conflict-free set of arguments S and an argumentation framework AF :

- S is a *complete extension* iff $S = F_{\mathrm{AF}}(S)$.
- 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 $Args \backslash S$.

Finally, we use Dung's definition of a *well-founded argumentation framework*. A well-founded AF is an AF without cycles or an infinite defeating chain of arguments. We will later prove the translation of Carneades to be well-founded.

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 (1995)) Every well-founded argumentation framework has exactly one complete extension which is grounded, preferred and stable.

2.2. *Structured argumentation frameworks*

The abstract argumentation frameworks by Dung (1995) keep the structure and nature of arguments and the attack relation unspecified. This allows for general reasoning about the acceptability status of arguments, but provides no guidance for the modelling of actual argumentation problems. Other research has therefore taken a structured approach to argumentation (Amgoud 2005, Gordon et al. 2007).

Structured argumentation frameworks (ASPIC$^+$) as defined by Prakken (2010) are a further development of the ASPIC framework as defined by Amgoud et al. (2006). Prakken's frameworks instantiate[3] the abstract argumentation model of Dung, defining the internal structure of arguments, defining multiple types of attack and adding preferences to the attack relation, and resulting in a defeat relation.

The basic building block of a structured argumentation framework is the concept of an *argumentation system*, extending the standard notion of a proof system. In argumentation systems, the logical language is left unspecified except for the existence of a contrariness relation (generalisation of logical negation to asymmetric conflict). Inference rules are divided into *strict* and *defeasible* rules, respectively, of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$. Strict rules are interpreted as "if the *antecedents* $\varphi_1, \ldots, \varphi_n$ hold, then *without exception* the *consequent* φ holds", where for defeasible rules this *presumably* holds. Finally, the relative strength of defeasible rules can be determined by means of a partial preorder.

DEFINITION 2.6 (Argumentation system (Prakken 2010, Def. 3.1)) An *argumentation system* is a tuple $\text{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \leq \rangle$ where

(i) $\mathcal{L}$ is a logical language,
(ii) $\bar{}$ 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*, that is, $\varphi \in \overline{\psi}$ and $\psi \in \overline{\varphi}$.

(iii) $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible inference rules ($\mathcal{R}_d$) such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$,
(iv) $\leq$ is a partial preorder on $\mathcal{R}_d$.

Since this definition leaves the nature of the logical language and the inference rules largely unspecified, it is possible to reformulate specific argumentation systems as instances of ASPIC$^+$. For example, Prakken (2010) has shown that assumption-based argumentation (Bondarenko et al. 1997, Dung, Mancarella, and Toni 2007), a structured argumentation approach using assumptions from which conclusions are drawn using strict inference rules, is a special case of ASPIC$^+$, and Modgil and Prakken (2011) have proved the same for variants of classical argumentation (cf. Besnard and Hunter 2008).

With the argumentation system defined, we can now look at the construction of arguments by means of a knowledge base in an argumentation system. The set of rules contains both a strict and defeasible kind and the knowledge base can be inconsistent. In addition to the possible inconsistency, the knowledge base also contains four different types of facts, inspired by a similar distinction of Gordon et al. (2007). Similar to the axioms in deductive logic, there are (unattackable) premises called *necessary axioms* ($\mathcal{K}_n$), (attackable) *ordinary premises* ($\mathcal{K}_p$), *assumptions* ($\mathcal{K}_a$) – which are a weak type of premise always defeated by an attack – and *issues* ($\mathcal{K}_i$) – which are premises that are not acceptable unless backed by further argument.

DEFINITION 2.7 (Knowledge base (Prakken 2010, Def. 3.5)) A *knowledge base* in an argumentation system $\langle \mathcal{L}, \bar{}, \mathcal{R}, \leq \rangle$ is a pair $\langle \mathcal{K}, \leq' \rangle$ where $\mathcal{K} \subseteq \mathcal{L}$ and $\leq'$ is a partial preorder on $\mathcal{K} \backslash \mathcal{K}_n$. 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.

With the knowledge base and inference rules defined as above, the construction of arguments can be defined by adopting Vreeswijk's (1993, 1997), definition of an argument. The smallest argument is simply a fact from the knowledge base. More complex arguments can be constructed by chaining inference rules on previous arguments, resulting in an argument in tree form (containing sub-arguments).

DEFINITION 2.8 (Arguments (Prakken 2010, Def. 3.6)) An *argument A* on the basis of a knowledge base $\langle \mathcal{K}, \leq' \rangle$ in an argumentation system $\langle \mathcal{L}, \bar{}, \mathcal{R}, \leq \rangle$ is

(1) ϕ if $\phi \in \mathcal{K}$ with:
$Prem(A) = \{\phi\}$,
$Conc(A) = \phi$,
$Sub(A) = \{\phi\}$,
$DefRules(A) = \emptyset$,
$TopRule(A) =$ undefined.
(2) $A_1, \ldots, A_n \rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict rule $Conc(A_1), \ldots, Conc(A_n) \rightarrow \psi$ in $\mathcal{R}_s$,
$Prem(A) = Prem(A_1) \cup \cdots \cup Prem(A_n)$,
$Conc(A) = \psi$,

$Sub(A) = Sub(A_1) \cup \cdots \cup Sub(A_n) \cup \{A\}$,
$Def\ Rules(A) = Def\ Rules(A_1) \cup \cdots \cup Def\ Rules(A_n)$,
$TopRule(A) = Conc(A_1), \ldots, Conc(A_n) \rightarrow \psi$.

(3) $A_1, \ldots, A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a defeasible rule $Conc(A_1), \ldots, Conc(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$,
$Prem(A) = Prem(A_1) \cup \cdots \cup Prem(A_n)$,
$Conc(A) = \psi$,
$Sub(A) = Sub(A_1) \cup \cdots \cup Sub(A_n) \cup \{A\}$,
$Def\ Rules(A) = Def\ Rules(A_1) \cup \cdots \cup Def\ Rules(A_n) \cup$
$\{Conc(A_1), \ldots, Conc(A_n) \Rightarrow \psi\}$,
$TopRule(A) = Conc(A_1), \ldots, Conc(A_n) \Rightarrow \psi$.

EXAMPLE 2.9 Given an argumentation system and a knowledge base in that argumentation system with the following rules and facts (where $q, r \rightarrow s$ means with q and r derived, derive s):

$$\mathcal{R}_s = \{z, s \rightarrow t; q, r \rightarrow s\},$$
$$\mathcal{R}_d = \{p, u \Rightarrow q\},$$
$$\mathcal{K}_n = \{p; z\},$$
$$\mathcal{K}_p = \{u\},$$
$$\mathcal{K}_a = \{r\},$$
$$\mathcal{K}_i = \{s\}.$$

An argument for t can be constructed by using an issue premise from $\mathcal{K}_i$, as seen in Figure 1. (The type of premise is used as superscript.)

Arguments using issue premises will result in not being acceptable. Therefore, to produce a possibly acceptable argument for t after evaluation, we will want to derive an argument for s instead of using the issue premise. Such an argument for t can be seen in Figure 2. Here double lines indicate a defeasible inference. This argument contains several sub-arguments which can formally be written as follows:

$$A_1 : p, \quad A_5 : A_1, A_2 \Rightarrow q,$$
$$A_2 : u, \quad A_6 : A_5, A_3 \rightarrow s,$$
$$A_3 : r, \quad A_7 : A_6, A_4 \rightarrow t.$$
$$A_4 : z,$$

Here A_7 is the argument from Figure 2.

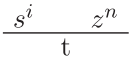

Figure 1. An argument for t using an issue premise.

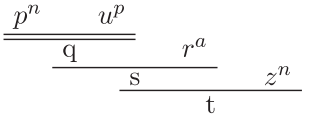

Figure 2. Another argument for t.

Now we can define the notion of an argumentation theory.

DEFINITION 2.10 (Argument theories (Prakken 2010, Def. 3.11)) 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.

With the internal structure of arguments defined, it is now possible to distinguish between types of attack.

Undermining attack is an attack on the premises on an argument and is the only attack possible in the context of strict rules. An *undercutting attack* is an attack on the (defeasible) inference step and is a way to provide "exceptions to the rule". Finally, *a rebutting attack* is done by constructing a contrary or contradictory conclusion for the attacked argument's (sub)conclusion.

DEFINITION 2.11 (Types of attack (Prakken 2010, Def. 3.16, 3.11, 3.14))

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

The types of attack can be combined with an argument ordering to define the notion of defeat. Similar to contraries, an undercutting attack does not take the ordering into account and hence always results in defeat. Intuitively, the undercutter contains an argument for an exception to the rule of the attacked argument, otherwise an undercutter and the attacked argument using that rule could be in the same extension.

DEFINITION 2.12 (Types of defeat (Prakken 2010, Def. 3.19, 3.20))

- 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 $A \not\prec \phi$.

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

DEFINITION 2.13 (Defeat (Prakken 2010, Def. 3.21)) 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.

To deal with issue premises, an argument is acceptable only if it contains no issue premises; therefore, changing Definition 2.2 to the following.
An argument $A \in Args$ is acceptable with respect to a set S of arguments, or alternatively S *defends* A, iff A *contains no issue premises and* for all arguments $B \in S$: if $defeats(B, A)$ holds then there is a $C \in S$ for which $defeats(C, B)$ holds.[4]

With arguments and the defeat relation fully defined, it is possible to link the argumentation theories of the structured approach to Dung's abstract argumentation frameworks, thereby formally making the correspondence between the structured and abstract approach.

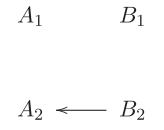

Figure 3. Corresponding argumentation framework.

DEFINITION 2.14 (Argumentation framework (Prakken 2010, Def. 3.22)) An *abstract argumentation framework AF corresponding to an argumentation theory AT* is a pair $\langle Args, defeats\rangle$ such that

- *Args* is the set of arguments on the basis of AT as defined by Definition 2.8,
- *defeats* is the relation on *Args* given by Definition 2.13.

Finally, the acceptability of conclusions (of a mathematical language $\mathcal{L}$) is defined in the corresponding argumentation framework.

DEFINITION 2.15 (Acceptability of conclusions (Prakken 2010, Def. 3.23)) For any semantics S and for any argumentation framework AF and formula $\phi \in \mathcal{L}_{\text{AF}}$:

(1) ϕ is *sceptically S-acceptable* in AF if and only if all S-extensions of AF contain an argument with conclusion ϕ;
(2) ϕ is *credulously S-acceptable* in AF if and only if there exists an S-extensions of AF that contain an argument with conclusion ϕ.

EXAMPLE 2.16 Given an argumentation system and a knowledge base in that argumentation system with the following rules and facts (where $q, r \rightarrow \neg r_1$ means that given q and r, rule r_1 does not apply, that is, undercut):

$$\begin{aligned}
\mathcal{R}_d &= \{bird \Rightarrow_r fly;\ penguin \Rightarrow \neg r\},\\
\mathcal{K}_p &= \{bird;\ penguin\},\\
\mathcal{K}_n &= \mathcal{K}_a = \mathcal{K}_i = \mathcal{R}_s = \preceq = \preceq' = \emptyset,\\
^- &= \{(r, \neg r)\}.
\end{aligned}$$

The arguments on the basis of this knowledge base are the following:

$$\begin{aligned}
&A_1 : bird, && B_1 : penguin,\\
&A_2 : A_1 \Rightarrow_r fly, && B_2 : B_1 \Rightarrow \neg r.
\end{aligned}$$

The defeat relation on basis of the argumentation system and knowledge base (independent of the ordering $\preceq$), $defeats = \{(B_2, A_2)\}$, can be visualised together with the arguments constructed (Figure 3).

For this argumentation framework, we have one complete (and thus one grounded, preferred and stable) extension, namely $E = \{A_1, B_1, B_2\}$. We can see that bird and penguin are sceptically and credulously justified in E.

3. Stage-specific Carneades

Carneades is a formal model of argumentation incorporating both static and dynamic aspects of argumentation. In Carneades, argumentation is seen as a dialogical process, determining the

acceptability of arguments by applying proof standards, where the assignment of proof standards to arguments is determined by the various proof burdens. The version of Carneades that we will discuss is the stage-specific part of the most recent version by Gordon and Walton (2009b). It is of primarily interest to us because of the existing reduction to abstract dialectical frameworks Brewka and Gordon 2010a; Brewka and Woltran 2010b). Contrary to how Carneades has been introduced in the literature (Gordon et al. 2007; Gordon and Walton 2009b), but similar to Brewka and Gordon (2010a), we will introduce proof standards to be part of the static, evaluative part of Carneades.

We will start with an introduction to Carneades' concept of arguments. Similar to structured abstract argumentation frameworks introduced in the previous section, arguments in Carneades are not left abstract but given certain structure. Arguments are constructed by linking premises and exceptions to a conclusion. The Arguments *pro* and *con* and conclusions are later aggregated and evaluated through proof standards.

DEFINITION 3.1 (Arguments (Gordon and Walton 2009b, Def. 1)) 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*. For simplicity, c and all members of P and E must be literals, that is, either an atomic proposition or a negated atomic proposition. 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 (2010a), we therefore only consider *stage-specific CAES*. To define them, the concepts of an *audience* and an *acyclic set* of arguments must be introduced.

DEFINITION 3.2 (Audience (adapted from Gordon and Walton 2009b, Def. 3)) 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[5] mapping arguments to real numbers in the range $0.0 \cdots 1.0$, representing the relative weights assigned by the audience to the arguments.

Carneades is defined with an acyclic set of arguments to simplify the definitions and time needed for the evaluation of arguments.[6] In the original definition by Gordon and Walton (2009b), sets of arguments were restricted to be acyclic in the sense that the chains of arguments $a_1, \ldots, a_n$ constructable from the set should be acyclic. A chain is constructed by sequencing arguments, intuitively giving the possibility to link the conclusion of an argument to a premise of a later argument. Acyclicity of a chain implied that a premise of an argument a_i could not be used in a conclusion of an argument a_j later in the chain. However, this results in a non-well-founded definition of acceptability in Carneades due to possible cycles in exceptions in arguments, for example, consider *arguments* $= \{a, b\}$ with $a = \langle \emptyset, \{p\}, q\rangle$ and $b = \langle \emptyset, \{q\}, p\rangle$. This is probably a small oversight, contrasting the correct definitions in Gordon et al. (2007).

Instead, we will use the more general concept of a dependency graph to determine acyclicity of a set of arguments, based on the definition of Brewka and Gordon (2010a).

DEFINITION 3.3 (Acyclic set of arguments (adapted from Brewka and Gordon 2010a, Def. 2.1)) 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 link to node q whenever p depends on q in the sense that there is an argument pro or con p that has q or $\bar{q}$ in its set of premises or exceptions.

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

DEFINITION 3.4 (Stage-specific CAES (adapted from Gordon and Walton 2009b, Def. 4)) A (*stage-specific*) CAES is a tuple $\langle arguments, audience, standard \rangle$, where the argument is an acyclic set of arguments, *audience* is an audience and *the standard* is a total function mapping of literals in $\mathcal{L}$ to their applicable proof standards.

In the (legal) dialogues that Carneades models, proof burdens, such as the burden of production or the burden of persuasion, can be assigned to a propositional literal. Intuitively, an assigned proof burden obliges a participant of the dialogue to provide proof for that proposition, under the condition that the proof satisfies a certain standard of proof. In the model, the assignment of the proof standard is determined by the function *standard* provided in the previous definition. This proof standard can be evaluated in a specific stage and thus can be taken as a static part of Carneades. Proof standards included in the Carneades model, originating from the work of Freeman and Farley (1996) and Farley and Freeman (1995), are *a scintilla of evidence*, *preponderance of the evidence*, *clear-and-convincing evidence*, *beyond-reasonable-doubt* and finally *dialectical-validity*. A proof standard is a function that given a literal p aggregates arguments pro and con p and evaluates to true or false depending on a specific audience.[7]

DEFINITION 3.5 (Proof standard (Gordon and Walton 2009b, Def. 4)) 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, we can define the acceptability of a literal. The acceptable set of literals can be seen as the collection of literals for which the proof standard is satisfied.

DEFINITION 3.6 (Acceptability of literals (adapted from Gordon and Walton 2009b, Def. 5)[8]) Given $C = \langle arguments, audience, standard \rangle$ 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 3.7 (Applicability of arguments (adapted from Gordon and Walton 2009b, Def. 6)) 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 [$\bar{p}$ is not an assumption and p is acceptable in C] and
- $e \in E$ implies e is not an assumption of the *audience* and [$\bar{e}$ is an assumption or e is not acceptable in C].

Now we can turn to the definition of Carneades' proof standards. There is one subtle matter concerning the first proof standard, *scintilla of evidence*. A literal that is assigned scintilla of evidence as its proof standard obliges the party who puts forward an argument for this literal to produce an applicable argument. So to satisfy the scintilla of evidence standard only an applicable

argument needs to be constructed; contradictory arguments do not influence the acceptability, thus allowing both p and its contradiction $\bar{p}$ to be acceptable in a given CAES.

DEFINITION 3.8 (Proof standards) Given a CAES $C = \langle arguments, audience, standard \rangle$ and a literal p in $\mathcal{L}$:

(i) $scintilla(p, arguments, audience) = true$ iff there exists at least one applicable argument pro p in *arguments*,
(ii) $preponderance(p, arguments, audience) = true$ 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,
(iii) $clear\text{-}and\text{-}convincing(p, arguments, audience) = true$ iff there is an applicable argument A, pro p for which:
- $preponderance(p, arguments, audience)$ holds,
- 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 threshold β.

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

The theory of a (stage-specific) CAES is constructed by combining the acceptable literals in that CAES with the assumptions of the audience. In Carneades, the propositional logic is assumed as the logical language, so the theory is taken to be closed under propositional inferences.

DEFINITION 3.9 (Theory of a CAES) Let $C = \langle arguments, audience, standard \rangle$ be a CAES. The *theory* of C is the deductive closure, in propositional logic, of the union of *assumptions* and the set of literals acceptable in C.

EXAMPLE 3.10 Let $\alpha = 0.3$, $\beta = 0.3$ and $\gamma = 0.6$. Consider a CAES $C = \langle arguments, audience, standard \rangle$ and $audience = \langle assumptions, weight \rangle$ with

$$
\begin{aligned}
arguments &= \{a_1, a_2, a_3, a_4\},\\
a_1 &= \langle \{p_1, p_2\}, \{e_1\}, c\rangle, a_2 = \langle \{p_2, p_3\}, \{e_2\}, \neg c\rangle,\\
a_3 &= \langle \{p_2\}, \{e_3\}, \neg c\rangle,\ a_4 = \langle \emptyset, \{e_4\}, \neg c\rangle,\\
assumptions &= \{p_1, p_2, e_4\},\\
weight(a_1) &= 0.4;\ weight(a_2) = 0.9;\ weight(a_3) = 0.5;\ weight(a_4) = 0.6,\\
standard(c) &= preponderance,\ standard(\neg c) = clear\text{-}and\text{-}convincing.
\end{aligned}
$$

We can visualise these arguments (arrows denote premises/inferences and open circles denote exceptions) (Figure 4).

Then we have that argument a_2 is not applicable because $p_3 \notin assumptions$ and p_3 is not acceptable because there is no argument with p_3 as the conclusion. Argument a_4 is not applicable because $e_4 \in assumptions$. Argument a_1 and a_3 are applicable.

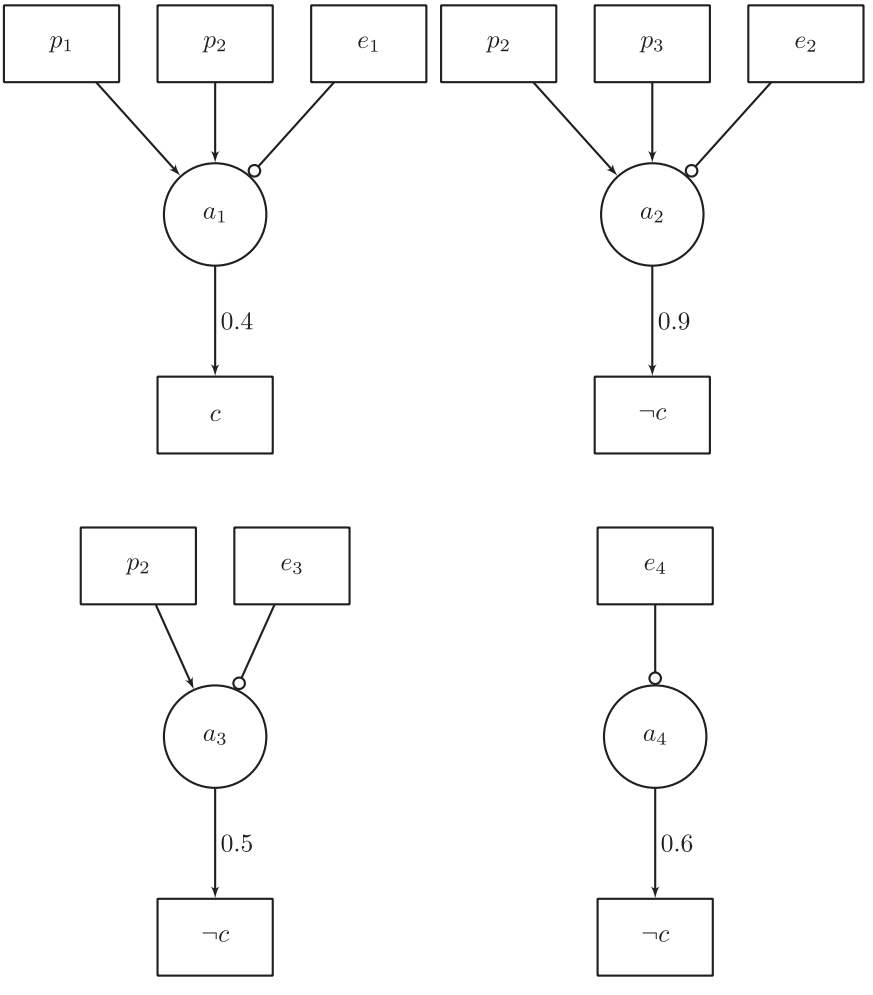

Figure 4. Arguments in Carneades.

The conclusion c (of argument a_1) is not acceptable because $standard(c) = preponderance$ and $weight(a_1) \not> weight(a_3)$ while a_3 is an applicable con argument for c. The conclusion $\neg c$ is also not acceptable because $standard(\neg c) = clear\text{-}and\text{-}convincing$ and when considering the argument a_4 it holds for the applicable con argument a_1 that: $weight(a_3) \not> weight(a_1) + \beta$ (although $weight(a_3) > \alpha$).

4. Relation between Carneades and Dung's frameworks

In the next subsections, Carneades will be related to Dung's model. First we will translate the static, stage-specific part of Carneades to structured argumentation frameworks. We will study properties of our translation, for instance, proving correspondence results and showing that the translation does not violate rationality postulates. Then, we will generalise our translation of Carneades, allowing to lift the restriction of acyclicity on a CAES. Finally, we will relate our translation to the existing translations of Carneades to ADF's (abstract dialectical framework) and the defeasible logic (Brewka and Gordon 2010a; Governatori 2011).

4.1. *Translation of stage-specific Carneades*

We will start with relating the premises and exceptions of the arguments in Carneades to a knowledge base in an argumentation system. The *assumptions* of the audience in a CAES are propositional literals which are unattackable and furthermore, as can be seen in Definition 3.9, part of the logical theory. Combining these characteristics, assumptions in Carneades are closely related to the concept of axioms in a knowledge base and thus will be modelled as necessary axioms, $\mathcal{K}_n$, in our knowledge base. Next, the use of conclusions as a premise in a later argument is similar to the chaining of sub-arguments to construct more complex arguments and can, therefore, be handled by the argument generation part of ASPIC$^+$. Finally, a premise with no backing, also called an *issue premise* in Gordon et al. (2007), maps exactly to the issue premises in our knowledge base.

Combining these insights, we can now define the knowledge base corresponding to a CAES.

DEFINITION 4.1 (Knowledge base corresponding to a CAES) Given a CAES $C = \langle \mathit{arguments}, \mathit{audience}, \mathit{standard} \rangle$ with $\mathit{audience} = \langle \mathit{assumptions}, \mathit{weight} \rangle$ and propositional language $\mathcal{L}_{\text{CAES}}$. Then the knowledge base in an argumentation system corresponding to C is a pair $\langle \mathcal{K}, \leq' \rangle$ where

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

There is no need to differentiate in the strength of premises, making our preference relation on premises just the reflexive closure on non-axiom premises.

As shown in our visualisation of Carneades' arguments in Example 3.10, the link between the premises, the exceptions and the conclusion is a two-part inference. The first part – applicability of the argument – is solely determined by the acceptability of the premises and exceptions. The second step – acceptability of the conclusion – requires the argument to be applicable and furthermore to satisfy the demands of the proof standard that is assigned to the conclusion.

So for every argument $a = \langle P, E, c \rangle$ in a CAES, a defeasible rule going from the premises to the applicability of the argument is added, $P \Rightarrow_{app_a} arg_a$, saying that if P then a is applicable.[9] The other inference is represented by a defeasible rule $arg_a \Rightarrow_{acc_a} c$, saying that if a is applicable, its conclusion is acceptable. As before, app_a and acc_a are rule names, which will need to be added to the language, $\mathcal{L}$, of the CAES (rule names are assumed to be disjoint with $\mathcal{L}$).

Exceptions in Carneades' arguments express exceptions to inferring the conclusion. If we have an argument containing an exception that is acceptable or assumed by the audience, then that argument is made inapplicable, so the argument cannot make the conclusion acceptable. Given an acceptable argument containing exception $\bar{p}$, it is not implied that p can be assumed to be true; so two arguments with conflicting exceptions can both be acceptable. This use of exceptions, similar to the concept of justifications in default logic (Reiter 1980), implies that negations of exceptions cannot be modelled as an assumption, but instead need to be modelled as an undercutter to the inference rule. So in our translation of argument a, for each exception $e \in E$, an undercutter $e \Rightarrow \neg app_a$ is added to $\mathcal{R}_d$.

Although it might seem natural to include the negation relation of Carneades into the contrariness relation of the corresponding argumentation system, this does not actually work. With scintilla of evidence as a proof standard that can determine acceptable literals of a CAES, both p and $\bar{p}$ are allowed to be acceptable, for example, $\mathit{arguments} = \{\langle \emptyset, \emptyset, p \rangle, \langle \emptyset, \emptyset, \neg p \rangle\}$. It is furthermore possible to construct an acceptable argument for $\neg c$ while $c \in \mathit{assumptions}$.[10] To retain the properties of the original Carneades system, the negation relation of Carneades will, therefore,

not be imported into the contrariness relation. Instead the contrariness relation in our argumentation system is used 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 4.2 (Argumentation system corresponding to a CAES) Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$ and propositional language $\mathcal{L}_{\text{CAES}}$, the corresponding argumentation system, AS, is a tuple $\langle \mathcal{L}, ^-, \mathcal{R}, \leq\rangle$ where

- $\mathcal{L} = \mathcal{L}_{\text{CAES}} \cup$ 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*:

$$\begin{aligned} \mathcal{R}_{d_a} &= \{P \Rightarrow_{app_a} arg_a; arg_a \Rightarrow_{acc_a} c\}\cup \\ &\quad \{e_i \Rightarrow \neg app_a \mid e_i \in E\} \\ ^-(app_a) &= \{\neg app_a\}. \end{aligned}$$

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 $standard(c) = preponderance$:

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

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

$$\begin{aligned} \mathcal{R}_{s_a} &= \{\rightarrow \neg acc_a | weight(a) \leq \alpha\}, \\ ^-(acc_a) &= \{arg_b | b = \langle P', E', \bar{c}\rangle \in arguments, weight(a) \leq weight(b) + \beta\} \\ &\quad \cup \{\neg acc_a\}. \end{aligned}$$

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

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

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

$$\begin{aligned} \mathcal{R}_{s_a} &= \emptyset, \\ ^-(acc_a) &= \{arg_b | b = \langle P', E', \bar{c}\rangle \in arguments\}. \end{aligned}$$

To illustrate the translation of one proof standard, notice that in a CAES, an argument a with standard *clear-and-convincing evidence* is unacceptable if either $weight(a) \leq \alpha$, or there is a

contradictory applicable argument b for which $weight(a) \leq weight(b) + \beta$ holds. This is then translated by extending the set of contraries for the acceptability, acc_a, with an arg_b for every contradictory argument b for which $weight(a) \leq weight(b) + \beta$ holds. If the weight of a is less than α, there is also an inference $\rightarrow \neg acc_a$ added to the strict rules, R_s, that together with the contrary, $\neg acc_a$, will undercut the acceptability.

Having built up the corresponding argumentation system, we can now relate an argumentation theory and consequently an argumentation framework to a CAES.

DEFINITION 4.3 (Argumentation theory corresponding to a CAES) Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$ and propositional language $\mathcal{L}_{\text{CAES}}$ the argumentation theory AT corresponding to C is a tuple $\langle \text{AS}, \text{KB}, \preceq\rangle$ where

- AS is the argumentation system corresponding to C according to Definition 4.2,
- KB is the knowledge base in the argumentation system AS corresponding to C according to Definition 4.1,
- $\preceq = \emptyset$.

DEFINITION 4.4 (Argumentation framework corresponding to a CAES) Given a CAES $C = \langle arguments, audience, standard\rangle$ with $audience = \langle assumptions, weight\rangle$, propositional language $\mathcal{L}_{\text{CAES}}$ and argumentation theory AT corresponding to C as given by Definition 4.3, the AF corresponding to C is the argumentation framework corresponding to AT as given by Definition 2.10.

To demonstrate our translation, we will show in detail how the CAES in Example 3.10 can be translated into its corresponding argumentation system, generating the corresponding argumentation framework.

EXAMPLE 4.5 First consider how the knowledge base in our argumentation system would correspond to the CAES given in Example 3.10. We have $K_n = assumptions = \{p_1, p_2, e_4\}$, while the other premises that are not a conclusion nor an assumption would be an issue premise, thus giving $K_i = \{p_3, e_1, e_2, e_3\}$.

Next we define the rules, $\mathcal{R}$, of the corresponding argumentation system. Every argument has a corresponding rule for applicability and for the acceptability of the conclusion, arguments containing an exception will have a corresponding inference rule, generating an undercutter and finally rules with the proof standard *clear-and-convincing* or *beyond-reasonable-doubt* can have a strict rule undercutting the acceptability, if the weight of the argument is below α. For example, argument a_2 will generate

$$R_{d_{a_2}} = \{\, p_2, p_3 \Rightarrow_{app_{a_2}} arg_{a_2}; arg_{a_2} \Rightarrow_{acc_{a_2}} \neg c; e_2 \Rightarrow \neg app_{a_2}\},$$
$$R_{s_{a_2}} = \emptyset.$$

Note that the set of strict rules is empty, because $weight(a_2) = 0.9 > 0.3 = \alpha$.

Given the previous rules, the structured arguments corresponding to the Carneades arguments can be visualised as given in Figure 5. These arguments contains several (sub)arguments. For example, the first argument can formally be written as follows:

$$A_1 : p_1, \quad A_3 : A_1, A_2 \Rightarrow_{app_{a_1}} arg_{a_1},$$
$$A_2 : p_2, \quad A_4 : A_3 \Rightarrow_{acc_{a_1}} c.$$

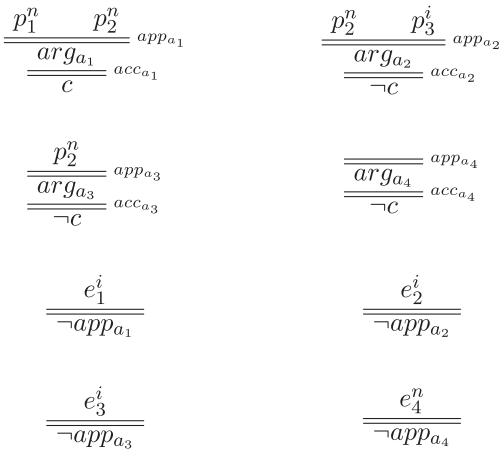

Figure 5. Structured arguments corresponding to Example 3.10.

What remains is the translation of the defeat relation, for which we will consider argument A_4, related to a_1 in the CAES. The argument A_4 first of all is undercut on app_{a_1}, by means of the argument $e_1 \Rightarrow \neg app_{a_1}$. The other attackers are undercutters on the acceptability, acc_{a_1}. The proof standard of the conclusion c of a_1 is *preponderance*, while the weight of a_1 minus β is less than the other arguments in the CAES, so by the translation, the argument nodes arg_{a_2}, arg_{a_3} and arg_{a_4} will all be a contrary of acc_{a_1}. Then the (sub)arguments $p_2, p_3 \Rightarrow_{app_{a_2}} arg_{a_2}$, $p_2 \Rightarrow_{app_{a_3}} arg_{a_3}$ and $\Rightarrow_{app_{a_4}} arg_{a_4}$ will undercut A_4 on acc_{a_1}. Since the translation does not consider preferences, every attack that will be made will result in a defeat.

Although there are some defeaters present that do not directly correspond to the original CAES, that is, the exception $e_1 \Rightarrow \neg app_{a_1}$ and the undercutter arg_{a_2}, this is not a problem since the arguments will not be deemed acceptable due to their issue premise. Thus in the end, the acceptable arguments will only be the non-issue premises, p_1, p_2 and e_4, the arguments for the applicability of a_1, a_3 and the argument for non-applicability of a_4 ($\neg app_{a_4}$), which is exactly what we want.

4.2. *Translation properties*

Now that we have defined the argumentation framework corresponding to a CAES, we can look at some interesting properties of the translation.

4.2.1. *Well-foundedness*

First of all, we will show that an argumentation framework corresponding to a CAES contains no cycles and therefore is actually a well-founded argumentation framework. From this property and a proposition proved by Dung, we can then immediately deduce that every argumentation framework corresponding to a CAES induces a unique Dung extension.

PROPOSITION 4.6 Every argumentation framework corresponding to a (finite) CAES according to Definition 4.4 is well-founded.

The next result follows directly from Proposition 4.6 and Theorem 2.5:

COROLLARY 4.7 Every argumentation framework corresponding to a CAES according to Definition 4.4 has exactly one complete extension which is grounded, preferred and stable.

Here we can see that contrary to the claim of Brewka and Gordon (2010a) – namely that modelling Carneades in Dung's approach could not be done cycle-free – we have proved that an argumentation framework corresponding to a CAES *is* well-founded and thus cycle-free. This means that the corresponding argumentation frameworks always induce a unique Dung extension which is the same in all Dung's semantics. Carneades' semantics is therefore essentially a single-status assignment approach.

We have seen that in Carneades the defeat relation that is generated through the translation depends on the audience and the proof standards. This use of audience is very similar to (and inspired by) the approach taken in value-based argumentation frameworks (Bench-Capon et al. 2007). More interestingly, just as the uniqueness of preferred extensions in VAF's (value-based argumentation framework) with respect to a single audience is guaranteed, in the translation of Carneades there is also a unique complete extension.

4.2.2. *Computational complexity*

The time to compute the extension of a well-founded argumentation framework can be determined to quadratic, by verifying that it is possible to topologically sort the acyclic dependency graph in $\Theta(|V| + |E|)$ (cf. Cormen, Leiserson, Rivest, and Stein 2001, pp. 549–552) and by checking that it is possible to compute the grounded extension in $\Theta(|V| + |E|)$ (by computing acceptability in order of dependency). We can, therefore, deduce that if the translation is polynomial, evaluating a CAES through our translation is also polynomial. Although this might seem immediate from our translation, there are some subtleties in the actual step that generates arguments. ASPIC$^+$ only declaratively states which arguments are to be generated from the argumentation system, but with a naive implementation/algorithm, the argumentation system corresponding to a CAES would actually generate an exponential number of arguments.

EXAMPLE 4.8 (Exponential explosion) Consider a class of CAES with $2n$ arguments, such that $\mathit{arguments} = \{a_{11}, a_{12}, a_{21}, a_{22}, \ldots, a_{n1}, a_{n2}\}$. Here the ith pair of arguments has the same conclusion c_i, with premises that depend on the previous, $i-1$th, pair of arguments. So we are building a large chain with pairs of arguments dependent on the previous pair. The start of the chain is not dependent on a previous conclusion, so $a_{11} = \langle\{p_{11}\}, \emptyset, c_1\rangle$, $a_{12} = \langle\{p_{11}\}, \emptyset, c_1\rangle$. The next arguments, for $i > 1$: $a_{i1} = \langle\{p_{i1}, c_{i-1}\}, \emptyset, c_i\rangle$, $a_{i2} = \langle\{p_{i2}, c_{i-1}\}, \emptyset, c_i\rangle$. And finally, we have that $\mathit{assumptions} = \{p_{11}, p_{12}, \ldots, p_{n1}, p_{n2}\}$, $\mathit{weight}(a_i) = 0.5$ and $\text{standard}(a_i) = \text{scintilla}$.

Now consider the corresponding defeasible rules (leaving out rule names):

$$
\begin{aligned}
R_d = \{ & p_{11} \Rightarrow \mathit{arg}_{a_{11}}; \mathit{arg}_{a_{11}} \Rightarrow c_1; \\
& p_{12} \Rightarrow \mathit{arg}_{a_{12}}; \mathit{arg}_{a_{12}} \Rightarrow c_1; \\
& \vdots \\
& p_{n1}, c_{n-1} \Rightarrow \mathit{arg}_{a_{n1}}; \mathit{arg}_{a_{n1}} \Rightarrow c_n; \\
& p_{n2}, c_{n-1} \Rightarrow \mathit{arg}_{a_{n2}}; \mathit{arg}_{a_{n2}} \Rightarrow c_n\}.
\end{aligned}
$$

Although the ith argument only needs the conclusion of one of the $i-1$th arguments to be acceptable, when generating arguments, we will generate every possible combination of sub-arguments, thereby generating 2^n arguments.

This exponential explosion is caused due to the implicit linking of arguments in Carneades that is made explicit when constructing arguments from the corresponding argumentation system.

However, due to the acyclicity of the arguments, this explicit linking is not needed to compute the acceptable conclusions. The exponential explosion can be solved by the following (sketched) polynomial algorithm.

DEFINITION 4.9 (CAES argument generation)

(1) $generatedArgs = \emptyset$.
(2) *sortedArgs* = Topological sort of *arguments* on its dependency graph.
(3) *while* $sortedArgs \neq \emptyset$:
    (a) pick the first argument in *sortedArgs*. Remove all arguments from *sortedArgs* that have the same conclusion, c, and put them in *argSet*;
    (b) translate *argSet* and generate arguments, building on previously *generatedArgs* as sub-arguments, and put the generated arguments in *tempArgs*;
    (c) if present, pick one acceptable argument in *tempArgs* that has the conclusion c and add it to *generatedArgs*;
    (d) $argSet = tempArgs = \emptyset$.

We leave it to future work to formally verify the complexity results of the translation.

This exponential explosion also sheds some light on the complexity of evaluation in Carneades. An important concept in the definition of Carneades is the "concept of a proof", where evaluating a proof (a CAES) should be possible in a tractable time. This tractability can now be proved by verifying the polynomiality of the translation.

4.2.3. *Correspondence results*

We can now prove the main theorem of this article, namely that every argumentation framework that corresponds to a CAES preserves the properties we would expect.

THEOREM 4.10 *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.*

From Part 2 of Theorem 4.10, we can directly relate the theory of a CAES with the corresponding argumentation framework:

COROLLARY 4.11 *A proposition p is part of the theory of C iff p is contained in the closure under propositional logic of the complete extension of AF.*

We have now formally shown that the argumentation framework corresponding to a CAES keeps the properties we wanted to preserve. This proves, in contrast to what was claimed by Gordon et al. (2007), that Carneades can be faithfully modelled in Dung's argumentation frameworks.

It can even be argued that it is too faithful a correspondence, keeping inconsistencies of the original model in the final translation. The inconsistencies caused by the use of the scintilla of evidence proof standard might suggest a change in the definition of proof standard in Carneades,

for instance, by disallowing proof standards that make contradictory conclusions true at the same time.

Regardless, the intermediate translation of a CAES into an argumentation system allows for an easy reparation. We can import the original negation relation of the CAES, generating attacks for any conflicting conclusion. To be precise, given an argumentation system AS corresponding to a CAES, we can make the new contrariness relation: $^-_{AS'} = {}^-_{AS} \cup {}^-_{CAES}$. This will generate additional defeats in the final argumentation framework, thereby breaking the correspondence results; however, it will ensure that no inconsistencies can be generated in the final theory.

4.2.4. *Ambiguity-blocking and ambiguity-propagating*

The stage-specific part of Carneades can be called "ambiguity-blocking" in contrast to "ambiguity-propagating" (see Gordon, Prakken, and Walton 2007, Section 7.1). Here a non-monotonic logic is ambiguity-blocking if, when a conflict between two lines of reasoning with contradictory conclusions cannot be resolved, both lines of reasoning are cut-off and neither of the conclusions can be used for further reasoning. In such logics, it may happen that other lines of reasoning remain undefeated even though one of the cut-off lines of reasoning interferes with it and is not weaker.

Consider the following example, containing an ambiguity between q and $\neg q$ that does not interfere with the inference of $\neg s$ even though $\neg q$ is used as an argument for s.

EXAMPLE 4.12 Consider the CAES $C = \langle arguments, audience, standard\rangle$ and $audience = \langle assumptions, weight\rangle$ with (Figure 6):

$$
\begin{aligned}
arguments &= \{a_1, a_2, a_3, a_4\},\\
a_1 &= \langle\{p\}, \emptyset, q\rangle, a_2 = \langle\{r\}, \emptyset, \neg q\rangle, \rangle,\\
a_3 &= \langle\{\neg q\}, \emptyset, s\rangle, a_4 = \langle\{t\}, \emptyset, \neg s\rangle,\\
assumptions &= \{p, r, t\},\\
weight(a_1) &= weight(a_2) = weight(a_3) = weight(a_4) = 0.5,\\
\text{standard}(q) &= \text{standard}(\neg q) = \text{standard}(s) = \text{standard}(\neg s) = \text{preponderance}.
\end{aligned}
$$

With the proof standard of q, $\neg q$, s and $\neg s$ being *preponderance*, we can see that q, $\neg q$ and s will not be acceptable, but $\neg s$ will be acceptable. Now consider a naive, direct translation of the arguments into defeasible inference rules in ASPIC$^+$, that is, $\mathcal{K}_n = \{p, r, t\}$ and $\mathcal{R}_d = \{p \Rightarrow q, r \Rightarrow \neg q, \neg q \Rightarrow s, t \Rightarrow \neg s\}$. This translation would instead make no corresponding arguments acceptable.

The translation according to Definition 4.4 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 the corresponding AF.

The main difficulty in finding the translation of Carneades to ASPIC$^+$ was dealing with the ambiguity-blocking nature of Carneades, while ASPIC$^+$ is ambiguity-propagating. We have largely solved this problem by introducing additional argument nodes, allowing for an explicit representation of applicability and acceptability. We note that to our knowledge, we are the first to have achieved a translation of an ambiguity-blocking non-monotonic logic to a standard Dung semantics.

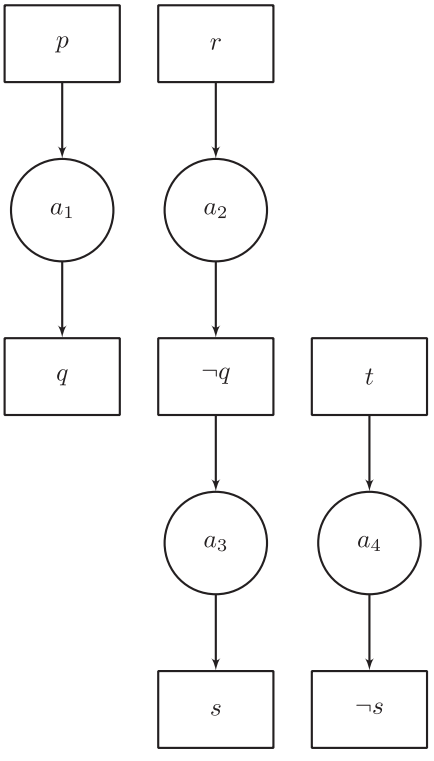

Figure 6. Ambiguity-blocking in Carneades.

4.2.5. *Rationality postulates*

The abstract argumentation model by Dung provides a basis for argumentation systems. The structure imposed, a set of arguments and a defeat (attack) relation, gives rise to multiple semantics to evaluate argumentation frameworks. These semantics, extensions of acceptable arguments, can be seen as rationality constraints on how to evaluate argumentation frameworks. Therefore, when additional structure is imposed on arguments, as done in Section 2.2, more rationality constraints can be imposed on the evaluation. This approach was taken by Caminada and Amgoud (2007), who introduced a set of *rationality postulates* for argumentation systems with more structure. Caminada and Amgoud argued that extensions of these systems should satisfy postulates regarding consistency and closure. We repeat here four postulates[11] in the formulation of Prakken (2010):

- *Closure under sub-arguments:* for every argument in an extension all its sub-arguments are in the extension.
- *Closure under strict rules:* the set of conclusions of all arguments in an extension is closed under strict-rule application.
- *Direct consistency:* the set of conclusions of all arguments in an extension is consistent.
- *Indirect consistency:* the closure of the set of conclusions of all arguments in an extension under strict-rule application is consistent.

We have shown in Section 4.1 that Carneades can be reduced to an ASPIC$^+$ argumentation theory. One advantage of going through this intermediate step is the possibility of applying existing results of ASPIC$^+$ regarding rationality postulates to translated Carneades. To formally verify the rationality postulates, we first need to repeat some properties of argumentation theories.

DEFINITION 4.13 (Properties of an argumentation theory (Modgil and Prakken 2011)) Given an argumentation theory AT with language $\mathcal{L}$, then

- AT is *closed under contraposition* iff for all $S \subseteq \mathcal{L}$, $s \in S$ and $\varphi \in \mathcal{L}$, if $S \vdash \varphi$ then $S \backslash \{s\} \cup \{-\varphi\} \vdash \neg s$;
- AT is *closed under transposition* iff for all $S \subseteq \mathcal{L}$, $s \in S$ and $\varphi_1 \cdots \varphi_n, \psi \in \mathcal{L}$, if $\varphi_1, \ldots, \varphi_n \rightarrow \psi \in \mathcal{R}_s$, then for $i = 1, \ldots, n$, $\varphi_1, \varphi_{i-1}, -\psi, \varphi_{i+1}, \ldots, \varphi_n \rightarrow \neg\varphi_i \in \mathcal{R}_s$;
- AT is *axiom-consistent* iff $Cl_{\mathcal{R}_s}(\mathcal{K}_n)$ is consistent (where $Cl_{\mathcal{R}_s}(P)$ is the smallest set containing P and the consequent of any strict rule in $\mathcal{R}_s$, whose antecedents are in $Cl_{\mathcal{R}_s}(P)$);
- AT is *well formed* iff if φ is a contrary of ψ then [$\psi \notin \mathcal{K}_n$ and ψ is not the consequent of a strict rule].

PROPOSITION 4.14 *Given a CAES C with corresponding argumentation theory AT and corresponding argumentation framework AF, then the four rationality postulates hold for the unique extension of AF.*

An important thing to note here is that although we have proven consistency for the extension of a corresponding argumentation framework, this consistency is relative to the contrariness relation of the argumentation theory. The achievement of gaining consistency in an argumentation framework corresponding to a CAES is mainly due to leaving out the negation relation of that CAES, in a sense circumventing the problem.

4.3. *Generalisation of the translation*

Important future work mentioned by Gordon and Walton (2009b) is to generalise Carneades to cycle-containing structures. Although it was claimed by Brewka and Gordon (2010a) that Carneades would need a cyclic representation in other frameworks, such as Dung's argumentation frameworks, our translation of Carneades translates to cycle-free, or well-founded argumentation frameworks. This same well-foundedness allows for an easy extension of Carneades's argument set to a possibly cycle-containing structure.

Since our translation of a CAES to an argumentation framework does not depend on possible cycles in the set of arguments, we can use the same translation for cycle-containing CAES and deal with the resulting cycles by using the standard Dung semantics.

DEFINITION 4.15 Given a CAES $C = \langle arguments, audience, standard \rangle$ without the acyclicity restriction, $\mathcal{L}_{\text{CAES}}$ 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 $c \in \mathcal{L}_{\text{CAES}}$ 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.

We will demonstrate our generalisation of Carneades by translating Example 2 of Brewka and Gordon (2010a) to an argumentation framework, showing intermediate steps.

EXAMPLE 4.16 (Cycle example) Assume, we have two possible destinations in mind for a summer vacation, Greece and Italy, but cannot afford to visit both destinations. We could formalise

$$\dfrac{\dfrac{\dfrac{}{arg_a}\,app_a}{Gr}\,acc_a}{\neg app_b} \qquad \dfrac{\dfrac{\dfrac{}{arg_b}\,app_b}{It}\,acc_b}{\neg app_a}$$

Figure 7. Greece versus Italy argument trees.

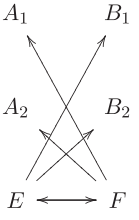

Figure 8. Greece versus Italy argumentation framework.

this as follows. Let *arguments* $= \{a, b\}$ with:

$$a = \langle \emptyset, \{It\}, Gr \rangle, \quad b = \langle \emptyset, \{Gr\}, It \rangle.$$

These arguments contain an exception cycle and therefore cannot directly be handled by Carneades. We can give the example semantics by using the generalisation of our translation. The translation of this CAES would give the following argument trees as shown in Figure 7, which can be written formally:

$$A_1 :\Rightarrow_{app_a} arg_a, \qquad B_1 :\Rightarrow_{app_b} arg_b,$$
$$A_2 : A_1 \Rightarrow_{acc_a} Gr, \quad B_2 : B_1 \Rightarrow_{acc_b} It,$$
$$E : A_2 \Rightarrow \neg app_b, \quad F : B_2 \Rightarrow \neg app_a.$$

From this formal description of arguments, together with the undercuts on applicability and acceptability, we would get the following argumentation framework as shown in Figure 8.

The argumentation framework above can be evaluated through Dung's semantics. For instance, under credulous stable and preferred semantics, both *Gr* and *It* are acceptable. Under sceptical stable, sceptical preferred or grounded semantics both would not be acceptable. These results are similar to the results in the generalisation by Brewka and Gordon (2010a).

4.4. *Related work*

Concurrent to the work done in this article and the paper by van Gijzel and Prakken (2011), there have been translations of Carneades to other argumentation approaches. First of all, there is the translation of Carneades to abstract dialectical frameworks by Brewka and Gordon (2010a). In this translation, premises and exceptions, respectively, have a support and attack relation with the argument node, much in the same way that sub-arguments and undercuts are used in our translation. Carneades' proof standards are encoded as acceptance conditions from the argument node, supporting the conclusion and attacking the contradictory conclusion.

Although the translation of Brewka and Gordon clarified the relation between Carneades and abstract argumentation by relating it to ADFs, one of the main concerns about this translation was that it needed the full power of abstract dialectical frameworks, thus obscuring the direct relation with Dung's argumentation frameworks. This connection has now been made explicit by the paper of Brewka et al. (2011), developing a translation of ADFs to AFs, using Boolean networks (Dunne 1988). The paper concerns itself mostly with the computational complexity of

the translation, to keep a polynomial complexity in both size and time. However, this translation introduces additional technical nodes in the final argumentation framework that have no intuitive meaning. So even though the translation gives a formal connection between the two argumentation models, the intuitive relation is mostly lost.

Very recently, Carneades has been translated to defeasible logic (Nute 1994) by Governatori (2011). Defeasible logic is a computational approach to non-monotonic reasoning with an argumentation-like flavour. Defeasible logic has the possibility to handle both ambiguity-blocking and ambiguity-propagating behaviour, allowing for a rather direct representation of Carneades' proof standards. The translation by Governatori maps proof standards to a single inference mechanism, giving a natural representation of the proof standards.

While Governatori thus establishes an intuitive relation between Carneades and defeasible logic, he only partly relates Carneades to abstract argumentation, since only the ambiguity-propagating part of defeasible logic has an established direct formal relation with Dung's argumentation frameworks. Its ambiguity-blocking variant has instead been translated to a Dung-like semantics using a different notion of acceptability (Governatori, Maher, Antoniou, and Billington 2000).

5. Conclusion

This article has shown that Carneades can be reconstructed, through ASPIC$^+$, as Dung's abstract argumentation frameworks. We have seen that the idea of varying proof standards can be modelled within a Dungean approach, while retaining a correspondence of properties between both systems. These results show that Dung's approach to argumentation is able to model complex argumentation issues such as proof standards. Furthermore, by first translating Carneades through an ASPIC$^+$ argumentation theory, we were able to prove and instantly gain a number of useful results. First of all, we were able to use results about rationality postulates from Prakken (2010) and directly apply these to the translated version of Carneades, proving consistency and strict closure of extensions. The translation also allows us to fully exploit the power of an ASPIC$^+$ argumentation theory, providing us, for instance, with an explicit distinction between strict and defeasible inference rules. So in addition to providing a correspondence, the translation allows us to integrate Carneades with an extra set of tools provided by ASPIC$^+$.

An important property of our reconstruction of Carneades is that our modelling gives a cycle-free argumentation framework, thus always inducing a unique Dung extension which is the same in all Dung's semantics. This shows that Carneades is essentially a single-status assignment approach. This property allowed us to generalise Carneades to cycle-containing structures by using Dung's standard grounded, preferred and stable semantics, thereby addressing the issue put forward by Gordon and Walton (2009b). This generalisation is done much in the same way as by Brewka and Gordon (2010a).

Finally, we note that our translation enables a standard Dung semantics for an "ambiguity-blocking" non-monotonic logic (see Gordon et al. 2007, Section 7.1); to our knowledge, we are the first to have achieved such a result.

5.1. *Future work*

Through this article and by the work of Prakken (2010), several approaches to structured argumentation have been developed and subsequently related through a single framework called ASPIC$^+$. Although the theoretical relations between these approaches have thus been clarified, we have seen that the actual step of generating arguments has not been given a concrete, efficient implementation. A useful path to take for future research would therefore be to develop a class of efficient

argument generation algorithms for ASPIC$^+$. The possibility to efficiently generate arguments for a large class of argumentation approaches would give a better integration of these approaches.

The translation of Carneades to ASPIC$^+$ gives us access to the full power of an ASPIC$^+$ argumentation theory, thereby gaining the possibility to use strict and defeasible rules or use different types of knowledge. This additional power could also be used to show how the concept of argument generators Gordon (2010) and the existing argument schemes of Carneades Gordon and Walton (2009a) relate to structured argumentation by translating them into schemes for defeasible inference rules (following the suggestion of Prakken 2010 that most argument schemes can be seen as such). This would make the relation between ASPIC$^+$ and the Carneades argumentation system as a whole more complete.

Our results raise the question whether it would now be better to use ASPIC+ directly, instead of Carneades, to model argumentation when variable proof standards and the other features of Carneades are required. The answer depends on whether Carneades is sufficient as a model of reasoning with variable proof standards. Prakken and Sartor (2011) claim that Carneades' ambiguity-blocking nature prevents an adequate modelling of the distinction between the burdens of production and persuasion. If they are right, then there is reason to change Carneades in the direction of ASPIC+.

Notes

1. Part of this work has appeared before in van Gijzel and Prakken (2011) and van Gijzel (2011).
2. Dung calls it "attack" but to unify terminology we rename it to "defeat".
3. Prakken (2010) calls the general argumentation model, defined by Dung (1995), argumentation frameworks. This is in contrast to the use of Dung, where an argumentation framework is a specific set of arguments and defeat relation. Instantiations of the abstract argumentation model by Dung are in ASPIC$^+$ called *argumentation systems* and argumentation frameworks are called *argumentation theories*.
4. This slightly changes the definition of Prakken (2010), disallowing arguments with issue premises to be acceptable at all, instead of only excluding them from extensions.
5. In contrast to how it was presented by Gordon and Walton (2009b), the *weight* function will not be partial. Evaluation of arguments without weights was, in exception of the rare case that the standard was *scintilla* for both c and $\bar{c}$, undefined, and therefore the partiality of the weight function will be assumed to be incorrect.
6. A possible generalisation to cyclic structures was left as future work (Gordon and Walton 2009b). This restriction was lifted in the translation done to abstract dialectical frameworks by Brewka and Gordon (2010a) and we will see that similar results hold for our translation done in Section 4.3.
7. This slightly generalises Carneades in that we allow complementary literals to be given different proof standards, in contrast to Gordon et al. (2007) and the implementations of Carneades.
8. This (stage-specific) definition fixes a small technical error in the original definition by Gordon and Walton (2009b). The function *standard* returning a proof standard was instead directly called with the arguments for a proof standard.
9. The idea to make the applicability step explicit by means of an argument node is adapted from Brewka and Gordon (2010a).
10. This is probably a technical mistake of Gordon and Walton (2009b). It can be fixed by slightly changing the definition of applicability of arguments, including the additional demand "there is not an assumption, $\bar{c}$ in the audience that is contradictory to the conclusion, c, of the argument".
11. The results of the other postulates follow directly from these four.

References

Amgoud, L. (2005), 'A Unified Setting for Inference and Decision: An Argumentation-Based Approach', in *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, AUAI Press, pp. 26–33.

Amgoud, L., Bodenstaff, L., Caminada, M., McBurney, P., Parsons, S., Prakken, H., van Veenen, J., and Vreeswijk, G. (2006), 'Final Review and Report on Formal Argumentation System', Deliverable D2.6, ASPIC IST-FP6-002307, ASPIC Consortium.

Atkinson, K., and Bench-Capon, T. (2007), 'Argumentation and Standards of Proof', in *Proceedings of the 11th International Conference on Artificial Intelligence and Law (ICAIL-07)*, Stanford, CA, New York: ACM, pp. 107–116.

Baroni, P., and Giacomin, M. (2007), 'On Principle-Based Evaluation of Extension-Based Argumentation Semantics', *Artificial Intelligence*, 171, 675–700.

Baroni, P., and Giacomin, M. (2009), 'Semantics of Abstract Argument Systems', in *Argumentation in Artificial Intelligence*, eds. G. Simari and I. Rahwan, Springer, New York, pp. 25–44.

Bench-Capon, T.J.M., Doutre, S., and Dunne, P.E. (2007), 'Audiences in Argumentation Frameworks', *Artificial Intelligence*, 171, 42–71.

Besnard, P., and Hunter, A. (2008), *Elements of Argumentation*, MIT Press, Cambridge, MA.

Bondarenko, A., Dung, P.M., Kowalski, R.A., and Toni, F. (1997), 'An Abstract, Argumentation-Theoretic Framework for Default Reasoning', *Artificial Intelligence*, 93, 63–101.

Brewka, G., and Gordon, T.F. (2010a), 'Carneades and Abstract Dialectical Frameworks: A Reconstruction', in *Computational Models of Argument. Proceedings of COMMA 2010*, eds. M. Giacomin and G.R. Simari, Amsterdam: IOS Press, pp. 3–12.

Brewka, G., and Woltran, S. (2010b), 'Abstract Dialectical Frameworks', in *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning*, AAAI Press, pp. 102–111.

Brewka, G., Dunne, P.E., and Woltran, S. (2011), 'Relating the Semantics of Abstract Dialectical Frameworks and Standard AFs', in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 780–785.

Caminada, M., and Amgoud, L. (2007), 'On the Evaluation of Argumentation Formalisms', *Artificial Intelligence*, 171, 286–310.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., and Stein, C., *Introduction to Algorithms* (2nd ed.), MIT Press and McGraw-Hill.

Dung, P.M. (1995), 'On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games', *Artificial Intelligence*, 77, 321–357.

Dung, P.M., Kowalski, R.A., and Toni, F. (2009), 'Assumption-Based Argumentation', in *Argumentation in Artificial Intelligence*, eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 199–218.

Dung, P.M., Mancarella, P., and Toni, F. (2007), 'Computing Ideal Sceptical Argumentation', *Artificial Intelligence*, 171, 642–674.

Dunne, P.E. (1988), *The Complexity of Boolean Networks*, San Diego, CA: Academic Press Professional, Inc.

Farley, A.M., and Freeman, K. (1995), 'Burden of Proof in Legal Argumentation', in *Proceedings of the 5th International Conference on Artificial Intelligence and Law (ICAIL-05)*, College Park, MD, New York: ACM, pp. 156–164.

Freeman, K., and Farley, A.M. (1996), 'A Model of Argumentation and Its Application to Legal Reasoning', *Artificial Intelligence and Law*, 4, 163–197, doi:10.1007/BF00118492.

van Gijzel, B. (2011), 'Relating Proof Standards and Abstract Argumentation', Master's thesis, Utrecht University.

van Gijzel, B., and Prakken, H. (2011), 'Relating Carneades with Abstract Argumentation', in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1113–1119.

Gordon, T.F. (2010), 'An Overview of the Carneades Argumentation Support System', in *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning*, eds. C. Tindale and C. Reed, College Publications, London, pp. 145–156.

Gordon, T.F., and Walton, D. (2009a), 'Legal Reasoning with Argumentation Schemes', in *Proceedings of the 12th International Conference on Artificial Intelligence and Law (ICAIL-09)*, Barcelona, New York: ACM, pp. 137–146.

Gordon, T.F., and Walton, D. (2009b), 'Proof Burdens and Standards', in *Argumentation in Artificial Intelligence*, eds. G. Simari and I. Rahwan, Springer, New York, pp. 239–258.

Gordon, T.F., Prakken, H., and Walton, D. (2007), 'The Carneades Model of Argument and Burden of Proof', *Artificial Intelligence*, 171, 875–896.

Governatori, G. (2011), 'On the Relationship Between Carneades and Defeasible Logic', in *Proceedings of the 13th International Conference on Artificial Intelligence and Law (ICAIL 2011)*, ed. T. van Engers, ACM Press, Pittsburgh, PA.

Governatori, G., Maher, M., Antoniou, G., and Billington, D. (2000), 'Argumentation Semantics for Defeasible Logics', in *Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence*, PRICAI'00, Melbourne, Berlin, Heidelberg: Springer-Verlag, pp. 27–37.

Modgil, S., and Prakken, H. (2011), 'Revisiting Preferences and Argumentation', in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1021–1026.

Nute, D. (1994), 'Defeasible Logic', in *Handbook of Logic in Artificial Intelligence and Logic Programming* (Vol. 3), eds. Dov M. Gabbay, C.J. Hogger and J.A. Robinson, Oxford University Press, New York, pp. 353–395.

Prakken, H. (2010), 'An Abstract Framework for Argumentation with Structured Arguments', *Argument & Computation*, 1, 93–124.

Prakken, H., and Sartor, G. (2011), 'On Modelling Burdens and Standards of Proof in Structured Argumentation', in *Legal Knowledge and Information Systems. JURIX 2011: The 24th Annual Conference*, ed. K. Atkinson, IOS Press, Amsterdam, pp. 83–92.

Reiter, R. (1980), 'A Logic for Default Reasoning', *Artificial Intelligence*, 13, 81–132.

Vreeswijk, G. (1993), 'Studies in Defeasible Argumentation', Doctoral dissertation, Free University Amsterdam.

Vreeswijk, G. (1997), 'Abstract Argumentation Systems', *Artificial Intelligence*, 90, 225–279.

Appendix: proofs

PROPOSITION 4.6 *Every argumentation framework corresponding to a (finite) CAES according to Definition* 4.4 *is well-founded.*

Proof *Given a CAES* $C = \langle arguments, audience, standard\rangle$ *with* $audience = \langle assumptions, weight\rangle$ *and corresponding argumentation framework AF. Assume AF is not well-founded and given that C is finite, there exists a sequence of arguments* $A_1, \ldots, A_n$ *in AF, such that* $\text{defeats}(A_n, A_1)$ *and for each* $i < n$, $\text{defeats}(A_i, A_{i+1})$ *hold.*

Given that for our translation every Carneades argument a_i *is assigned a unique argument node* arg_i*, we define ASPIC*$^+$ *arguments of the form* $A_i : C_1, \ldots C_l \Rightarrow_{app_i} arg_i$*, and the possible extension* $A_{i'} : A_i \Rightarrow c$*, to correspond with* a_i*. Note that by construction every* $Conc(C_i)$ *will be a premise of* a_i *and the set of contraries of* arg_i *will contain the exceptions of* a_i*. Then, denoting the direct and indirect parents of a node, c in a dependency graph as* $ancestors(c)$*, we will show that for every pair of arguments* A_i, A_j *for which* $defeats(A_i, A_j)$ *holds,* $conc(a_i) \in ancestors(conc(a_j))$ *holds for the corresponding Carneades arguments* a_i *and* a_j*. By proving this statement, we can infer that for every pair of arguments* $ancestors(conc(a_i)) \subseteq ancestors(conc(a_j))$ *holds, entailing* $ancestors(conc(a_i)) = ancestors(conc(a_j))$ *by cyclity of the defeat sequence. Then by combining both statements we can infer that* $conc(a_i) \in ancestors(conc(a_i))$*, inducing a cycle in the dependency graph, contradicting our initial acylicity assumption of arguments, thereby proving what we want. (Notice that we are talking about defeat cycles in ASPIC*$^+$ *and dependency cycles in a CAES, which although related, are not of the same nature.)*

We will prove the above property by considering the shapes of an arbitrary defeating argument A_i *and its target* A_j *in our defeat sequence. By construction of our translation we can see that a defeating argument can only be of the following shapes*:

Case 1: $A_i :\rightarrow \neg acc_j$*. This argument cannot be defeated, and will therefore never be part of a defeat sequence, contradicting our assumption.*

Case 2: $A_i : C_1, \ldots, C_l \Rightarrow_{app_i} arg_i$ *(with possible superargument:* $A'_i \Rightarrow_{acc_i} c$*).* A_i *can undercut an argument of the following form:* $A_{subj} : D_1, \ldots, D_m \Rightarrow_{app_{\text{subj}}} arg_{\text{subj}} \Rightarrow_{acc_{\text{subj}}} \bar{c}$*, assuming that* $arg_i \in \overline{acc_{subj}}$*. However, a conclusion* $\bar{c}$ *cannot defeat the next argument* A_k *in the defeat chain and therefore* A_i *must have defeated a proper sub-argument of* A_j*. So* A_j *either extended* A_{subj} *to an argument that is an exception to the next argument* ($\neg app_j$) *or it was extended to an argument that has an argument node* (arg_j) *as the*

conclusion (note that A_{subj} can be one of the many sub-arguments of A_j). In both cases, we see that the argument a_j corresponding to A_j must have had $\bar{c}$ as a premise, inducing that a_j has c and $\bar{c}$ as dependencies. This establishes $\text{conc}(a_i) \in ancestors(conc(a_j))$.

Case 3: *$A_i : e_i \Rightarrow \neg app_j$ with $e_i \in$ assumptions. This argument cannot be defeated, and will therefore never be part of a defeat sequence, contradicting our assumption.*

Case 4: *$A_i : C_1, \ldots C_l \Rightarrow_{app_i} arg_i \Rightarrow_{acc_i} e_i \Rightarrow \neg app_j$. A_i can undercut A_j, on its sub-argument $D_1, \ldots, D_m \Rightarrow_{app_j} arg_j$. This does not have to be a proper sub-argument of A_j. For $\neg app_j$ to be a contrary of app_j, e_i has to be an exception of the argument corresponding to A_j, a_j. This immediately establishes that $conc(a_i) \in ancestors(conc(a_j))$.* ■

THEOREM 4.10 *Let C be a CAES, $\langle arguments, audience, standard\rangle$, $\mathcal{L}_{\text{CAES}}$ 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 $c \in \mathcal{L}_{\text{CAES}}$ 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 sub-arguments (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 $\bar{c}$. Then, again by the directionality criterion, we have (2. $\Rightarrow$) left to prove for c and $\bar{c}$ in C.

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

If a is applicable and $d = c$, 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 $\bar{c}$ unsatisfiable in C, by construction of AF, A_1 will defeat any argument b with conclusion $\bar{c}$ on its inference rule arg_b. So by conflict-freeness no such argument will be in CE_{AF}, correctly preserving acceptability of $\bar{c}$.

If a is applicable and $d = \bar{c}$, then we only need to consider the influence of the applicability of a on the acceptability of c, since the acceptability of $\bar{c}$ is irrelevant. First, by (1.) there exists an argument A_1 with conclusion arg_a in CE_{AF}. Consider standard(c) = clear-and-convincing. Take an applicable argument b for c in C' with the highest weight. Since c is acceptable, b satisfies all (weight) conditions for clear-and-convincing. By the induction hypothesis there is an argument B_i, with $Conc(B_i) = c$, in $CE_{AF'}$. Then notice A_1 defeats B_i iff arg_a is a contrary of acc_b which holds iff $weight(b) \not> weight(a) + \beta$, correctly preserving acceptability of 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 a is not applicable or d is neither c nor $\bar{c}$.*

*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) = clear-*and-convincing. *Then either $weight(a) \leq \alpha$ or $weight(a) \leq weight(b) + \beta$ for another applicable argument b with conclusion $\bar{c}$. 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}.*

If $d = \bar{c}$ and c is not acceptable, then applicability of a will not influence acceptability of c. ■

PROPOSITION 4.14 *Given a CAES C with corresponding argumentation theory AT and corresponding argumentation framework AF, then the following rationality postulates hold for the unique extension of AF:*

(1) *Closure under sub-arguments*,
(2) *Closure under strict rules*,
(3) *Direct consistency*,
(4) *Indirect consistency*.

Proof The first two rationality postulates follow directly from the translation of C to an ASPIC$^+$ argumentation theory and Propositions 6.1 and 6.2 of Prakken (2010). For the other two postulates, by Theorem 6.9 and 6.10 of Prakken (2010), we will have to prove our argumentation theory to be closed under contraposition or transposition, axiom-consistent, well-formed and finally have a reasonable argument ordering (Definition 6.7 of Prakken 2010).

The corresponding AT of C has only one type of strict rule: $\rightarrow \neg acc_a$, for every argument a in arguments. Closure under contraposition (and transposition) is immediately satisfied by the lack of strict rules starting with a premise acc_a. Premises from $\mathcal{K}_n$ cannot be used as a premise for a strict rule, trivially satisfying closure under strict rules, given the consistency of $\mathcal{K}_n$. If φ is a contrary of ψ, then by construction of the translation, ψ is always of the form app_a or arg_a. Neither are in $\mathcal{K}_n$, nor a consequent of a strict rule, thus satisfying well-formedness. Finally, preferences are not used in the corresponding AT, allowing us to take any reasonable argument ordering. ■