Argument and Computation
Vol. 1, No. 2, June 2010, 93–124

Taylor & Francis
Taylor & Francis Group

An abstract framework for argumentation with structured arguments

Henry Prakken*

Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands; Faculty of Law, University of Groningen, Groningen, The Netherlands

(*Received 18 September 2009; final version received 9 December 2009*)

An abstract framework for structured arguments is presented, which instantiates Dung's ('On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and *n*-Person Games', *Artificial Intelligence*, 77, 321–357) abstract argumentation frameworks. Arguments are defined as inference trees formed by applying two kinds of inference rules: strict and defeasible rules. This naturally leads to three ways of attacking an argument: attacking a premise, attacking a conclusion and attacking an inference. To resolve such attacks, preferences may be used, which leads to three corresponding kinds of defeat: undermining, rebutting and undercutting defeats. The nature of the inference rules, the structure of the logical language on which they operate and the origin of the preferences are, apart from some basic assumptions, left unspecified. The resulting framework integrates work of Pollock, Vreeswijk and others on the structure of arguments and the nature of defeat and extends it in several respects. Various rationality postulates are proved to be satisfied by the framework, and several existing approaches are proved to be a special case of the framework, including assumption-based argumentation and DefLog.

Keywords: argumentation frameworks; structured arguments; rationality postulates

1. Introduction

In 1995, Phan Minh Dung introduced an abstract formalism for argumentation-based inference (Dung 1995), which assumes as input nothing else but a set (of arguments) ordered by a binary relation (of attack). Although he thus fully abstracted from the structure of arguments and the nature of the attack relation, he was still able to develop an extremely interesting theory. His article was a breakthrough in three ways: it provided a general and intuitive semantics for the consequence notions of argumentation logics (and for non-monotonic logics in general); it made a precise comparison possible between different systems (by translating them into his abstract format) and it made a general study of formal properties of systems possible, which are inherited by instantiations of his framework. In consequence, Dung's work has given an enormous boost to research in computational argumentation. Yet it has also been criticised for not specifying the structure of arguments and the nature of the attack relation, which makes it less suitable for modelling specific argumentation problems. I believe that such criticism fails to appreciate the nature of Dung's formalism. It is best seen not as a formalism for directly representing argumentation-based inference problems but as a tool for analysing particular argumentation systems and for developing a meta-theory of such systems. As such it has been very successful: differences between particular systems can be characterised in terms of some simple notions, and formal results established for the framework are inherited by its instantiations. This was already illustrated by Dung (1995) with reconstructions of

*Email: henry@cs.uu.nl

ISSN 1946-2166 print/ISSN 1946-2174 online
© 2010 Taylor & Francis
DOI: 10.1080/19462160903564592
http://www.informaworld.com

Pollock's (1987) system, various logic-programming semantics and Reiter's (1980) default logic in his formalism.

Nevertheless, it is true that when actual argumentation-based inference has to be modelled, Dung's framework is by itself usually too abstract and instead an instantiated version of his approach should be used. However, here too abstraction is still possible and worthwhile. The aim of this paper is to instantiate Dung's abstract approach with a general account of the structure of arguments and the nature of the defeat relation.[1] The framework defines arguments as inference trees formed by applying two kinds of inference rules, strict and defeasible rules. This naturally leads to three ways of attacking an argument: attacking a premise, a conclusion and an inference. To resolve such attacks, preferences may be used, which leads to three corresponding kinds of defeat: undermining, rebutting and undercutting defeats. To characterise them, some minimal assumptions on the logical object language must be made, namely that certain well-formed formulas are a contrary or contradictory of certain other well-formed formulas. Apart from this, the framework is still abstract: it applies to any set of inference rules, as long as it is divided into strict and defeasible ones, and to any logical language with a contrary relation defined over it.

The choice for tree-structured arguments based on two types of inference rules arguably is very natural both in light of logic and argumentation theory and when looking at argumentation as it occurs in human thinking and dialogue. The notion of arguments as trees of inferences is very common in standard logic and in argumentation theory and is the basis of many software tools for argument visualisation. Moreover, in actual argumentation, humans often express their arguments as claims supported with one or more premises, which can in turn be supported with further premises, and so on. Finally, as will be further explained in Section 4, the setup with general defeasible inference rules is very suited for modelling reasoning with argumentation schemes (Walton, Reed, and Macagno 2008).

The account offered in this paper is not completely new. In fact, a rhetorical aim of the paper is to counter the idea that the computational study of argumentation started with Dung's abstract approach and that only then researchers made it more concrete with accounts of the structure of arguments and the nature of defeat. As a matter of fact, much work on these two issues was already done or going on at the time when Dung wrote his paper, and some of this work is still state-of-the-art. For instance, both Pollock (1987, 1994) and Vreeswijk (1993, 1997) did important work on the structure of arguments, while Pollock (1974, 1987) introduced an important distinction between two kinds of defeat, namely rebutting defeat (attack on a conclusion) and undercutting defeat (attack on an inference rule). One aim of the present paper is to profit from, integrate and build on this and other important work as much as possible. As such, this paper is a further development of the integration attempt that was undertaken in the European ASPIC project (Amgoud et al. 2006). In this project, Vreeswijk's formalisation of the structure of arguments was combined with Pollock's definitions of rebutting and undercutting defeat in a way that also used insights from other work. The result was a characterisation of a set of tree-structured arguments ordered with a binary defeat relation, so that an instantiation of Dung's abstract approach was achieved and any of Dung's semantics could be used to compute the acceptability status of the structured arguments.

The ASPIC framework was developed by Leila Amgoud, Martin Caminada, Claudette Cayrol, Marie-Christine Lagasquie-Schieux, myself and Gerard Vreeswijk and was first reported in a European project deliverable (Amgoud et al. 2006). The added expressiveness compared with Dung's abstract formalism gave rise to further work by Caminada and Amgoud (2007) on rationality postulates for systems instantiating the ASPIC framework. The aim of this work was to propose the idea of rationality postulates and to criticise some

specific rule-based argumentation systems for failing to satisfy them. For this aim, only a simplified version of the ASPIC framework was needed, without preferences and without the notion of a knowledge base. Moreover, the examples discussed by Caminada and Amgoud (2007) were all with domain-specific inference rules instead of with general inference patterns, which in effect somewhat obscured the potential of the framework to be a general account of structured argumentation.

In contrast, the present paper aims to present the ASPIC framework as a general abstract model of argumentation with structured arguments.[2] To achieve this aim, the ASPIC framework will be extended and generalised in four respects.

(1) A third way of argument attack, namely premise attack or 'undermining', will be added, in a way inspired by Vreeswijk's (1993, chap. 8) combination of 'plausible' and 'defeasible' argumentation. Apart from the naturalness of having all three kinds of attack in a general framework for argumentation, this will make it easier to formalise argument schemes in the framework and it will make it possible to regard existing systems with premise attack as special cases of the framework.
(2) The three notions of attack will be generalised from the notion of contradiction between formulas φ and $\neg\varphi$ to an abstract relation of contrariness between formulas which is not necessarily symmetric. This idea is taken from Bondarenko, Dung, Kowalski, and Toni (1997) and Verheij (2003a) and will help in showing that their systems are a special case of the present framework.
(3) Four types of premises will be distinguished, inspired by a similar distinction of Gordon, Prakken, and Walton (2007).
(4) Attack relations will be partly resolved with preference orderings on arguments, defeasible rules and the knowledge base (although Amgoud et al. (2006) also have preferences, the results of Caminada and Amgoud (2007) do not cover them).

It will then be investigated to what extent the results of Caminada and Amgoud (2007) on rationality postulates generalise to the thus extended ASPIC framework. The final aim of this paper is to compare the resulting framework with recent related work. It will turn out that assumption-based argumentation (Bondarenko et al. 1997; Dung, Kowalski, and Toni 2006; Dung, Mancarella, and Toni 2007), DefLog (Verheij 2003) and Amgoud and Cayrol (2002)'s version of deductive argumentation are special cases of this paper's version of the ASPIC framework.

2. Dung's abstract argumentation frameworks

First without explanation, the basic concepts and insights of Dung's abstract argumentation approach are listed. For a state-of-the-art introduction, see Baroni and Giacomin (2009).

DEFINITION 2.1 (*abstract argumentation framework*) An *abstract argumentation framework* (*AF*) is a pair $\langle \mathcal{A}, Def \rangle$. $\mathcal{A}$ is a set arguments and $Def \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of defeat. We say that an argument A *defeats* an argument B iff $(A,B) \in Def$.

DEFINITION 2.2 (conflict-free, defence) Let $\mathcal{B} \subseteq \mathcal{A}$.

- A set $\mathcal{B}$ is *conflict-free* iff there exist no A_i, A_j in $\mathcal{B}$ such that A_i defeats A_j.
- A set $\mathcal{B}$ *defends* an argument A_i iff for each argument $A_j \in \mathcal{A}$, if A_j defeats A_i, then there exists A_k in $\mathcal{B}$ such that A_k defeats A_j.

DEFINITION 2.3 (acceptability semantics) Let $\mathcal{B}$ be a conflict-free set of arguments, and let $\mathcal{F}$: $2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ be a function such that $\mathcal{F}(\mathcal{B}) = \{A \mid \mathcal{B}$ defends $A\}$.

- $\mathcal{B}$ is *admissible* iff $\mathcal{B} \subseteq \mathcal{F}(\mathcal{B})$.
- $\mathcal{B}$ is a *complete extension* iff $\mathcal{B} = \mathcal{F}(\mathcal{B})$.
- $\mathcal{B}$ is a *grounded extension* iff it is the smallest (w.r.t. set inclusion) complete extension.
- $\mathcal{B}$ is a *preferred extension* iff it is a maximal (w.r.t. set inclusion) complete extension (or, equivalently, if $\mathcal{B}$ is a maximal (w.r.t. set inclusion) admissible set).
- $\mathcal{B}$ is a *stable extension* iff it is a preferred extension that defeats all arguments in $\mathcal{A}\backslash\mathcal{B}$.

Note that this implies that each grounded, preferred or stable extension of an *AF* is also a complete extension of that *AF*. Some other known results are that

- the grounded extension is indeed unique but all other semantics allow for multiple extensions of an *AF*;
- each *AF* has a grounded and at least one preferred and complete extension, but there are *AF*s without stable extensions;
- the grounded extension of an *AF* is contained in all other extensions of that *AF*.

3. Argumentation systems with structured arguments

In this section, the arguments of Dung's argumentation frameworks are given structure and its defeat relation is defined in terms of the structure of arguments plus external preference information. Apart from this, the resulting formalism is still as abstract as possible, allowing for different logical languages, different sets of inference rules for building arguments and different preference orderings. The framework uses Vreeswijk's (1993, 1997) definition of the structure of arguments and then adds Pollock's (1987, 1994) distinction between rebutting and undercutting attack, as well as a variant of the notion of premise attack proposed by Vreeswijk (1993, chap. 8). These notions are then generalised to languages with arbitrary relations of contrariness and contradiction between well-formed formulas. Then the three notions of attack are combined into a notion of defeat in a way inspired by Vreeswijk (1993, chap. 8) and Prakken and Sartor (1997). It is this combination that makes it possible to regard the system as an instantiation of Dung's abstract framework.

The resulting framework unifies two ways to capture the defeasibility of reasoning. Some, e.g. Amgoud and Cayrol (2002), Besnard and Hunter (2008), Bondarenko et al. (1997), Verheij (2003a), locate the defeasibility of arguments in the uncertainty of their premises, so that arguments can only be attacked on their premises. Others, e.g. Pollock (1994), Vreeswijk (1997), instead locate the defeasibility of arguments in the riskiness of their inference rules: in these logics, inference rules are of two kinds, being either deductive or defeasible, and arguments can only be attacked on their applications of defeasible inference rules. Typically, in this approach inconsistency of the knowledge base makes the system collapse. Vreeswijk (1993, chap. 8) called these two approaches *plausible* and *defeasible* reasoning: he described plausible reasoning as sound (i.e. deductive) reasoning on an uncertain basis and defeasible reasoning as unsound (but still rational) reasoning on a solid basis. In Chapter 8, Vreeswijk attempted to combine both forms of reasoning in a single formalism, but since then most formal accounts of argumentation have modelled either only plausible or only defeasible reasoning.

3.1. *Basic definitions*

The basic notion of the present framework is that of an argumentation system, which extends the familiar notion of a proof system with a distinction between strict and defeasible inference rules[3] and a preference ordering on the defeasible inference rules.

DEFINITION 3.1 (argumentation system) *An argumentation system* is a tuple $AS = (\mathcal{L}, ^{-}, \mathcal{R}, \leq)$ where

- $\mathcal{L}$ is a logical language,
- $^{-}$ is a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$,
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$,
- $\leq$ is a partial preorder on $\mathcal{R}_d$.

Amgoud et al. (2006) and Caminada and Amgoud (2007) assume that arguments are expressed in a logical language that is left unspecified except that it is closed under classical negation. In this paper, this assumption will be generalised in two ways. First, non-symmetric conflict relations between formulas will be allowed, such as the contrariness relation of Bondarenko et al. (1997) (which captures, for instance, negation as failure), and its inverse, the dialectical negation of Verheij (2003a) (which means 'it is defeated that'). Second, in addition to classical negation, other symmetric conflict relations will be allowed, so that, for instance, formulas like 'bachelor' and 'married' can, if desired, be declared contradictory without having to reason with an axiom $\neg$(bachelor $\wedge$ married).

DEFINITION 3.2 (logical language) Let $\mathcal{L}$, a set, be a logical language and $^{-}$ a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$. If $\varphi \in \overline{\psi}$ then if $\psi \notin \overline{\varphi}$, then φ is called a *contrary* of ψ, otherwise φ and ψ are called *contradictory*. The latter case is denoted by $\varphi = -\psi$ (i.e. $\varphi \in \overline{\psi}$ and $\psi \in \overline{\varphi}$).

In examples with classical negation $\neg$, it will be assumed that $\neg\varphi \in \overline{\varphi}$ and $\varphi \in \overline{\neg\varphi}$.

Now that the notion of negation has been generalised, the same must be done with the notion of consistency.

DEFINITION 3.3 (consistent set) Let $\mathcal{P} \subseteq \mathcal{L}$. $\mathcal{P}$ is *consistent* iff $\nexists\, \psi, \varphi \in \mathcal{P}$ such that $\psi \in \overline{\varphi}$, otherwise it is *inconsistent*.

Note that this is a weak form of consistency, determined by whether a set contains contrary or contradictory formulas. Caminada and Amgoud (2007) call this *direct consistency* and they call consistency of the closure of a set under strict inference *indirect consistency*.

Arguments are built by applying inference rules to subsets of $\mathcal{L}$. Inference rules are either *strict* or *defeasible*. This distinction goes back to Lin and Shoham (1989), Pollock (1987) and Vreeswijk (1993), as does the idea of abstracting from their nature.

DEFINITION 3.4 (strict and defeasible rules) Let $\varphi_1, \ldots, \varphi_n, \varphi$ be elements of $\mathcal{L}$.

- A *strict rule* is of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$, informally meaning that if $\varphi_1, \ldots, \varphi_n$ hold, then *without exception* it holds that φ.
- A *defeasible rule* is of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$, informally meaning that if $\varphi_1, \ldots, \varphi_n$ hold, then it *presumably* holds that φ.

$\varphi_1, \ldots, \varphi_n$ are called the *antecedents* of the rule and φ its *consequent*.

As usual in logic, inference rules will often be specified by schemes in which a rule's antecedents and consequent are metavariables ranging over $\mathcal{L}$.

Arguments are constructed from a knowledge base which, inspired by Gordon et al. (2007), is assumed to contain four kinds of formulas.

DEFINITION 3.5 (knowledge bases) A *knowledge base* in an argumentation system $(\mathcal{L}, ^{-}, \mathcal{R}, \leq)$ is a pair $(\mathcal{K}, \leq')$, 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 and

- $\mathcal{K}_n$ is a set of (necessary) *axioms*. Intuitively, arguments cannot be attacked on their axiom premises.
- $\mathcal{K}_p$ is a set of *ordinary premises*. Intuitively, arguments can be attacked on their ordinary premises, and whether this results in defeat must be determined by comparing the attacker and the attacked premise (in a way specified below).
- $\mathcal{K}_a$ is a set of *assumptions*. Intuitively, arguments can be attacked on their assumptions, where these attacks always succeed.
- $\mathcal{K}_i$ is a set of *issues*. Intuitively, arguments of which the premises include an issue are never acceptable: an issue must always be backed with a further argument.

(Gordon et al. (2007) call ordinary premises 'assumptions', they regard assumptions as the contradictories of 'exceptions' and they call issues 'ordinary premises'. Their counterpart to axioms is 'accepted' and 'rejected' statements.) As explained by Gordon et al. (2007), the category of issue premises is useful if an argumentation system is embedded in a dialogical context, defining the acceptability status of arguments relative to a stage in a dialogue. For example, in legal proceedings, legal claims that are not backed by factual evidence usually do not stand: for instance, an argument 'we have a contract by Section X of the Civil Code since I made an offer and you accepted' will be unacceptable as long as no factual evidence for the offer and acceptance is provided. In the present framework, this can be captured by giving the non-supported premises issue status.

3.2. *Arguments*

Next the arguments that can be constructed from a knowledge base in an argumentation system are defined. Arguments can be constructed step-by-step by chaining inference rules into trees. Arguments thus contain subarguments, which are the structures that support intermediate conclusions (plus the argument itself and its premises as limiting cases). In what follows, for a given argument, the function Prem returns all the formulas of $\mathcal{K}$ (called *premises*) used to build the argument, Conc returns its conclusion, Sub returns all its subarguments, DefRules returns all the defeasible rules of the argument and, finally, TopRule returns the last inference rule used in the argument.

DEFINITION 3.6 (argument) An *argument* A on the basis of a knowledge base $(\mathcal{K}, \leq')$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is

(1) φ if $\varphi \in \mathcal{K}$ with
$\text{Prem}(A) = \{\varphi\}$,
$\text{Conc}(A) = \varphi$,
$\text{Sub}(A) = \{\varphi\}$,
$\text{DefRules}(A) = \emptyset$,
$\text{TopRule}(A) =$ undefined.

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

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

Example 3.7 Consider a knowledge base in an argumentation system with

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

An argument for w is displayed in a traditional proof-tree format in Figure 1, where a single line stands for a strict inference and a double line for a defeasible inference. The type of a premise is indicated with a superscript. Formally, the argument and its subarguments are written as follows:

A_1: p $\quad$ A_5: $A_1 \Rightarrow t$
A_2: q $\quad$ A_6: $A_1, A_2 \rightarrow s$
A_3: r $\quad$ A_7: $A_5, A_3, A_6 \Rightarrow v$
A_4: u $\quad$ A_8: $A_7, A_4 \rightarrow w$

We have that

$\text{Prem}(A_8) = \{p, q, r, u\}$
$\text{Conc}(A_8) = w$
$\text{Sub}(A_8) = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$
$\text{DefRules}(A_8) = \{p \Rightarrow t;\ s, r, t \Rightarrow v\}$
$\text{TopRule}(A_8) = v, u \rightarrow w$

DEFINITION 3.8 (argument properties) An argument A is

- strict if $\text{DefRules}(A) = \emptyset$;
- *defeasible* if $\text{DefRules}(A) \neq \emptyset$;
- *firm* if $\text{Prem}(A) \subseteq \mathcal{K}_n$;
- *plausible* if $\text{Prem}(A) \not\subseteq \mathcal{K}_n$.

We write $S \vdash \varphi$ if there exists a strict argument for φ with all premises taken from S, and $S \mid\sim \varphi$ if there exists a defeasible argument for φ with all premises taken from S.

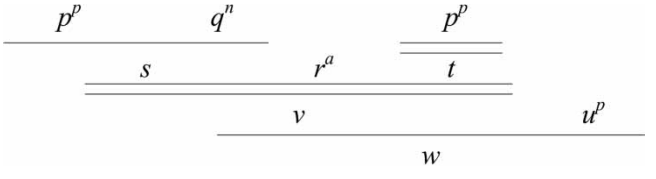

Figure 1. An argument.

Example 3.9 In Example 3.7 the argument A_2 is strict and firm, while A_1, A_3, A_4 and A_6 are strict and plausible and A_5, A_7 and A_8 are defeasible and plausible. Furthermore, we have that $\mathcal{K} \vdash p$, $\mathcal{K} \vdash q$, $\mathcal{K} \vdash r$, $\mathcal{K} \vdash u$, $\mathcal{K} \vdash s$ and $\mathcal{K} \mathrel{|\!\sim} t$, $\mathcal{K} \mathrel{|\!\sim} v$, $\mathcal{K} \mathrel{|\!\sim} w$.

(From hereon, the theory will be left implicit if there is no danger for confusion.)

Now that the notion of an argument has been defined, orderings on arguments can be considered. Below $\preceq$ is a partial preorder such that $A \preceq B$ means that B is at least as 'good' as A. As usual $A \prec B$ means $A \preceq B$ and $B \not\preceq A$.

In Section 6, two ways will be discussed to define $\preceq$ as a function from the orderings $\leq$ on $\mathcal{R}_d$ and $\leq'$ on $\mathcal{K}$. However, the present framework allows for any partial preorder on arguments that satisfies two basic assumptions (taken from Vreeswijk (1993)).

DEFINITION 3.10 (admissible argument orderings) Let $\mathcal{A}$ be a set of arguments. Then a partial preorder $\preceq$ on A is an *argument ordering* iff

(1) if A is firm and strict and B is defeasible or plausible, then $B \prec A$;
(2) if $A = A_1, \ldots, A_n \rightarrow \psi$, then for all $1 \leq i \leq n$, $A \preceq A_i$ and for some $1 \leq i \leq n$, $A_i \preceq A$.

(Vreeswijk also assumes that an argument cannot be stronger than its weakest subargument but in Section 6 the so-called 'last-link' principle will be discussed, which violates this assumption.) The first condition says that strict-and-firm arguments are stronger than all other arguments, while the second condition says that a strict inference cannot make an argument weaker or stronger.

DEFINITION 3.11 (argumentation theories) An *argumentation theory* is a triple $AT = (AS, \{KB, \preceq)$, where AS is an argumentation system, KB a knowledge base in AS and $\preceq$ an argument ordering on the set of all arguments that can be constructed from KB in AS (below called the set of arguments on the basis of AT).

3.3. *Attack and defeat*

Dung's use of the term 'attack' might at first sight lead to the belief that Dung's framework has no place for preferences. However, Dung's attack relation can also be seen as *abstracting* from the use of preferences: in this view, an attack relation in his framework may be the result of applying preferences to a syntactic conflict. This view on Dung's attack relation was, to my knowledge, first used by Prakken and Sartor (1997), it was also employed by Amgoud and Cayrol (2002) and it was the basis of Bench-Capon's (2003) value-based *AF*s. It was also the reason why Prakken and Sartor (1997) and Prakken and Vreeswijk (2002) replaced Dung's term 'attack' with 'defeat', to reflect that it may incorporate evaluative considerations. This convention will also be adopted in the present paper, while the term 'attack' will be reserved for non-evaluative syntactic notions of conflict. The idea then is that defeat is determined by attack plus preference (except in some cases, where attack automatically leads to defeat).

The notion of a defeasible inference rule naturally leads to two notions of rebutting and undercutting attack, introduced by Pollock (1974) and first formalised by Pollock (1987). The third kind of attack, premise attack (in this paper called undermining), is a natural addition (and for deductive inferences it is the only kind of attack) but highlights the philosophical distinction between plausible and defeasible reasoning discussed above. It was independently introduced by Vreeswijk (1993, chap. 8) and Elvang-Göransson, Fox, and Krause (1993). In line with Prakken and Sartor (1997), rebutting and undercutting attacks can also be launched on subarguments. This is essential in making the system an instantiation of Dung's abstract framework.

3.3.1. Attack

First the ways in which arguments can be attacked are defined. Recall that these are just syntactic categories and do not reflect any preference between arguments. The first way of attack corresponds to the case where one argument uses a defeasible rule of which another argument says that it does not apply to the case at hand. Its definition assumes that inference rules can be named in the object language; the precise nature of this naming convention will be left implicit.

DEFINITION 3.12 (undercutting attack) Argument A *undercuts* argument B (on B') iff $\text{Conc}(A) \in \overline{B'}$ for some $B' \in \text{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \psi$.

Example 3.13 In Example 3.7, argument A_8 can be undercut in two ways: by an argument with conclusion $\overline{A_5}$, which undercuts A_8 on A_5, and by an argument with conclusion $\overline{A_7}$, which undercuts A_8 on A_7.

Undercutting attackers only say that there is some exceptional situation in which a defeasible inference rule cannot be applied, without drawing the opposite conclusion. Rebutting attacks do the latter: they provide a contrary or contradictory conclusion for a defeasible (sub-)conclusion of the attacked argument.

DEFINITION 3.14 (rebutting attack) Argument A *rebuts* argument B on (B') iff $\text{Conc}(A) \in \overline{\varphi}$ for some $B' \in \text{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$. In such a case A *contrary-rebuts* B iff $\text{Conc}(A)$ is a contrary of φ.

Example 3.15 In Example 3.7, argument A_8 can be rebutted on A_5 with an argument for $\overline{t}$ and on A_7 with an argument for $\overline{v}$. Moreover, if $\overline{t} = -t$ then A_5 in turn rebuts any argument for $\overline{t}$ with a defeasible top rule. However, A_8 itself does not rebut that argument, except in the special case where $w \in \overline{\overline{t}}$. This shows that for three reasons rebutting attack is not symmetric: the rebuttal can have a strict top rule, rebutting can be contrary-rebutting and rebutting can be launched on a subargument. However, the present example also shows that in the latter case, if the rebutting attack has a defeasible top rule and is not of the contrary-rebutting kind, the directly rebutted subargument in turn rebuts its attacker.

The final way of attack is an attack on a (non-axiom) premise.

DEFINITION 3.16 (undermining attack) Argument A *undermines* B (on φ) iff $\text{Conc}(A) \in \overline{\varphi}$ for some $\varphi \in \text{Prem}(B) \backslash \mathcal{K}_n$. In such a case, argument A *contrary-undermines* B iff $\text{Conc}(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$.

Example 3.17 In Example 3.7, argument A_8 can be undermined with an argument that has conclusion $\overline{p}$, $\overline{r}$ or $\overline{u}$. If that attacker has a defeasible top rule and, say, a conclusion $\overline{p}$ and does not contrary-undermine A_8, then p as an argument in turn rebuts the attacker.

The following example (based on Example 4 of Caminada and Amgoud (2007)) illustrates the interplay between strict and defeasible rules in rebutting attack.

Example 3.18

$$\begin{array}{lll} A_1\text{: } \textit{WearsRing} & A_2\text{: } A_1 \Rightarrow \textit{Married} & A_3\text{: } A_2 \rightarrow \neg \textit{Bachelor} \\ B_1\text{: } \textit{Partyanimal} & B_2\text{: } B_1 \Rightarrow \textit{Bachelor} & B_3\text{: } B_2 \rightarrow \neg \textit{Married} \end{array}$$

A_3 rebuts B_3 on its subargument B_2 while B_3 rebuts A_3 on its subargument A_2. Note that A_2 does not rebut B_3, since B_3 applies a strict rule; likewise for B_2 and A_3.

3.3.2. *Defeat*

Now that we know how arguments can be attacked, the argument ordering can be used to define which attacks result in defeat. For undercutting attack, no preferences will be needed to make it result in defeat, since otherwise a weaker undercutter and its stronger target might be in the same extension. This would be strange since then the extension contains an argument that applies an inference rule of which another argument in the same extension says that it should not be applied.[4] The same holds for the other two ways of attack as far as they involve contraries (i.e. non-symmetric conflict relations between formulas). The reason for this is that otherwise if a rebutting or undermining attacker is weaker than its target, both may be in the same extension. For the remaining forms of attack, the argument ordering will be used to determine whether they result in defeat.

Definition 3.19 (successful rebuttal) Argument A successfully rebuts argument B if A rebuts B on B' and either A contrary-rebuts B' or $A \not\prec B'$.

This definition determines whether a rebutting attack is successful by comparing the conflicting arguments at the points where they conflict. Thus, in Example 3.18, the conflict between A_3 and B_3 is resolved by comparing A_3 with B_2 and comparing B_3 with A_2. Now if $B_2 \prec A_3$ (for example, since the married-rule is given priority over the bachelor-rule) then A_3 successfully rebuts B_2 and B_3 while B_3 does not successfully rebut A_2 or A_3. If, in contrast, $A_2 \not\prec B_3$ and $B_2 \not\prec A_3$ then both A_3 and B_3 successfully rebut each other (while A_3 still successfully rebuts B_2 and not vice versa, and likewise for B_3 and A_2). Note also that if A_3 is deleted from the example, then if $B_3 \prec A_2$, no argument in the example is defeated. This may at first sight seem counterintuitive but this is due to the fact that the example violates closure of R_s under transposition (cf. Section 5).

As noted by Caminada and Amgond (2007), Example 3.18 also illustrates why Definitions 3.14 and 3.19 should not allow that a defeasible argument with a strict top rule can be (successfully) rebutted on its final conclusion. The reason is that otherwise if all defeasible rules in the example are of equal preference, the set $\{A_1, A_2, B_1, B_2\}$ is admissible, which violates the rationality postulate of indirect consistency (see Section 6).

Definition 3.20 (successful undermining) Argument A successfully undermines B if A undermines B on φ and either A contrary-undermines B or $A \not\prec \varphi$.

This definition exploits that an argument premise is also defined to be a subargument.

In Example 3.7, any argument for $\overline{r}$ successfully undermines A_8 since it contrary-undermines it since $r \in \mathcal{K}_a$. The same holds for any argument for a contrary of p or u while for arguments for contradictories of p or u this depends on the argument ordering (which may in turn depend on the ordering $\preceq'$ on $\mathcal{K}$; see Definitions 6.14 and 6.17).

It remains to be discussed how the framework should deal with arguments that have issue premises. As explained above, the idea is that arguments with issue premises are always unacceptable. There are various ways to formalise this idea. One would be to let a special designated argument, or perhaps all strict-and-firm arguments, defeat any argument with an issue premise (as in Modgil (2009) and Prakken and Sartor (1997)). Here another solution is adopted: an argument can defeat another only if it has no issue premises. Then in Definition 2.1, only sets $\mathcal{B}$ with no issue premises will be considered, so that no argument with issue premises is in any extension.

The three defeat relations can now be combined into an overall definition of 'defeat'.

Definition 3.21 (defeat) Argument A *defeats* argument B iff no premise of A is an issue and A undercuts or successfully rebuts or successfully undermines B. Argument A *strictly defeats* argument B if A defeats B and B does not defeat A.

In the literature other combinations of these kinds of attack have been considered. For example, Prakken and Sartor (1997) (who have no undermining) give precedence to undercutting defeat over rebutting defeat, so that if A successfully undercuts B while B successfully rebuts A, nevertheless A strictly defeats B. It remains to be investigated how crucial the present definition is for the results below.

Finally, argumentation theories can be linked to Dung-style argumentation frameworks.

DEFINITION 3.22 (AF) An abstract argumentation framework AF corresponding to an argumentation theory AT is a pair $< \mathcal{A}, Def >$ such that:

- $\mathcal{A}$ is the set of arguments on the basis of AT as defined by Definition 3.6,
- Def is the relation on $\mathcal{A}$ given by Definition 3.21.

To leave arguments with issue premises out of any extension, Definition 2.1 should now start with 'Let $\mathcal{B}$ be a conflict-free set of arguments that have no issue premises ...'.

It is now also possible to define a consequence notion for well-formed formulas. Several definitions are possible. One is as follows.

DEFINITION 3.23 (acceptability of conclusions) For any semantics S and for any argumentation theory AT and formula $\varphi \in \mathcal{L}_{AT}$:

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

An alternative definition of skeptical acceptability is

(1) φ is *skeptically S-acceptable* in AT if and only if there exists an argument with conclusion φ that is contained in all S-extensions of AT.

While the original definition allows that different extensions contain different arguments for a skeptical conclusion, the alternative definition requires that there is one argument for it that is in all extensions.

4. Using the framework: domain-specific *vs.* general inference rules

The framework defined in the previous section can be used in two ways, depending on whether the inference rules are domain-specific or not. The inference rules of argumentation systems are not part of the logical language $\mathcal{L}$ but are metalevel constructs. The usual practice in standard logic is that inference rules express general patterns of reasoning, such as *modus ponens*, universal instantiation and so on. Yet Caminada and Amgoud (2007) use the inference rules to represent domain knowledge, in line with a long tradition in non-monotonic logic of using domain-specific inference rules (e.g. Reiter 1980; Loui 1987; Nute 1994; Garcia and Simari 2004). The difference between both approaches is illustrated with the following example. Consider the information that all Frisians are Dutch, that the Dutch are usually tall and that Wiebe is Frisian. With domain-specific inference rules, this can in a propositional language be represented as follows:

$\mathcal{R}_s = \{Frisian \rightarrow Dutch\}$
$\mathcal{R}_d = \{Dutch \Rightarrow Tall\}$
$\mathcal{K}_p = \{Frisian\}$

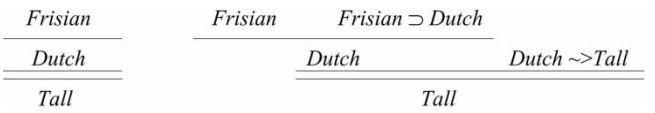

Figure 2. Domain-specific *vs.* general inference rules.

The argument that Wiebe is tall then has the form as displayed on the left in Figure 2.

With general inference rules, the two rules must instead be represented in the object language $\mathcal{L}$. The first one can be represented with the material implication but for the second one a connective for defeasible conditionals must be added to $\mathcal{L}$ and a defeasible *modus-ponens* inference rule must be added for this connective. For example:

$\mathcal{R}_s = \{\varphi, \varphi \supset \psi \rightarrow \psi \text{ (for all } \varphi, \psi \in \mathcal{L}), \ldots\}$
$\mathcal{R}_d = \{\varphi, \varphi \rightsquigarrow \psi \Rightarrow \psi \text{ (for all } \varphi, \psi \in \mathcal{L}), \ldots\}$
$\mathcal{K}_p = \{Frisian \supset Dutch, Dutch \rightsquigarrow Tall, Frisian\}$

Then the argument that Wiebe is tall has the form as displayed on the right in Figure 2.

Although the present system can be used both ways, both Vreeswijk and Pollock intended their inference rules to express general patterns of reasoning, which is much more in line with the role of inference rules in standard logic. Indeed, an important part of John Pollock's work was the study of general patterns of (epistemic) defeasible reasoning, which he called prima facie reasons. He formalised prima facie reasons for reasoning patterns involving perception, memory, induction, temporal persistence and the statistical syllogism, as well as undercutters for these reasons. The ASPIC framework allows for such general use of inference rules, by expressing the rules through schemes (in the logical sense, with metavariables ranging over $\mathcal{L}$). When used thus, the framework becomes a general framework for argumentation with structured arguments. It thus is also suitable for modelling reasoning with argument schemes, which currently is an important topic in the computational study of argument (cf. Walton et al. 2008). Argument schemes are stereotypical non-deductive patterns of reasoning, consisting of a set of premises and a conclusion that is presumed to follow from them. Uses of argument schemes are evaluated in terms of critical questions specific to the scheme. An example of an epistemic argument scheme is the scheme from expert opinion (Walton et al. 2008, p. 310):

E is an expert in domain *D*
E asserts that *P* is true
P is within *D*
P is true

This scheme has six critical questions:

(1) How credible is *E* as an expert source?
(2) Is *E* an expert in domain *D*?
(3) What did *E* assert that implies *P*?
(4) Is *E* personally reliable as a source?
(5) Is *P* consistent with what other experts assert?
(6) Is *E*'s assertion of *P* based on evidence?

A natural way to formalise reasoning with argument schemes is to regard them as defeasible inference rules and to regard critical questions as pointers to counterarguments (this approach was earlier defended by Bex, Prakken, Reed, and Walton (2003) and Verheij (2003b). More

precisely, the three kinds of attack on arguments correspond to three kinds of critical questions of argument schemes. Some critical questions challenge an argument's premise and therefore point to undermining attacks, others point to undercutting attacks, while again other questions point to rebutting attacks. In the scheme from expert opinion questions (2) and (3) point to underminers (of, respectively, the first and second premise), questions (4), (1) and (6) point to undercutters (the exceptions that the expert is biased or incredible for other reasons and that he makes scientifically unfounded statements) while question (5) points to rebutting applications of the expert opinion scheme. Thus, we also see that Pollock's prima facie reasons are examples of epistemic argument schemes and that his undercutters are negative answers to one kind of critical question.

Now one benefit of having undermining attack in addition to rebutting and undercutting attack can be discussed in more detail: if the inference rules are supposed to be domain-independent, then representing facts with non-conditional inference rules (as done by Caminada and Amgoud (2007)) does not make sense.

5. Transposition and contraposition

Before it can be studied to what extent the present framework satisfies the rationality postulates of Caminada and Amgoud (2007), first some technicalities concerning strict inference rules must be discussed. To start with, Caminada and Amgoud define the notions of a transposition of a strict rule and closure of sets of strict rules under transposition.

DEFINITION 5.1 (transposition) A strict rule s is a *transposition* of $\varphi_1, \ldots, \varphi_n \rightarrow \psi$ iff $s = \varphi_1, \varphi_{i-1}, \varphi_{i-1}, -\psi, \varphi_{i+1}, \ldots, \varphi_n \rightarrow -\varphi_i$ for some $1 \leq i \leq n$.

DEFINITION 5.2 (transposition operator) Let $\mathcal{R}_s$ be a set of strict rules. $Cl_{tp}(\mathcal{R}_s)$ is the smallest set such that:

- $\mathcal{R}_s \subseteq Cl_{tp}(\mathcal{R}_s)$ and
- If $s \in Cl_{tp}(\mathcal{R}_s)$ and t is a transposition of s, then $t \in Cl_{tp}(\mathcal{R}_s)$.

We say that $\mathcal{R}_s$ is *closed under transposition* iff $Cl_{tp}(\mathcal{R}_s) = \mathcal{R}_s$.

Now the subclass of argumentation systems closed under transposition can be defined.

DEFINITION 5.3 (closure under transposition) An argumentation system $(\mathcal{L}, ^-, \mathcal{R}, \leq)$ is *closed under transposition* if $\mathcal{R}_s = Cl_{tp}(\mathcal{R}_s)$. An argumentation theory is closed under transposition if its argumentation system is.

Caminada and Amgoud (2007) also define the closure of a set of formulas under application of strict rules.

DEFINITION 5.4 (closure of a set of formulas) Let $\mathcal{P} \subseteq \mathcal{L}$. The closure of $\mathcal{P}$ under the set $\mathcal{R}_s$ of strict rules, denoted $Cl_{\mathcal{R}s}(\mathcal{P})$, is the smallest set such that:

- $\mathcal{P} \subseteq Cl_{\mathcal{R}s}(\mathcal{P})$
- if $\varphi_1, \ldots, \varphi_n \rightarrow \psi \in \mathcal{R}_s$ and $\varphi_1, \ldots, \varphi_n \in Cl_{\mathcal{R}s}(\mathcal{P})$, then $\psi \in Cl_{\mathcal{R}s}(\mathcal{P})$.

If $\mathcal{P} = Cl_{\mathcal{R}s}(\mathcal{P})$, then $\mathcal{P}$ is said to be *closed*.

It is also relevant whether strict inference satisfies contraposition.

DEFINITION 5.5 (closure under contraposition) An argumentation system is *closed under contraposition* if for all $S \subseteq \mathcal{L}$, all $s \in S$ and all φ it holds that if $S \vdash \varphi$ then $S \backslash \{s\} \cup \{-\varphi\} \vdash - s$. An argumentation theory is closed under contraposition if its argumentation system is.

Closure under transposition does not imply closure under contraposition, as shown by the following counterexample (in all examples below, sets which are empty are not listed).

Example 5.6 Let $\mathcal{R}_s = Cl_{tp}(\{p \rightarrow q; p \rightarrow r; p, r \rightarrow s\})$. Then $\{p\} \vdash s$ but $\{-s\} \not\vdash - p$.

In general, it neither holds that closure under contraposition implies closure under transposition, as shown by the following counterexample.

Example 5.7 Let $\mathcal{R}_s = \{p \rightarrow q;\ \neg q \rightarrow r;\ r \rightarrow \neg p;\ \neg r \rightarrow q;\ p \rightarrow \neg r\}$. Then $\mathcal{R}_s$ is not closed under transposition, since it does not include $\neg q \rightarrow \neg p$. Still we have

$$\{p\} \vdash q \text{ and } \{\neg q\} \vdash \neg p \qquad \{p\} \vdash \neg r \text{ and } \{r\} \vdash \neg p$$

$$\{\neg r\} \vdash q \text{ and } \{\neg q\} \vdash r \qquad \{\neg q\} \vdash r \text{ and } \{\neg r\} \vdash q$$

So $\mathcal{R}_s$ satisfies contraposition.

However, contraposition does imply transposition in the following special case.

PROPOSITION 5.8 *Consider any argumentation theory with $\mathcal{L}$ closed under classical negation and $^-$ defined correspondingly. Then if $\mathcal{R}_s$ consists of all valid propositional inferences, then $\mathcal{R}_s$ is closed under contraposition and transposition.*

Note that the proposition does not hold if the condition '$\mathcal{R}_s$ consists of all valid propositional inferences' is changed to '$\vdash$ corresponds to propositional logic'. A counterexample is any argumentation theory with a sound and complete axiomatisation of propositional logic with *modus ponens* as the only inference rule.

6. Rationality postulates

Dung's semantics can be seen as rationality constraints on evaluating arguments in abstract argumentation frameworks. The refinement of his abstract approach with structured arguments naturally leads to the question whether this additional structure gives rise to additional rationality constraints. Caminada and Amgoud (2007) gave a positive answer to this question by proposing a number of 'rationality postulates' for what they called 'rule-based argumentation'. Four of their postulates formulate constraints on any extension of an argumentation framework corresponding to an argumentation theory:[5]

- *Closure under subarguments:* for every argument in an extension also all its subarguments 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.

Caminada and Amgoud (2007) proved for their version of the ASPIC framework that the first two postulates are always satisfied while the two consistency postulates are satisfied if the set of strict rules is consistent and closed under transposition. However, their version of the ASPIC framework is considerably simpler than the present one. First, it has no knowledge base and facts must be represented as inference rules with empty antecedents; because of this, arguments cannot be undermined. Furthermore, it assumes just a basic ordering on arguments, according to which strict arguments are strictly preferred over defeasible ones and nothing else. Finally, it has a special case of the present $^-$ function from $\mathcal{L}$ to $2^{\mathcal{L}}$, corresponding to classical negation.

The task now is to investigate to which extent the results of Caminada and Amgoud (2007) can be generalised to the present case.

The postulates of closure under subarguments and strict-rule application still hold unconditionally for the present framework. (Here that a given semantics is subsumed by complete semantics means that any of its extensions also is a complete extension).

PROPOSITION 6.1 *Let $< \mathcal{A}, Def >$ be an argumentation framework as defined in Definition 3.22 and E any of its extensions under a given semantics subsumed by complete semantics. Then for all $A \in E$: if $A' \in \text{Sub}(A)$ then $A' \in E$.*

PROPOSITION 6.2 *Let $< \mathcal{A}, Def >$ be an argumentation framework corresponding to an argumentation theory and E any of its extensions under a given semantics subsumed by complete semantics. Then* $\{\text{Conc}(A) | A \in E\} = \mathcal{C}_{Rs}(\{\text{Conc}(A) | A \in E\})$.

As for the two consistency postulates, Caminada and Amgoud's results do not generalise unconditionally. Consider the following example.

Example 6.3 Let $\mathcal{R}_d = \{\Rightarrow p;\ \Rightarrow q\}$ and $\mathcal{R}_s = \{q \rightarrow \neg p;\ p \rightarrow \neg q\}$. Then we have

A: $\Rightarrow p$
B': $\Rightarrow q$ $\quad B$: $B' \rightarrow \neg p$

Now assume that $A \succ B$, so B does not defeat A. However, A neither defeats B, since B's last inference is strict. At first sight, it would seem that A can be extended with the transposition of $q \rightarrow \neg p$ (i.e. with $p \rightarrow \neg q$) to an argument

A^+: $A \rightarrow \neg q$

that rebuts B's subargument B' for q. Then since by condition (2) of Definition 3.10 a strict continuation of an argument cannot make it weaker, $B' \prec A^+$ so A^+ defeats B'. Moreover, by the same conditions any argument defeats A if and only if it defeats A^+ so if A is in an extension E then by Proposition 6.2 A^+ will be in E and therefore B will not be in E since extensions are conflict-free.

However, this line of reasoning does not hold without a further assumption on the argument ordering. Consider a more complex variant of Example 6.3.

Example 6.4 Let $\mathcal{R}_d = \{\Rightarrow p;\ \Rightarrow q;\ \Rightarrow r\}$ and $\mathcal{R}_s = \{q, r \rightarrow \neg p;\ q, p \rightarrow \neg r;\ p, r \rightarrow \neg q\}$. Then we have

A: $\Rightarrow p$
B': $\Rightarrow q$ $\quad B''$: $\Rightarrow r$ $\quad B$: B', $\quad B'' \rightarrow \neg p$

The problem is that A cannot be extended with any transposition of $q, r \rightarrow \neg p$ to obtain A^+ unless it is combined with either B' or B'' but then A is extended with a defeasible rule, so A^+ might be weaker than A. This problem holds whenever B has more than one maximal defeasible or plausible subargument.

However, assuming contraposition or transposition, direct consistency can still be proved if it can also be assumed that there is a way to extend A with all but one of B's maximal defeasible subarguments that is not weaker than the remaining one. In our example, this means that either A extended with B' is not weaker than B'' or A extended with B'' is not weaker than B'. Intuitively, this assumption seems acceptable given that A is stronger than both B' and B''. It is therefore to be expected that it will be satisfied by many reasonable argument orderings. Since similar situations can arise with undermining attack, the notion of a maximal fallible subargument is needed.

DEFINITION 6.5 (maximal fallible subarguments) For any argument A, an argument $A' \in \text{Sub}(A)$ is a *maximal fallible subargument* of A if

(1) A''s final inference is defeasible or A' is a non-axiom premise; and
(2) there is no $A'' \in \text{Sub}(A)$ such that $A'' \neq A$ and $A' \in \text{Sub}(A'')$ and A'' satisfies condition (1).

The set of maximal fallible subarguments of an argument A will be denoted by $M(A)$.

COROLLARY 6.6 *For any argument A, it holds that* $\text{Conc}(M(A)) \vdash \text{Conc}(A)$.

DEFINITION 6.7 (reasonable argument orderings) Argument ordering $\preceq$ is *reasonable* if it satisfies the following condition. Let A and B be arguments with contradictory conclusions such that $B \prec A$. Then there exists a $B_i \in M(B)$ and an A^+ with $A \in \text{Sub}(A^+)$ such that $\text{Conc}(A^+) = -\text{Conc}(B_i)$ and $A^+ \not\prec B_i$.

A final problem to deal with is that in Example 6.3, $\text{Conc}(A)$ could be a contrary of $\text{Conc}(B)$; the problem is that the solution with closure under contraposition and transposition does not apply to this case. Therefore, the focus must be restricted to argumentation theories that respect the intended use of assumptions and contraries.

DEFINITION 6.8 An argumentation theory is *well formed* if:

(1) no consequent of a defeasible rule is a contrary of the consequent of a strict rule;
(2) if $\varphi \in \mathcal{K}_a$ and φ is a contrary of ψ, then $\psi \notin \mathcal{K}_n \cup \mathcal{K}_p$ and ψ is not the conclusion of a rule in $\mathcal{R}$.

Condition (2) in effect says that assumptions can only be contraries of other assumptions. An example of an argumentation theory that is not well formed is

$$\mathcal{R}_s = \{p \rightarrow q\}, \quad \mathcal{R}_d = \{r \Rightarrow s, t \Rightarrow u\}, \quad \mathcal{K}_p = \{p, r\}, \quad \mathcal{K}_a = \{v\}$$

and such that s is a contrary of q and v is a contrary of u. Then condition (1) of Definition 6.8 is violated since we have arguments A: $p \rightarrow q$ and B: $r \Rightarrow s$. Moreover, condition (2) is violated since $v \in \mathcal{K}_a$ and $t \Rightarrow u \in \mathcal{R}_d$.

Now it can be proved that under certain conditions an argumentation theory satisfies the postulate of direct consistency.

THEOREM 6.9 *Let* $< \mathcal{A}$, *Def* $>$ *be an argumentation framework corresponding to a well-formed argumentation theory that is closed under contraposition or transposition and has a reasonable argument ordering and a consistent* $Cl_{Rs}(\mathcal{K}_n)$, *and let E be any of its extensions under a given semantics subsumed by complete semantics. Then the set* $\{\text{Conc}(A) | A \in E\}$ *is consistent.*

Caminada and Amgoud (2007) also prove that their system satisfies the postulate of indirect consistency. This follows from their Proposition 7, which says that if an argumentation theory satisfies closure and direct consistency, it also satisfies indirect consistency. Since in the present case, the conditions of the proof of direct consistency had to be strengthened, the same holds for indirect consistency.

THEOREM 6.10 *Let* $<A$, *Def* $>$ *be an argumentation framework corresponding to a well-formed argumentation theory that is closed under contraposition or transposition and has a reasonable argument ordering and a consistent* $Cl_{Rs}(K_n)$, *and let E be any of its extensions under a given semantics subsumed by complete semantics. Then the set* $Cl_{Rs}(\{\text{Conc}(A) | A \in E\})$ *is consistent.*

COROLLARY 6.11 *If the conditions of Theorem 6.10 are satisfied, then for any extension E under a given semantics subsumed by complete semantics the set* $\{\varphi|\varphi$ *is a premise of an argument in* $E\}$ *is consistent.*

Concluding this section, two intuitively plausible argument orderings will be shown to be reasonable, namely, the weakest-link and last-link orderings from Amgoud et al. (2006). The versions below are slightly revised to make the principles arguably more intuitive. Both orderings define a strict partial order $\prec_s$ on sets in terms of a partial preorder $\preceq_e$ on their elements, as follows: $S_1 \prec_s S_2$ iff there exists an $e_1 \in S_1$ such that for all $e_2 \in S_2$ it holds that $e_1 <_e e_2$.

The *last-link principle* prefers an argument A over another argument B if the last defeasible rules used in B are less preferred than the last defeasible rules in A or, in case both arguments are strict, if the premises of B are less preferred than the premises of A. The concept of 'last defeasible rules' is defined as follows and is essentially the same as Prakken and Sartor's (1997) notion of a 'relevant set'.

DEFINITION 6.12 (last defeasible rules) Let A be an argument.

- $\text{LastDefRules}(A) = \emptyset$ iff $\text{DefRules}(A) = \emptyset$.
- If $A = A_1, \ldots, A_n \Rightarrow \phi$, then $\text{LastDefRules}(A) = \{\text{Conc}(A_1), \ldots, \text{Conc}(A_n) \Rightarrow \phi\}$, otherwise $\text{LastDefRules}(A) = \text{LastDefRules}(A_1) \cup \ldots \cup \text{LastDefRules}(A_n)$.

COROLLARY 6.13 $\text{LastDefRules}(A) = \{\text{TopRule}(A')|A' \in M(A)\}$.

An example with more than one last defeasible rule is with $\mathcal{K} = \{p; q\}$ and $\mathcal{R}_d = \{p \Rightarrow r;\ q \Rightarrow s\}$. Then for argument A for $r \wedge s$, we have $\text{LastDefRules}(A) = \{p \Rightarrow r;\ q \Rightarrow s\}$.

The above definition is now used to compare pairs of arguments as follows.

DEFINITION 6.14 (last link principle) Let A and B be two arguments. Then $A \prec B$ iff either

(1) condition (1) of Definition 3.10 holds or
(2) $\text{LastDefRules}(A) \prec_s \text{LastDefRules}(B)$ or
(3) $\text{LastDefRules}(A)$ and $\text{LastDefRules}(B)$ are empty and $\text{Prem}(A) \prec_s \text{Prem}(B)$.

(Amgoud et al. 2006 do not include the second condition so if both arguments are strict the ordering on the knowledge base is ignored.) This definition in effect compares sets on their weakest elements.

PREPOSITION 6.15 *The last-link argument ordering is reasonable.*

Consider the following example (taken from Prakken 1997) on whether people misbehaving in a university library may be denied access to the library.

Example 6.16 Let $\mathcal{K}_p = \{Snores;\ Professor\}$, $\mathcal{R}_d =$

$\{Snores \Rightarrow_{r_1} Misbehaves;$
$Misbehaves \Rightarrow_{r_2} AccessDenied;$
$Professor \Rightarrow_{r_3} \neg AccessDenied\}$.

Assume that $Snores <' Professor$ and $r_1 < r_2$, $r_1 < r_3$, $r_3 < r_2$ and consider the following arguments.

A_1: *Snores* $\quad A_2$: $A_1 \Rightarrow Misbehaves$ $\quad A_3$: $A_2 \Rightarrow AccessDenied$
B_1: *Professor* $\quad B_2$: $B_1 \Rightarrow \neg AccessDenied$

To resolve the conflict between A_3 and B_2, the rule sets to be compared are $\text{LastDefRules}(A_3) = \{r_2\}$ and $\text{LastDefRules}(B_2) = \{r_3\}$. Since $r_3 < r_2$ we have that $B_2 \prec_s A_3$ so A_3 strictly defeats B_2.

The *weakest-link principle* considers not the last but all uncertain elements in an argument. It prefers an argument A over an argument B if A is preferred to B on both their premises and their defeasible rules.

DEFINITION 6.17 (weakest link principle) Let A and B be two arguments. Then $A \prec B$ iff either condition (1) of Definition 3.10 holds or

(1) $\text{Prem}(A) \prec_s \text{Prem}(B)$ and
(2) If $\text{DefRules}(B) \neq \emptyset$, then $\text{DefRules}(A) \prec_s \text{DefRules}(B)$.

(Amgoud et al. (2006) do not have condition (2), so that with two strict arguments neither of them can be preferred.)

PROPOSITION 6.18 *The weakest-link argument ordering is reasonable.*

Example 6.19 Consider again Example 6.16. With the weakest-link principle, the outcome is different. To resolve the conflict between A_3 and B_2, the rule sets to be compared are now $\text{DefRules}(A_3) = \{r_1, r_2\}$ and $\text{DefRules}(B_2) = \{r_3\}$. Since $r_1 < r_3$, we have that $\text{DefRules}(A_3) \prec_s \text{DefRules}(B_2)$. Moreover, since $\textit{Snores} <' \textit{Professor}$, we also have that $\text{Prem}(A_3) \prec_s \text{Prem}(B_2)$. Hence, B_2 now strictly defeats A_3.

Example 6.20 We finally return to Example 1. Let

$$r_1 = \textit{WearsRing} \Rightarrow \textit{Married}$$
$$r_2 = \textit{PartyAnimal} \Rightarrow \textit{Bachelor}$$

Note that since both arguments apply just one defeasible rule and no premise is attacked, the weakest- and last-link ordering produce the same result. Now if $r_1 < r_2$, we have that A_3 strictly defeats B_3 by successfully rebutting it on B_2, while if both $r_1 \not< r_2$ and $r_2 \not< r_1$ then A_3 and B_3 defeat each other since A_3 successfully rebuts B_3 on B_2 while B_3 successfully rebuts A_3 on A_2.

7. Self-defeat

As discussed by Pollock (1994) and Caminada and Amgoud (2007), self-defeating arguments can cause problems if argumentation systems are not carefully defined, particularly if they include standard propositional logic. In the present framework, two types of self-defeating arguments are possible: serial self-defeat occurs when an argument defeats one if its earlier steps, while parallel self-defeat occurs when the contradictory conclusions of two or more arguments are taken as the premises for $\bot$. Pollock (1994) gives an example of serial self-defeat of the following form.

Example 7.1 Let $\mathcal{R}_d = \{p \Rightarrow q\}$, $\mathcal{R}_s = \{\, q \rightarrow \neg A_2\}$ and $\mathcal{K} = \{p, q\}$. Then, we have

$$A_1\colon p \quad A_2\colon A_1 \Rightarrow q \quad A_3\colon A_2 \rightarrow \neg A_2$$

(Read p as 'witness John says that he is unreliable' and q as 'witness John is unreliable'). Argument A_3 is self-defeating since it undercuts itself on A_2. This example is arguably handled properly by preferred and grounded semantics, who both have $E = \{A_1\}$ as the only extension.

One of Pollock's (1994) examples of parallel self-defeat has the following form.

Example 7.2 Let $\mathcal{R}_d = \{p \Rightarrow q;\ r \Rightarrow \neg q;\ t \Rightarrow s\}$ and $\mathcal{K} = \{p, r, t\}$ while $\mathcal{R}_s$ contains all propositionally valid inferences. Then:

$$\begin{array}{ll} A_1\colon p & A_2\colon A_1 \Rightarrow q \\ B_1\colon r & B_2\colon B_1 \Rightarrow \neg q \\ C_1\colon A_2, B_2 \rightarrow \bot & C_2\colon C_1 \rightarrow \neg s \\ D_1\colon t & D_2\colon D_1 \Rightarrow s \end{array}$$

Here a problem arises since s can be any formula, so any defeasible argument unrelated to A_2 or B_2, such as D_2, can, depending on the rule priorities, be rebutted by C_2. Clearly, this is extremely harmful, since the existence of just a single case of mutual rebutting defeat, which is very common, could trivialise the system. In fact, of the semantics defined by Durg (1995), this is only a problem for grounded semantics. Since all preferred/stable extensions contain either A_2 or B_2, argument C_2 is not in any of these extensions so D_2 is. However, if neither of A_2 and B_2 strictly defeats the other, then neither of them is in the grounded extension so that extension does not defend D_2 against C_2 and therefore does not contain D_2.

Pollock (1994) also discusses the following variant of this example (with the same argumentation theory):

$$\begin{array}{lll} A_1\colon p & A_2\colon A_1 \Rightarrow q & A_3\colon A_2 \rightarrow q \vee \neg s \\ B_1\colon r & B_2\colon B_1 \Rightarrow \neg q & \\ C_1\colon A_2, B_2 \rightarrow \neg s & & \\ D_1\colon t & D_2\colon D_1 \Rightarrow s & \end{array}$$

Again with grounded semantics, the problem is that s can be any formula, so any defeasible argument unrelated to A_2 or B_2 can be rebutted by C_1.

According to Caminada (personal communication), the only way to solve this problem is to make parallel self-defeat impossible. One way to implement this solution is to disallow arguments with a contradictory set of subconclusions. However, this affects the proof of Theorems 6.9 and 6.10. The reason is that for such systems the argument A^+ that according to Lemma A1 can be constructed sometimes has to have contradictory sub-conclusions, as the following example (with a system closed under transposition) shows.

Example 7.3 Let $p \in \mathcal{K}_n$, $q \in \mathcal{K}_a$ and $\mathcal{R}_s = Cl_{tp}(\{p \rightarrow t;\ q \rightarrow r;\ q \rightarrow s;\ r, s \rightarrow \neg t\})$.

$$\begin{array}{llll} A_1\colon p & A_2\colon A_1 \rightarrow t & & \\ B_1\colon q & B_2\colon B_1 \rightarrow r & B_3\colon B_1 \rightarrow s & B_4\colon B_2, B_3 \rightarrow \neg t \end{array}$$

Now if A_2 is to be extended to an argument A^+ that undermines B_4, then B_1 must be included in A^+.

A similar example for systems closed under contraposition is as follows.

Example 7.4 Let $\mathrm{K}_p = \{p, q, \neg p, \neg q\}$ and let $\mathcal{R}_s$ consist of all valid propositional inferences. Then

$$\begin{array}{lll} A_1\colon p & A_1\colon q & A_3\colon A_1, A_2 \rightarrow p \wedge q \\ B_1\colon \neg p & B_2\colon \neg q & B_3\colon B_1, B_2 \rightarrow \neg(p \wedge q) \end{array}$$

Note that $M(B_3) = \mathrm{Prem}(B_3)$. Now any addition of a premise of B_3 to $\mathrm{Prem}(A_3)$ makes $\mathrm{Prem}(A_3)$ inconsistent.

Since these problems only arise in particular argumentation systems and with particular semantics, no general solution will be pursued here; instead, such solutions are left for future research on instantiations of the framework. Note also that Examples 7.3 and 7.4 only contain strict rules, so that the problem may also arise in assumption-based frameworks, which will in the next section be proved to be a special case of the ASPIC framework.

8. The relation with assumption-based argumentation

After having presented his fully abstract approach to argumentation, Dung joined Kowalski, Toni and others in their development of a more concrete version of his approach (e.g. Bondarenko et al. 1997; Dung et al. 2006, 2007). In this approach, arguments essentially are sets of formulas called 'assumptions', from which conclusions can be drawn with strict inference rules. Arguments can be attacked with arguments that conclude to the 'contrary' of one of their assumptions. In fact, the extensions defined by the various semantics of Bondarenko et al. (1997) are not sets of arguments but sets of assumptions. However, Dung et al. (2007) showed that an equivalent fully argument-based formulation can be given.

In this section, it will be shown that assumption-based argumentation is a special case of the present framework with only strict inference rules, only assumption-type premises and no preferences. The proof will be given for the argument-based version of Dung et al. (2007) and carries over to Bondarenko et al. (1997) by the equivalence result of Dung et al. (2007).

First the main definitions of ABA are recalled (in the formulation of Dung et al. (2007)).

DEFINITION 8.1 (Dung et al. 2007, Definition 2.3) A *deductive system* is a pair $(\mathcal{L}, \mathcal{R})$ where

- $\mathcal{L}$ is a formal language consisting of countably many sentences, and
- $\mathcal{R}$ is a countable set of inference rules of the form $\alpha_1, \ldots, \alpha_n \rightarrow \alpha$.[6] $\alpha \in \mathcal{L}$ is called the *conclusion* of the inference rule, $\alpha_1, \ldots, \alpha_n \in \mathcal{L}$ are called the *premises* of the inference rule and $n \geq 0$.

DEFINITION 8.2 (Dung et al. 2007, Definition 2.5) An *assumption-based argumentation framework* (*ABF*) is a tuple $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ where

- $(\mathcal{L}, \mathcal{R})$ is a deductive system,
- $\mathcal{A} \in \mathcal{L}, \mathcal{A} \neq \emptyset$. $\mathcal{A}$ is the set of *candidate assumptions*,
- If $\alpha \in \mathcal{A}$, then there is no inference rule of the form $\alpha_1, \ldots, \alpha_n \rightarrow \alpha \in \mathcal{R}$,
- $^-$ is a total mapping from $\mathcal{A}$ into $\mathcal{L}$. $\overline{\alpha}$ is the *contrary* of α.

The third condition amounts to a restriction to so-called flat ABFs. This restriction is not entirely innocent, since in debates it may occur that someone first assumes a premise and, after it is defeated, constructs an argument for it, in an attempt to rebut the defeater. To make Dung et al.'s analysis apply to all stages of such a debate, assumptions should be deleted from $\mathcal{A}$ as soon as they are supported with an argument.

Since the notion of an argument is central to the present concerns, the informal explanation of Dung et al. (2007, p. 646) will be quoted in (almost) full.

> Deductions can be understood as proof trees: the root of the tree is labelled by the conclusion of the deduction and the leaves are labelled by the premises supporting the deduction. For every non-terminal node in the tree, there is an inference rule whose conclusion matches the sentence labelling the node, and the children of the node are labelled by the premises of the inference rule. (...) we define deductions as sequences of frontiers $S_1, \ldots, S_m$ of the proof trees. Each frontier is represented by a multi-set, in which the same sentence can have several occurrences, if it is generated more than once as a premise of different inference steps. In order to generate proof trees, a selection strategy is

needed to identify which node to expand next. We formalise this selection strategy by means of a selection function, as in the formalisation of SLD resolution. A selection function, in this context, takes as input a sequence of multi-sets S_i and returns as output a sentence occurrence in S_i. We restrict the selection function so that if a sentence occurrence is selected in a multi-set in a sequence then it will not be selected again in any later multi-set in that sequence.

Essentially, a backward deduction thus presents one particular order in which an argument in the sense of Definition 3.6 can be constructed by reasoning backwards from the conclusion to the premises.

DEFINITION 8.3 (Dung et al. 2007, Definition 2.4) Given a selection function f, a (*backward*) deduction of a conclusion α based on (or supported by) a set of premises P is a sequence of multi-sets $S_1, \ldots, S_m$, where $S_1 = \{\alpha\}$, $S_m = P$, and for every $1 \leq i < m$, where σ is the sentence occurrence in S_i selected by f:

(1) If σ is not in P, then $S_{i+1} = S_i - \{\sigma\} \cup S$ for some inference rule of the form $S \rightarrow \sigma \in \mathcal{R}$.
(2) If σ is in P, then $S_{i+1} = S_i$.

Each S_i is a step in the deduction.

Now an assumption-based argument is defined as follows:

DEFINITION 8.4 (Dung et al. 2007, Definition 2.6) An *argument* for a conclusion on the basis of an *ABF* is a deduction of that conclusion whose premises are all assumptions (in $\mathcal{A}$).

As for notation, the existence of an argument for a conclusion α supported by a set of assumptions A is denoted by $A \vdash \alpha$, or by $A \vdash_{ABF} \alpha$ if it has to be distinguished from the existence of a strict argument according to Definition 3.6 with the same premises and conclusion; the latter will below be denoted by $A \vdash_{AT} \alpha$.

Finally, Dung et al.'s notion of argument attack is defined as follows.

DEFINITION 8.5 (Dung et al. 2007, Definition 2.7)

- an argument $A \vdash \alpha$ *attacks an argument* $B \vdash \beta$ if and only if $A \vdash \alpha$ attacks an assumption in B;
- an argument $A \vdash \alpha$ *attacks an assumption* β if and only if α is the contrary $\overline{\beta}$ of β.

The argumentation theory corresponding to an assumption-based framework is now defined as follows.

DEFINITION 8.6 Given an assumption-based framework $ABF = (\mathcal{L}_{ABF}, \mathcal{R}_{ABF}, \mathcal{A}, \bar{}_{ABF})$, the *corresponding argumentation theory* $AT_{ABF} = (AS, KB)$, where $AS = (\mathcal{L}_{AT}, \bar{}_{AT}, \mathcal{R}_{AT}, \leq)$ and $KB = (\mathcal{K}, \leq')$, is defined as follows:

- $\mathcal{L}_{AT} = \mathcal{L}_{ABF}$
- $\varphi \in \overline{\psi}_{AT}$ iff $\varphi = \overline{\psi}_{ABF}$
- $\mathcal{R}_{AT} = \mathcal{R}_s = \mathcal{R}_{ABF}$
- $\mathcal{K}_n = \mathcal{K}_p = \mathcal{K}_i = \emptyset$
- $\mathcal{K}_a = \mathcal{A}$
- $\leq = \leq' = \preceq = \emptyset$

Note that AT_{ABF} is well formed and all AT_{ABF} arguments are strict and plausible.

The main task now is to prove that there is an *ABF*-argument for α from P if and only if there is an AT_{ABF}-argument for α with premises P. In fact, this can only be proved for the special case

of argumentation theories that do not allow for arguments with an infinite number of subarguments. Technically, the present framework allows for such arguments even if they are non-circular. For example, an AT with $\mathcal{R}_s = \{p_{i+1} \rightarrow p_i | i \geq 1\}$ allows for an argument for p_1 with an infinite number of subarguments (and an empty set of premises). So far no proof has depended on finiteness of arguments. In an *ABF*, however, arguments are by definition finite even if the set of inference rules allows for infinite ones, as in the just-given example.

PROPOSITION 8.7 *For all ABF such that $AT = AT_{ABF}$ does not allow arguments with an infinite number of subarguments, there exists an argument $A \vdash_{ABF} \alpha$ if and only if there exists an argument $A \vdash_{AT} \alpha$.*

From this it follows that

PROPOSITION 8.8 *For all ABF such that $AT = AT_{ABF}$ does not allow arguments with an infinite number of subarguments, it holds for every argument $A \vdash_{ABF} \alpha$ and every argument $A \vdash_{AT} \alpha$ that $A \vdash_{ABF} \alpha$ is defeated by an argument $B \vdash_{ABF} \beta$* if and only if *$A \vdash_{AT} \alpha$ is defeated by an argument $B \vdash_{AT} \beta$.*

Now the main correspondence result can be proved.

THEOREM 8.9 *For all ABF, any semantics S subsumed by complete semantics and any set E*:

(1) *if E is an S-extension of ABF then E_{AT} is an S-extension of AT, where $E_{AT} = \{A \vdash_{AT} \alpha | A \vdash_{ABF} \alpha \in E\}$;*
(2) *if E is an S-extension of AT then E_{ABF} is an S-extension of ABF, where $E_{ABF} = \{A \vdash_{ABF} \alpha | A \vdash_{AT} \alpha \in E\}$.*

Theorem 8.9 in fact says that there is a one-to-one correspondence between the extensions of an *ABF* and those of its corresponding *AT*. From this we have the following:

COROLLARY 8.10 *For any* ABF, *any semantics S subsumed by complete semantics, and for any formula φ it holds that φ is skeptically (credulously) S-acceptable in ABF if and only if φ is skeptically (credulously) S-acceptable in AT_{ABF}.*

9. Other related research

As was said above, the present framework is inspired by the work of Pollock (1987, 1994) and Vreeswijk (1993, 1997). Essentially, it takes from both the idea that defeasible reasoning proceeds by chaining two kinds of inference rules into inference trees. The present mathematical formulation of this idea is directly adopted from Vreeswijk (1993, 1997). The present notions of undercutting and rebutting defeat are taken from Pollock's work and then generalised to arbitrary preference relations on arguments (Pollock only has a notion of probabilistic strength), and to logical languages with arbitrary contrary mappings. They are then combined with a notion of undermining defeat.

In fact, the system of Pollock (1994) is not formalised in terms of arguments but in terms of the so-called 'inference graphs', in which nodes are connected either by inference links (applications of inference rules) or by defeat links. The nodes are 'lines of argument', which are propositions plus an encoding of the argument lines from which they are derived. So if a proposition is derived in more than one way, it occurs in more than one line of argument. Such duplications cannot be avoided, since defeat relations depend on the strength of a proposition, which in turn depends on the way in which it is derived. Nodes are evaluated in terms of the recursive structure of the graph. Jakobovits and Vermeir (1999) proved that

Pollock's system can be given an equivalent formulation as an instance of Dung's abstract argumentation frameworks with preferred semantics.

With Vreeswijk's framework, the relation with Dung-style semantics is still an open issue, since it models conflict not as a relation between two individual arguments but as a property of *sets* of arguments: a set of arguments is said to be in conflict if there exists a strict argument from their conclusions for $\bot$. Vreeswijk then defines a notion of warrant for arguments which resembles stable semantics.

Gordon et al. (2007) proposed the Carneades framework 'of argument and burden of proof'. Carneades' main structure is that of an argument graph, which, despite its name, is similar to Pollock's inference graphs. Statement nodes are linked to each other via argument nodes, which record the inferences from one or more nodes to another. This notion of an argument does not have the recursive structure of Definition 3.6 but instead stands for a single inference step. As explained in Section 3.1, the premises of an argument can be of three types: presumptions (similar to the present issues), assumptions (similar to the present ordinary premises) and exceptions (similar to contradictories of the present assumptions). Carneades has no distinction between strict and defeasible inference rules and, unlike Pollock, does not express conflicts as a special type of link between statement nodes. Instead, inferences (i.e. arguments) can be either pro or con a statement. Because of this, statements occur only once in the graph. Also, attack relations are thus expressed either as arguments pro and con the same statement or as an argument pro an exception-type premise of another argument. Carneades thus allows for rebutting and undermining but not for undercutting; instead, undercutters are simulated by arguments pro exceptions. Carneades' inference graphs are assumed to contain no cycles, which excludes the representation of mutual attack relations through exceptions.

In Carneades, the evaluation of statements in an argument graph is, as with Pollock's inference graphs, defined in terms of the recursive structure of the graph. Statements are acceptable if they satisfy their 'proof standard'. The general framework abstracts from their nature but Gordon et al. (2007) give several examples of proof standards. The proof standards are at the heart of Carneades' acceptability notion, just like the notions of defence and admissibility are at the heart of Dung-style semantics. None of the examples given by Gordon et al. (2007) have a known relation with any existing Dung-style semantics or the present framework, which thus is an issue for future research. Here it is also relevant that Carneades incorporates dialogical elements since it matters whether a statement is 'stated', 'questioned', 'accepted' or 'rejected'. These statuses of a statement are assumed to be provided by a dialogical context in which Carneades is embedded.

Verheij (2003a) presents a 'sentence-based' (as opposed to 'argument-based') logic for defeasible reasoning, called DefLog. Verheij assumes a logical language with just two connectives, a unary connective $\times$ which informally stands for 'it is defeated that' and a binary connective $\rightsquigarrow$ for expressing defeasible conditionals. He then assumes a single inference scheme for this language, namely, *modus ponens* for $\rightsquigarrow$. A set of sentences T is said to *support* a sentence φ if 'φ is in T or follows from T by repeated application of $\rightsquigarrow$-*modus ponens*' (Verheij 2003a, p. 327). It seems reasonable to formalise this as the backward deductions of assumption-based argumentation or the strict arguments of the present framework. Moreover, T is said to *attack* φ if T supports $\times\varphi$. Verheij then considers partitions (J, D) of sets of sentences Δ which he calls *dialectical interpretations* and which are such that J (the 'justified' sentences) is conflict-free and attacks every sentence in D (the 'defeated' sentences).

As already suggested by Verheij, there is a close formal relation between DefLog and assumption-based argumentation. First, dialectical interpretations are easily proved to be equivalent to stable labellings, which are known to be equivalent to stable semantics (first proved by Verheij (1996); see also Jakobovits and Vermeir (1999), and Caminada (2006)). Furthermore, DefLog

theories can be mapped onto assumption-based frameworks by letting an *ABF* contrary mapping be $\times\varphi = \overline{\varphi}$ for any φ, by regarding any set of dialectically interpreted sentences as the assumptions $\mathcal{A}$ of an *ABF* and by having $\varphi, \varphi \rightsquigarrow \psi \rightarrow \psi$, for any φ and ψ in DefLog's language, as the set $\mathcal{R}$ of inference rules of the *ABF*. The result is an assumption-based framework in the sense of Definition 8.2 with stable semantics. The correspondence results of Dung et al. (2007) with Bondarenko et al. (1997) then also apply to the special case of a DefLog-style *ABF* so that by the above Theorem 8.9 DefLog is a special case of the present framework with only strict arguments and only undermining defeat.

Several argumentation systems model deductive argumentation. Here arguments are proofs according to some deductive logic with consistent premises taken from a possibly inconsistent knowledge base expressed in the language of the logic (usually taken to be standard propositional or first-order logic). In Amgoud and Cayrol (2002), which is based on propositional logic, the structure of arguments is left undefined, except that the premises imply the conclusion according to propositional logic. Several notions of defeat are then considered. One of them corresponds to the present undermining defeat, where arguments are compared in terms of a partial preorder on the belief base from which their premises are taken. Argument acceptability is defined according to grounded semantics.

This variant of Amgoud and Cayrol (2002) can be reconstructed as a special case of the present framework as follows. First, $\mathcal{L}$ is any propositional language closed under classical negation, where $\varphi = \overline{\psi}$ if $\varphi = \neg\psi$ or $\psi = \neg\varphi$. Then $\mathcal{R}_s$ consists of all valid propositional inferences while $\mathcal{R}_d$ is empty. The knowledge base equals $\mathcal{K}_p$. Finally, as with Deflog, it seems reasonable to formalise arguments as the strict arguments of the present framework, although the extra constraint must be added that such arguments have classically consistent premises. This consistency constraint makes that not all results of this paper hold without further qualification. It is easy to verify that Propositions 5.8, 6.1 and 6.2 still hold with this constraint (for Proposition 5.8 note that in this case $S \vdash \varphi$ by definition implies that the strict argument that exists for φ has consistent premises). However, the proofs of Theorems 6.9 and 6.10 do not apply to this case, for similar reasons as explained above in Section 7 with Example 7.4. It remains to be investigated whether these theorems can be proved for this case under alternative conditions.

Besnard and Hunter's (2008) version of deductive argumentation is similar to that of Amgoud and Cayrol (2002), except for a generalised notion of undermining: an argument is undermined by any argument of which the conclusion negates the conjunction if its premises. It remains to be seen whether this version of undermining can be reduced to the present version.

Two other logics for defeasible reasoning with both (domain-specific) strict and defeasible inference rules are Defeasible Logic (DL), first proposed by Nute (1994), and Defeasible Logic Programming (DeLP; e.g. Garcia and Simari 2004). In both systems, the logical language is restricted in logic-programming style. DL is not explicitly argument-based but defines the notion of a proof tree, which interleaves support and attack. Governatori, Maher, Antoniou, and Billington (2004) investigated the relation with Dung-style semantics. One variant of DL is proved to instantiate grounded semantics. In DeLP, the only way to attack an argument is on a (sub-)conclusion. DeLP's notion of argument acceptability has no known relation to any of the current argumentation semantics.

Prakken and Sartor (1997) presented an argument-based version of extended logic programming, designed as an instance of Dung's abstract argumentation frameworks with grounded semantics. Their system comes close to being a special case of the present framework. It has (domain-specific) strict and defeasible inference rules and allows for rebutting and undercutting defeat. Furthermore, its notion of an argument comes close to a 'deduction' version of Definition 3.6, i.e. it represents a particular order in which an argument can be constructed. A difference is that in Prakken and Sartor (1997) two parallel subarguments do not need to be completed with an

inference from their conclusion, so that, for example (in the present notation), $p, p \Rightarrow q, r, r \Rightarrow s$ is an argument with conclusions q and s. In Prakken and Sartor (1997), this was convenient for modelling reasoning about defeasible priorities in the system. A more substantial difference is that while the present framework considers rebutting and undercutting attack on equal footing, Prakken and Sartor (1997) give priority to undercutting attack, so that if A undercuts B while B rebuts A, A strictly defeats B. It seems that the present results do not crucially rely on this difference, but this should be further investigated.

A final difference with the present framework is that in Prakken and Sartor (1997) the role of strict rules in defeat is different. As in the present framework, only defeasible inferences can be attacked, but an argument A with conclusion φ rebuts an argument B with conclusion φ' if there exists sets of strict rules S_a and S_b and a formula ψ such that (with present notation) $S_a \cup \{\varphi\} \vdash \psi$ and $S_b \cup \{\varphi'\} \vdash \overline{\psi}$. The difference can be best explained with Examples 3.18 and 6.20. The motivation behind the definition of Prakken and Sartor (1997) was that intuitively the 'real' conflict is between the two defaults on whether someone is a bachelor or married. This is captured by their definition of rebutting attack, since A_2 can be extended with A_3 to contradict B_2's conclusion and vice versa. Hence the rule priorities are applied to A_2 and B_2. By contrast, in the present framework these arguments do not rebut each other since their top rules are strict. Instead, we saw that their conflict is decided indirectly, by comparing A_3 with B_2 and B_3 with A_2. The present treatment of such examples can be defended by saying that conflicts are recognised only when they are made explicit in an argument's conclusion, which seems to better respect the general nature of argumentation as providing explicit grounds for conclusions. It remains to be investigated whether this difference affects the present results on the rationality postulates (note that, although Prakken and Sartor (1997) do not assume that the strict rules are closed under transposition, this assumption can be easily added).

In one respect, Prakken and Sartor (1997) go beyond the present framework, namely, in making the preference relation on the set of defeasible inference rules defeasible and derivable within the framework. In this respect, the system is a forerunner of Modgil's (2009) extended *AF*s.

10. Conclusion

The main rhetorical aim of this paper has been to present the ASPIC framework as a general abstract framework for rule-based argumentation. In previous publications on the ASPIC framework its unifying potential was underexposed because of a focus on domain-specific inference rules instead of on general inference patterns. Here it has been argued that ASPIC, although it can be used as a specific logic at the same level of abstraction as systems such as DeLP, DL and Prakken and Sartor (1997), can also be used as an abstract framework for reasoning with general inference rules, including argument schemes. Moreover, it has been shown that by including undermining attack and generalising negation to arbitrary contrary mappings, the ASPIC framework unifies rule- and assumption-based approaches to argumentation. The latter claim has been backed by a formal proof that assumption-based argumentation (Bondarenko et al. 1997; Dung et al. 2007) is a special case of the framework and by semi-formal explanations that the same holds for Verheij's (2003) DefLog and (to a large extent) Amgoud and Cayrol's (2002) version of deductive argumentation.

In addition, the following technical contributions have been made:

- a generalisation of the ASPIC framework to arbitrary relations of contrariness between well-formed formulas;
- an extension of the ASPIC framework with preference information for resolving conflicts between arguments;

- an extension of the ASPIC framework with four types of premises and with undermining attack;
- proof that Caminada and Amgoud's (2007) rationality postulates still hold for the thus generalised and extended framework, and that they hold not only for systems closed under transposition but also for systems closed under contraposition.

The framework can be further extended and investigated in several ways. First as indicated above in Section 3.3.2, several alternative ways to define the relation between the three kinds of defeat are possible. It could be investigated to what extent such alternatives affect the present results. The same holds for the use of preferences to resolve undercutting attack (also discussed in Section 3.3.2), for the constraint that arguments have consistent premises (cf. the discussion of deductive argumentation in Section 9) and for alternative ways to define argument conflicts involving strict rules (cf. the discussion of Prakken and Sartor (1997) in Section 9).

Finally, as touched upon at the end of Section 9, an important extension of the present framework is making the preference relations that are used for resolving conflicts defeasible and derivable within the framework. This could be done along the lines of Prakken and Sartor (1997), after which it should be investigated whether Modgil's (2009) reconstruction of Prakken and Sartor (1997) as an instance of his extended argumentation frameworks can be adapted to the extended ASPIC framework.

Acknowledgements

This work was partially supported by a Distinguished Visitor grant from the Scottish Informatics and Computer Science Alliance (SICSA). I thank Chris Reed and the School of Computing, University of Dundee, Scotland, for their hospitality during the summer of 2009 and Chris Reed for encouraging me to write this paper. Floris Bex, Phan Minh Dung, Tom Gordon, Sanjay Modgil, Leon van der Torre, Bart Verheij and Gerard Vreeswijk gave useful feedback on earlier versions of this paper. Finally, I thank my former collaborators in the ASPIC project for working with me on previous versions of the ASPIC framework.

Notes

1. For reasons explained in Section 3, this paper will rename Dung's attack relations to 'defeat' relations and reserve the term 'attack' for something else.
2. In this paper, the term 'framework' will be used to denote the general model, to highlight that it can be instantiated in various ways (such instantiations will in turn be called argumentation systems). This contrasts with Dung's (1995) use of the term 'argumentation framework', which denotes a specific set of arguments with a specific attack relation. In the present paper, such specific inputs to an argumentation system will be called argumentation theories.
3. Pollock (1987, 1994) calls these 'conclusive' and '*prima facie* reasons'.
4. Modgil (2009) argued that in some contexts such extensions make sense. It seems that the formal results in Section 6 on rationality postulates also hold for undercutting defeat with preferences, but this should be formally verified.
5. Caminada and Amgoud (2007) proposed similar postulates for the intersection of extensions but since their results on these postulates directly follow from the ones for individual extensions, they will be ignored.
6. In Dung et al. (2007), the arrows are from right to left.

References

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.

Amgoud, L., and Cayrol, C. (2002), 'A Model of Reasoning Based on the Production of Acceptable Arguments', *Annals of Mathematics and Artificial Intelligence*, 34, 197–216.
Baroni, P., and Giacomin, M. (2009), 'Semantics of Abstract Argument Systems', in *Argumentation in Artificial Intelligence*, eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 25–44.
Bench-Capon, T. (2003), 'Persuasion in Practical Argument Using Value-based Argumentation Frameworks', *Journal of Logic and Computation*, 13, 429–448.
Besnard, P., and Hunter, A. (2008), *Elements of Argumentation*, Cambridge, MA: MIT Press.
Bex, F., Prakken, H., Reed, C., and Walton, D. (2003), 'Towards a Formal Account of Reasoning about Evidence: Argumentation Schemes and Generalisations', *Artificial Intelligence and Law*, 12, 125–165.
Bondarenko, A., Dung, P., Kowalski, R., and Toni, F. (1997), 'An Abstract, Argumentation-theoretic Approach to Default Reasoning', *Artificial Intelligence*, 93, 63–101.
Caminada, M. (2006), 'On the Issue of Reinstatement in Argumentation', in *Proceedings of the 11th European Conference on Logics in Artificial Intelligence (JELIA 2006)*, no. 4160 in Springer Lecture Notes in AI, Berlin: Springer Verlag, pp. 111–123.
Caminada, M., and Amgoud, L. (2007), 'On the Evaluation of Argumentation Formalisms', *Artificial Intelligence*, 171, 286–310.
Dung, P. (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., Kowalski, R., and Toni, F. (2006), 'Dialectic Proof Procedures for Assumption-based, Admissible Argumentation', *Artificial Intelligence*, 170, 114–159.
Dung, P., Mancarella, P., and Toni, F. (2007), 'Computing Ideal Sceptical Argumentation', *Artificial Intelligence*, 171, 642–674.
Elvang-Ġoransson, M., Fox, J., and Krause, P. (1993). 'Acceptability of Arguments as Logical Uncertainty', in *Proceedings of the 2nd European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU93)*, Berlin: Springer Verlag, pp. 85–90.
Garcia, A., and Simari, G. (2004), 'Defeasible Logic Programming: An Argumentative Approach', *Theory and Practice of Logic Programming*, 4, 95–138.
Gordon, T., Prakken, H., and Walton, D. (2007), 'The Carneades Model of Argument and Burden of Proof', *Artificial Intelligence*, 171, 875–896.
Governatori, G., Maher, M., Antoniou, G., and Billington, D. (2004), 'Argumentation Semantics for Defeasible Logic', *Journal of Logic and Computation*, 14, 675–702.
Jakobovits, H., and Vermeir, D. (1999), 'Robust Semantics for Argumentation Frameworks', *Journal of Logic and Computation*, 9, 215–261.
Lin, F., and Shoham, Y. (1989), 'Argument Systems. A Uniform Basis for Nonmonotonic Reasoning', in *Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference*, San Mateo, CA: Morgan Kaufmann Publishers, pp. 245–255.
Loui, R. (1987), 'Defeat among Arguments: A System of Defeasible Inference', *Computational Intelligence*, 2, 100–106.
Modgil, S. (2009), 'Reasoning about Preferences in Argumentation Frameworks', *Artificial Intelligence*, 173, 901–934.
Nute, D. (1994), 'Defeasible Logic', in *Handbook of Logic in Artificial Intelligence and Logic Programming*, eds. D. Gabbay, C.J. Hogger and Robinson, Oxford: Clarendon Press, pp. 253–395.
Pollock, J. (1974), *Knowledge and Justification*, Princeton: Princeton University Press.
Pollock, J. (1987), 'Defeasible Reasoning', *Cognitive Science*, 11, 481–518.
Pollock, J. (1994), 'Justification and Defeat', *Artificial Intelligence*, 67, 377–408.
Prakken, H. (1997), *Logical Tools for Modelling Legal Argument. A Study of Defeasible Argumentation in Law*, Law and Philosophy Library, Dordrecht/Boston/London: Kluwer Academic Publishers.
Prakken, H., and Sartor, G. (1997), 'Argument-based Extended Logic Programming with Defeasible Priorities', *Journal of Applied Non-classical Logics*, 7, 25–75.
Prakken, H., and Vreeswijk, G. (2002), 'Logics for Defeasible Argumentation', in *Handbook of Philosophical Logic* (Vol. 4, 2nd ed.), eds. D. Gabbay and F. Ġunthner, Dordrecht/Boston/London: Kluwer Academic Publishers, pp. 219–318.
Reiter, R. (1980), 'A Logic for Default Reasoning', *Artificial Intelligence*, 13, 81–132.
Verheij, B. (1996), 'Two Approaches to Dialectical Argumentation: Admissible Sets and Argumentation Stages', in *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC-96)*, Utrecht: The Netherlands, pp. 357–368.
Verheij, B. (2003a), 'DefLog: On the Logical Interpretation of Prima Facie Justified Assumptions', *Journal of Logic and Computation*, 13, 319–346.

Verheij, B. (2003b), 'Dialectical Argumentation with Argumentation Schemes: An Approach to Legal Logic', *Artificial Intelligence and Law*, 11, 167–195.
Vreeswijk, G. (1993), 'Studies in Defeasible Argumentation', doctoral dissertation, Vrije University Amsterdam.
Vreeswijk, G. (1997), 'Abstract Argumentation Systems', *Artificial Intelligence*, 90, 225–279.
Walton, D., Reed, C., and Macagno, F. (2008), *Argumentation Schemes*, Cambridge: Cambridge University Press.

Appendix: Proofs

PROPOSITION 5.8 *Consider any argumentation theory with $\mathcal{L}$ closed under classical negation and $^-$ defined accordingly. Then if $\mathcal{R}_s$ consists of all valid propositional inferences then $\mathcal{R}_s$ is closed under contraposition and transposition.*

Proof Note first that if $\mathcal{R}_s$ consists of all valid propositional inferences, then $\vdash$ satisfies the deduction theorem, i.e. it satisfies

$$\{p_1, \ldots, p_n\} \vdash q \;\Leftrightarrow\; \vdash (p_1 \wedge \cdots p_n) \supset q$$

Now consider any rule $p_1, \ldots, p_n \rightarrow q$. Then $\{p_1, \cdots, p_n\} \vdash q$ so by the deduction theorem $\vdash (p_1 \wedge \ldots p_n) \supset q$. Then also (by propositional reasoning) $\vdash (\neg q \wedge p_2 \wedge \ldots p_n) \supset \neg p_1$. But then by the deduction theorem $\{\neg q, p_2, \ldots, p_n\} \vdash \neg p_1$ so since $\mathcal{R}_s$ contains all valid propositional inferences, $\mathcal{R}_s$ contains $\neg q, p_2, \ldots, p_n \rightarrow \neg p_1$. ■

PROPOSITION 6.1 Let $<\mathcal{A}, Def>$ *be an argumentation framework as defined in Definition 3.22 and E any of its extensions under a given semantics subsumed by complete semantics. Then for all $A \in E$*: if $A' \in \text{Sub}(A)$, *then $A' \in E$.*

Proof The proof is a trivial adaptation of the proof of Proposition 1 of Caminada and Amgoud (2007), taking the possibility of undermining defeat into account. ■

PROPOSITION 6.2 Let $<\mathcal{A}, Def>$ *be an argumentation framework corresponding to an argumentation theory, and E any of its extensions under a given semantics subsumed by complete semantics. Then* $\{\text{Conc}(A) | A \in E\} = Cl_{\mathcal{R}s}(\{\text{Conc}(A) | A \in E\})$.

Proof Caminada and Amgoud's proof of their Proposition 8 depends on Proposition 6.1, which also holds for the present framework, and makes no assumptions on the use of priorities. Therefore, the proof also holds for the present version. ■

THEOREM 6.9 *Let $< \mathcal{A}, Def >$ be an argumentation framework corresponding to a well-formed argumentation theory that is closed under contraposition or transposition and has a reasonable argument ordering and a consistent $Cl_{\mathcal{R}s}(\mathcal{K}_n)$, and let E be any of its extensions under a given semantics subsumed by complete semantics. Then the set* $\{\text{Conc}(A) \mid A \in E\}$ *is consistent.*

Proof Let E be a complete extension. Suppose that $\{\text{Conc}(A) \mid A \in E\}$ is inconsistent. This means that $\exists A, B \in E, \text{Conc}(A) = \overline{\text{Conc}(B)}$. Since E is a complete extension, E is conflict-free. This means that A does not defeat B and B does not defeat A. It will be shown that this leads to a contradiction.

First the following lemmas are proved.

LEMMA A1 *Let A be an argument and B a plausible or defeasible argument in an argumentation theory that is closed under contraposition or transposition such that* Conc(A) *and* Conc(B) *are contradictories. Then A can be extended to an argument* A^+ *that rebuts or undermines B.*

Proof Consider first systems closed under contraposition. By Corollary 6.6, it holds that $\text{Conc}(M(B)) \vdash \text{Conc}(B)$ so with contraposition (which is assumed to hold) and since Conc(A) and Conc(B) contradict each other we have for any $B_i \in M(B)$ that $\text{Conc}(M(B)\backslash\{B_i\}) \cup \text{Conc}(A) \vdash -\text{Conc}(B_i)$. Then clearly $M(B)\backslash\{B_i\}$ and $M(A)$ are the maximal fallible subarguments of an argument A^+ for $-\text{Conc}(B_i)$. Since by construction of $M(B)$ either B_i is a non-axiom premise or ends with a defeasible inference, A^+ either undermines or rebuts B_i. But then A also undermines or rebuts B.

For systems closed under transposition, the existence of arguments A^+ and B_i is proved by a straightforward generalisation of Lemma 6 of Caminada and Amgoud (2007). Then the proof can be completed as above. ■

COROLLARY A2 *If the argumentation theory has a reasonable argument ordering then if* $B \prec A$, *then* A^+ *defeats B.*

Proof (continuing the proof of Lemma A1) Since $\preceq$ is reasonable, there exist such a B_i and A^+ such that $A^+ \not\prec B_i$. Then A^+ defeats B_i so A^+ defeats B. ■

Now for proving Theorem 6.9, the following cases must be distinguished.

(1) $A \in \mathcal{K}_i$. Then A is not in any extension.
(2) A is an assumption. If A is a contradictory of $Conc(B)$, then B defeats A. If instead A is a contrary of $Conc(B)$, then since the argumentation theory is well formed B is also an assumption so A defeats B. Contradiction.
(3) A is firm and strict. If B is also firm and strict, then $Cl_{Rs}(K_n)$ is inconsistent, which contradicts the assumption that it is consistent. If B is plausible or defeasible, then A defeats B by condition (1) of Definition refpreceq. Contradiction.
(4) A is plausible or defeasible. If B is firm and strict then this is case (3). If B's top rule is defeasible and Conc(A) is a contrary of Conc(B), then A defeats B, while if Conc(A) and Conc(B) contradict each other, either A defeats B or B defeats A. If B's top rule is strict, then by the assumption that the argumentation theory is well formed, Conc(A) and Conc(B) contradict each other. If $B \not\prec A$ then B defeats A while otherwise A can by Lemma A1 and Corollary A2 be extended to an argument A^+ that defeats B. It is then left to prove that $A^+ \in E$. Any defeater C of A^+ will by construction of A^+ do so by defeating an element of $M(A)$ or $M(B)$ (since all inferences that are not in $M(A)$ or $M(B)$ are strict and there are no new premises). However, this defeated element is in E by Proposition 6.1, so since E is conflict-free, $C \notin E$. But then $A^+ \in E$, which contradicts the fact that E is conflict-free.

■

THEOREM 6.10 *Let* $<\mathcal{A}, Def>$ *be an argumentation framework corresponding to a well-formed argumentation theory that is closed under contraposition or transposition and has a reasonable argument ordering and a consistent* $Cl_{\mathcal{R}s}(\mathcal{K}_n)$, *and let E be any of its extensions under a given semantics subsumed by complete semantics. Then the set* $Cl_{\mathcal{R}s}(\{\text{Conc}(A) \mid A \in \mathcal{E}\})$ *is consistent.*

Proof As in Caminada and Amgoud (2007). ■

COROLLARY 6.11 *If the conditions of Theorem 6.10 are satisfied, then for any extension E under a given semantics subsumed by complete semantics the set* $\{\varphi | \varphi$ *is a premise of an argument in* $E\}$ *is consistent.*

Proof Let A be any argument in E and φ any premise of A. By definition of an argument, φ is a subargument of A so by Proposition 6.1 we have that $\varphi \in E$. Then the corollary follows from Theorem 6.10 and the fact that subsets of consistent sets are consistent. ■

PROPOSITION 6.15 *The last-link argument ordering is reasonable.*

Proof

LEMMA A3 *Consider any ordering* $\preceq_s$ *on sets ordered by a partial preorder* $\leq_e$ *such that* $S_1 \prec_s S_2$ *iff there exists an* $e_1 \in S_1$ *such that for all* $e_2 \in S_2$ *it holds that* $e_1 <_e e_2$. *Then if* $S_1 \prec_s S_2$ *and* e_1 *is a non-smallest element of* S_1 *(w.r.t.* $\leq_e$*), then* $S_2 \cup \{e_1\} \not\prec_s S_1$.

Proof Straightforward. ■

Now by Corollary 6.13 that $B \prec A$ means that there exists a $B_i \in M(B)$ with top rule b such that for all $A' \in M(A)$ with top rule a it holds that $b < a$. Choose such a B_i with minimal b (w.r.t. $\leq_e$) to form A^+ as in the proof of Corollary A2. Then by Lemma A3 $\text{LastDefRules}(A^+) \not\prec_s \text{LastDefRules}(B_i)$. But then $A^+ \not\prec B_i$. ■

PROPOSITION 6.18 *The weakest-link argument ordering is reasonable.*

Proof That $B \prec A$ now means that $\text{Prem}(B) \prec_s \text{Prem}(A)$ and $\text{DefRules}(B) \prec_s \text{DefRules}(A)$.

If $\text{DefRules}(B) \neq \emptyset$, then there exists a $B_i \in \text{DefRules}(B)$ with top rule b such that for all $A' \in \text{DefRules}(A)$ with top rule a it holds that $b < a$. Choose such a B_i with minimal b (w.r.t. $\leq$) in the construction of A^+ and B_i in the proof of Corollary A2. Then since all new defeasible rules of the corresponding A^+ are from elements of $M(B)$, by Lemma A3 $\text{DefRules}(A^+) \not\prec_s \text{DefRules}(B)$. But then $A^+ \not\prec B_i$.

If $\text{DefRules}(B) = \emptyset$, then $\text{DefRules}(A) = \emptyset$. Since $\text{Prem}(B) \prec_s \text{Prem}(A)$ there exists a premise p in $\text{Prem}(B)$ such that for all premises p' in $\text{Prem}(A)$ it holds that $p' < p$. Then in the construction of A^+ and B_i in the proof of Corollary A2, choose B_i to be an argument containing a minimal such p. Then since all new premises of the corresponding A^+ are from $\text{Prem}(B)$, by Lemma A3 $\text{Prem}(A^+) \not\prec_s \text{Prem}(B)$. But then $A^+ \not\prec B_i$. ■

PROPOSITION 8.7 For all ABF such that $AT = AT_{ABF}$ does not allow arguments with an infinite number of subarguments, there exists an argument $A \vdash_{ABF} \alpha$ if and only if there exists an argument $A \vdash_{AT} \alpha$.

Proof $\Rightarrow$ For the only-if part, let $S_1, \ldots, S_n$ be a backward deduction of α. It will be shown by induction on the structure of backward deductions that there exists an AT-argument with conclusion α and premises S_n.

Note first that since all elements of S_n are in $\mathcal{A}$ so in $\mathcal{K}_a$, by clause (1) of Definition 3.6 they are all an AT-argument and their premises are all in S_n.

Consider next any set S_i such that all elements of S_{i+1} are the conclusion of an AT-argument with premises from S_n. Then for any element α_i of S_i, if α_i is also in S_{i+1}, then trivially α_i is the conclusion of an AT-argument with premises in S_n, otherwise for some set $S = \{\beta_1, \ldots, \beta_m\} \subseteq S_{i+1}$ there exists a rule $\beta_1, \ldots, \beta_m \rightarrow \alpha_i$ in $\mathcal{R}_{ABF}$. But then this rule is also in $\mathcal{R}_s$. Let, furthermore, the AT-arguments for $\beta_1, \ldots, \beta_m$ (which exist by the induction hypothesis) be $B_1, \ldots, B_m$: then by clause (2) of Definition 3.6, $B_1, \ldots, B_m \rightarrow \alpha_i$ is an AT-argument for α_i with all its premises in S_n.

Next it is proved that for any S_i the union of all premises of all AT-arguments for elements in S_i is S_n. Note that for any pair S_i, S_{i+1}, the set S_{i+1} is formed by replacing at most one element σ in S_i with a set S in S_{i+1}. As just proved, there exists an AT-argument $B_1, \ldots, B_m \rightarrow \alpha_i$, where $B_1, \ldots, B_m$ are the AT-arguments for all elements in S. By clause (2) of Definition 3.6, the premises of this argument are the union of the premises of the arguments $B_1, \ldots, B_n$. But then no premises have been added or deleted by creating S_{i+1} from S_i. Note finally, that the union of the premises of all AT-arguments for any element in S_n (which are these elements themselves) trivially equals S_n. But then this set equals S_n for all S_i.

$\Leftarrow$ For the if-part, suppose $P \vdash_{AT} \alpha$. A backward deduction with multi-sets $S_1, \ldots, S_n$ such that $S_1 = \{\alpha\}$ and $S_n = P$ can be created as a maximal sequence such that:

(1) $S_1 = \{\alpha\}$,
(2) For all $S_i (i \geq 1)$: create S_{i+1} by selecting one element σ from S_i not selected before and:
(a) if $\sigma \in P$ then $S_{i+1} = S_i$; otherwise
(b) $S_{i+1} = S_i - \{\sigma\} \cup S$ for some $S = \{\text{Conc}(B_1), \ldots, \text{Conc}(B_n)\}$ such that there exists an argument $B \in \text{Sub}(A)$ of the form $B_1, \ldots, B_n \rightarrow \sigma$.

It is now proved that for any S_i and any $\sigma \in S_i$ one of these two conditions is satisfied, i.e. either $\sigma \in P$ or σ is the conclusion of an argument in $\text{Sub}(A)$. The proof is with induction on the structure of $S_1, \ldots, S_n$. Consider first $S_1 = \{\alpha\}$. Then if $A = \alpha \in \mathcal{K}_a$, then trivially $\alpha \in P$, otherwise $A = A_1, \ldots, A_n \rightarrow \alpha$ so trivially $A \in \text{sub}(A)$. Consider next any S_i such that all its elements satisfy conditions (2)a and (2)b. Then if $S_{i+1} = S_i$ this trivially also holds for S_{i+1}, otherwise if S replaces σ in S_{i+1} then by the induction hypothesis this is since there exists a subargument $B \in \text{Sub}(A)$ of the form $B_1, \ldots, B_n \rightarrow \sigma$ such that $S = \{\text{Conc}(B_1), \ldots, \text{Conc}(B_n)\}$. Then clearly for any new element $\text{Conc}(B_i) \in S$, there exists a subargument for it in $\text{Sub}(A)$, namely, B_i.

Next, since all steps in the sequence apply an inference rule from $\mathcal{R}_s$, which by Definition 8.6 is also in $\mathcal{R}_{ABF}$, the sequence clearly is a backward deduction.

Finally, it is proved that the sequence ends with $S_n = P$. Let $\text{Sub}^*(A)$ be the multi-set consisting of, for all $A' \in \text{Sub}(A)$, as many occurrences as there are inferences in A that use A'. Note that by the assumption that $\text{Sub}(A)$ is finite, $\text{Sub}^*(A)$ is also finite. Then let for any S_i the set $\text{UnusedSub}(S_i)$ be the subset of all arguments in $\text{Sub}^*(A)$ that were not used to create S_i from S_1. (So $\text{UnusedSub}(S_1) = \text{Sub}^*(A)$ and, e.g. $\text{UnusedSub}(S_2) = \text{Sub}^*(A) - \{A\}$). Then note that by any application of condition (2)b this multi-set loses one element. Then since $S_1, \ldots, S_n$ is a maximal sequence of elements satisfying conditions (1) and (2), we have that $\text{UnusedSub}(S_n) = \emptyset$. Then since $P \subseteq \text{Sub}^*(A)$, we have that $P \subseteq S_n$. Assume next for contradiction that there is an element $\sigma \in S_n$ which is not in P: then, as proved above, σ can be replaced by a set S such that $S \rightarrow \sigma$ is an inference in A, so $S_1, \ldots, S_n$ is not maximal. Contradiction, so $S_n = P$. ■

PROPOSITION 8.8 *For all ABF such that $AT = AT_{ABF}$ does not allow arguments with an infinite number of subarguments it holds for every argument $A \vdash_{ABF} \alpha$ and every argument $A \vdash_{AT} \alpha$ that*

$B \vdash_{AT} \beta$ is defeated by an argument $B \vdash_{ABF} \beta$ if and only if $A \vdash_{AT} \alpha$ is defeated by an argument $B \vdash_{AT} \beta$.

Proof $\Rightarrow$ Assume $A \vdash_{ABF} \alpha$ and $B \vdash_{ABF} \beta$ defeats $A \vdash_{ABF} \alpha$. Then according to the contrariness mapping in *ABF* we have that $\beta = \overline{p}$ for some $p \in A$. Furthermore, by Proposition 8.7, there exists an $A \vdash_{AT} \alpha$ and an argument $B \vdash_{AT} \beta$. Then by identity of the contrariness mappings we also have that $\beta = \overline{p}$ for some $p \in A$ according to AT. Then since $p \in \mathcal{K}_a$, clearly $B \vdash_{AT} \beta$ defeats $A \vdash_{AT} \alpha$.

$\Leftarrow$ Assume $A \vdash_{AT} \alpha$ and $B \vdash_{AT} \beta$ defeats $A \vdash_{AT} \alpha$. Then since all arguments in *AT* are strict, B undermines A, and according to the contrariness mapping in AT, we have that $\beta = \overline{p}$ for some $p \in A$. Furthermore, by Proposition 8.7, there exists an $A \vdash_{ABF} \alpha$ and an argument $B \vdash_{ABF} \beta$. Then by identity of the contrariness mappings, we also have that $\beta = \overline{p}$ for some $p \in A$ according to *ABF*. Then since $p \in A$, clearly $B \vdash_{ABF} \beta$ defeats $A \vdash_{ABF} \alpha$. ■

THEOREM 8.9 *For all ABF, any semantics S subsumed by complete semantics and any set E*:

(1) *if E is an S-extension of ABF then E_{AT} is an S-extension of AT, where* $E_{AT} = \{A \vdash_{AT} \alpha | A \vdash_{ABF} \sigma \in E\}$;
(2) *if E is an S-extension of AT then E_{ABF} is an S-extension of ABF, where* $E_{ABF} = \{A \vdash_{ABF} \alpha | A \vdash_{AT} \alpha \in E\}$.

Proof As before, the proof for complete semantics suffices.

(1) Consider any complete extension E of *ABF*. It is first proven that any member of E_{AT} is defended by E_{AT}. Since E is conflict-free, by construction of E_{AT} and Proposition 8.8 also E_{AT} is conflict-free. Consider next any $A \vdash_{AT} \alpha \in E_{AT}$ defeated by some $B \vdash_{AT} \beta$. By construction of E_{AT}, there exists an $A \vdash_{ABF} \alpha \in E$. Then by Propositions 8.8 and 8.8 there exists a $B \vdash_{ABF} \beta$ defeating $A \vdash_{ABF} \alpha$. But since E is a complete extension, $B \vdash_{ABF} \beta$ is in turn defeated by some $C \vdash_{ABF} \gamma \in E$. Then by construction of E_{AT} and Proposition 8.8, also $C \vdash_{AT} \gamma \in E_{AT}$ and by Proposition 8.7, $C \vdash_{AT} \gamma$ defeats $B \vdash_{AT} \beta$. So $A \vdash_{AT} \alpha$ is defended by E_{AT}.
Next, to prove that any argument defended by E_{AT} is a member of E_{AT}, assume $A \vdash_{AT} \alpha$ is defended by E_{AT}. Then any of its defeaters $B \vdash_{AT} \beta$ is in turn defeated by an element $C \vdash_{AT} \gamma \in E_{AT}$. But then by Proposition 8.7, the same holds for their corresponding *ABF*-arguments, which exist by Proposition 8.7. Moreover, by construction of E_{AT} we have that $C \vdash_{ABF} \gamma \in E$ so, since E is a complete extension, also $A \vdash_{ABF} \alpha \in E$. But then $A \vdash_{AT} \alpha \in E_{AT}$ by construction of E_{AT} and Proposition 8.8
(2) The proof of (2) is entirely similar and therefore omitted.

■

COROLLARY 8.10 *For any ABF, any semantics S subsumed by complete semantics, and for any formula φ it holds that φ is skeptically (credulously) S-acceptable in* ABF *if and only if φ is skeptically (credulously) S-acceptable in AT_{ABF}.*

Proof Straightforward. ■