UTRECHT UNIVERSITY

Paraconsistent Logics in Argumentation Systems

by

Diana Grooters
ICA-3470857

Supervised by Henry Prakken

A thesis submitted in partial fulfilment for the degree of Master of Science

in the
Faculty of Science
Department of Information and Computing Sciences

July 2014

"How do you reason from inconsistent sets?"

"You don't, since every formula follows in that case, you reason from consistent subsets."

Payette and Schotch (2007)

UTRECHT UNIVERSITY

Abstract

Faculty of Science
Department of Information and Computing Sciences

Master of Science

by Diana Grooters

Supervised by Henry Prakken

This thesis is about argumentation theory and the problems that can arise when two contradicting conclusions are obtained during an argument. These conclusions can be taken as premises for $\perp$ from which everything can be derived (in non-paraconsistent logics). This is a problem since for every sentence φ, an argument can be constructed which takes $\perp$ as a premise and concludes with φ.

The instantiation is examined of the $ASPIC^+$ framework with paraconsistent logics for the strict inference rules in order to prevent the system from becoming trivial in case of an inconsistency. For the paraconsistent Da Costa's C_ω system, the Logic of Paradox, the logic W and the relevant logic **R**, it is examined whether the closure and consistency postulates are satisfied. These postulates impose relevant requirements on any extension of an argumentation framework. It is shown that the first two logics do not satisfy these postulates. The relevant logic **R** is also not applicable since this logic is non monotonic. However, the rationality postulates are satisfied for the logic W. For this, the $ASPIC^\star$ framework is introduced.

Furthermore, the generation of arguments of which the strict rules are non-minimal can be seen as an efficiency problem. It is investigated whether these 'unnecessary' subarguments can be removed without affecting the conclusions of the extensions.

Like for the $ASPIC^+$ framework, the obtained extensions, when conflict-free is defined in terms of defeat or attack free, are compared for minimal $ASPIC^\star$ frameworks and minimal $ASPIC^+$ frameworks. It turns out that these extensions are the same for both definitions.

Acknowledgements

I wish to thank various people for their contribution to this thesis. First of all, I would like to thank my thesis supervisor Henry Prakken for his valuable and useful critiques. He also helped me with finding the right direction while letting me free in doing my research. I wish to acknowledge the help provided by Graham Priest, Edwin Mares and Gillman Payette. My special thanks are extended to Rick Barneveld and my family for giving me support during my thesis research.

Contents

List of Figures

Abbreviations

ABA	**A**ssumption-**B**ased **A**rgumentation
AF	**A**bstract argumentation **F**ramework
AFSAF	**A**bstract argumentation **F**ramework corresponding to a **S**tructured **A**rgumentation **F**ramework
AL	**A**bstract **L**ogic
AS	**A**rgumentation **S**ystem
AT	**A**rgumentation **T**heory
KB	**K**nowledge **B**ase
LAF	**L**ogic-associated **A**rgumentation **F**ramework
LP	**L**ogic of **P**aradox
PBLAF	**P**reference **B**ased **L**ogic-associated **A**rgumentation **F**ramework
SAF	**S**tructured **A**rgumentation **F**ramework

Symbols

$\mathcal{A}$	Arguments in an argumentation framework
att	Attack relation in an argumentation framework
$\mathcal{D}$	Defeat relation in an argumentation framework
$\mathcal{K}$	Knowledge base in an argumentation framework
$\mathcal{K}_n$	Necessary premises in an argumentation framework
$\mathcal{K}_p$	Ordinary premises in an argumentation framework
$\mathcal{L}$	Logical language closed under negation
NR	Non Redundant arguments
$\mathcal{R}$	Inference rules in an argumentation framework
$\mathcal{R}_d$	Defeasible inference rules in an argumentation framework
$\mathcal{R}_s$	Strict inference rules in an argumentation framework

Functions

Base($\star$)	Base($\star$) function	$\mathcal{A} \to 2^{\mathcal{A}}$
$BD(\star)$	Basic defeasible arguments($\star$) function	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$
$Cl_{\mathcal{R}_s}(\star)$	Closure function	$2^{\mathcal{L}} \to 2^{\mathcal{L}}$
CN	Consequence operator	$2^{\mathcal{L}} \to 2^{\mathcal{L}}$
Cnl	Conclusion function	$\mathcal{A} \to \mathcal{L}$
Conc	Conclusion function	$2^{\mathcal{A}} \to 2^{\mathcal{L}}$
DefRules	Defeasible rules function	$2^{\mathcal{A}} \to 2^{\mathcal{R}_d}$
F	Acceptability function	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$
GN	Generated arguments function	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$
LastDefRules	Last defeasible rules function	$2^{\mathcal{A}} \to 2^{\mathcal{R}_d}$
$M(\star)$	Maximal fallible subarguments($\star$) function	$\mathcal{A} \to 2^{\mathcal{A}}$
n	Partial function	$\mathcal{R}_d \to \mathcal{L}$
$NP(\star)$	Necessary premise function	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$
Prem	Premise function	$2^{\mathcal{A}} \to 2^{\mathcal{K}}$
Sub	Subargument function	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$
TopRule	Top rule function	$2^{\mathcal{A}} \to \mathcal{R}$
$-$	Minimal argument($\star$)	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$
$+$	Extended argument($\star$)	$\mathcal{A} \to \mathcal{A}$
#	Basic defeasible arguments($\star$) together with all necessary premises	$2^{\mathcal{A}} \to 2^{\mathcal{A}}$

Chapter 1

Introduction

Over the years argumentation has become more and more important within artificial intelligence (Bench-Capon and Dunne (2007)). Argumentation theory studies how conclusions can be drawn from a set of premises by logical reasoning and whether these arguments are acceptable or justified. The main purpose of argumentation is to resolve a conflict between opinions by using arguments of both sides.

The conflicts between arguments can be used to describe an attack relation between the arguments. Dung (1995) introduced a theory of abstract argumentation frameworks, which only takes arguments with a binary attack relation between them as input. These frameworks provide intuitive semantics for the central notion of acceptability of arguments. The abstract argumentation frameworks of Dung (1995) can be instantiated with a wide range of logical formalisms. One is free to choose a logical language $\mathcal{L}$ and define what constitutes an argument and attack between arguments.

Since it is not specified how these attack relations are obtained, the framework of Dung (1995) is by itself usually too abstract. Instead, an instantiated version of his approach should be used such as the $ASPIC$ framework which was developed by Amgoud et al. (2006); Caminada and Amgoud (2007). The $ASPIC$ framework is a characterisation of a set of tree-structured arguments ordered with a binary defeat relation and preferences. The arguments are built from a knowledge base by using strict and defeasible rules. The approach of Dung (1995) can be instantiated with the $ASPIC$ framework, so each of the semantics of Dung (1995) can be used to compute the acceptability status of the structured arguments. Caminada and Amgoud (2007) defined rationality postulates for the sound definition of an argumentation system and will avoid anomalous results. They showed that $ASPIC$, without taking preferences of arguments into account, satisfies the rationality postulates under the condition that the argumentation theory is closed under transposition or contraposition.

In Modgil and Prakken (2013), this $ASPIC$ framework is modified by generalising $ASPIC$ to accommodate a broader range of instantiations, which is called the $ASPIC^+$ framework. Sets of sufficient conditions were identified for satisfying the postulates when preferences are taken into account. Modgil and Prakken (2013) further developed the $ASPIC^+$ framework. The modified version of $ASPIC^+$ accommodates the possibility for instantiating the framework with a range of concrete logics, to the extent that one can prove satisfaction of the rationality postulates of Caminada and Amgoud (2007).

Pollock (1994) described the problems arising from self-defeating arguments, for example when the contradictory conclusions of two or more arguments can be taken as the premises for $\perp$. In classical logic, everything can be derived from $\perp$. This means that there is an argument for every statement when for example the $ASPIC^+$ framework is instantiated with classical logic. It follows, that depending on the preferences between arguments, there might be unrelated arguments which are defeated by an argument which has $\perp$ as a premise. This is also known as the problem of self-defeat or the trivialisation problem. This problem will arise for every logic for which $\{A, \neg A\} \vdash B$ holds.

Wu and Podlaszewski (2014) have proven that if the set of conclusions of all subarguments of an argument is required to be indirectly consistent, this problem does not arise, while all results on the rationality postulates still hold. This holds for the special case where a propositional or first-order language with a classical interpretation is used. However, it does not hold in the general case where preferences for arguments are taken into account.

This thesis tries to find a solution for the problem of self-defeat for argumentation frameworks in which preferences are taken into account. The problem lies in the explosion in case of an inconsistency: $\{A, \neg A\} \vdash B$. Paraconsistent logics attempt to deal with inconsistencies without having an explosion. In non paraconsistent logics, there is only one inconsistent theory: the trivial theory that has every sentence as a theorem. Paraconsistent logics make it possible to distinguish between inconsistent theories and to reason with them. The primary goal is to find out whether instantiating the $ASPIC^+$ framework with paraconsistent logics will prevent the trivialisation of a system in case of a single inconsistency while still satisfying the rationality postulates. This research focuses on the instantiation of the 'Logic of Paradox' (Priest (1989)), the relevance logic **R** (Mares (2004)), 'Da Costa's C_ω system' (Sylvan (1990)) and the logic W (Rescher and Manor (1970)). For the latter, the $ASPIC^\star$ framework is introduced.

Another research question that will be answered is whether it makes a difference for extensions if arguments are required to be *minimal*. This idea is based on the intuition that the addition of unnecessary subarguments should not make a difference. The reason for being in favour of only allowing minimal arguments, is that the number of arguments is reduced, which results in a more efficient system. It is investigated whether the conclusions that can be drawn from an argumentation framework are not affected in case arguments are required to be minimal.

For the $ASPIC^+$ framework, Modgil and Prakken (2013) proved that under certain circumstances the conclusions of the extensions are the same when conflict-free extensions are defined in terms of defeat or attack free. This thesis checks the same for minimal $ASPIC^\star$ and minimal $ASPIC^+$ frameworks.

This thesis is structured as follows. In Chapter 2, abstract argumentation frameworks and the $ASPIC^+$ frameworks are briefly described. After these frameworks are presented, the trivialisation and efficiency problem are further clarified with an example in Chapter 3. Then the four paraconsistent logics are discussed in Chapter 4, which are further examined for their applicability in the $ASPIC^+$ framework. In Chapter 5, the rationality postulates of Caminada and Amgoud (2007) are described which impose requirements on any extension of an argumentation framework. Then some results of Modgil and Prakken (2013) and Dung and Thang (2014) are used to test whether these rationality postulates still hold for $ASPIC^+$ frameworks, instantiated with one of these paraconsistent logics. The results of Dung and Thang (2014) are also generalised on some points. Chapter 6 is focused on the question whether non-minimal arguments can be removed without getting different conclusions. In Chapter 7, it is checked whether conclusions of the extensions of minimal $ASPIC^\star$ and minimal $ASPIC^+$ frameworks are the same when conflict-free is defined in terms of defeat or attack free.

Chapter 2

Argumentation Frameworks

Roughly, the idea of argumentative reasoning is that a statement is acceptable if it can be argued successfully against attacking arguments. To understand the acceptability of arguments, Dung (1995) provided a theory of abstract argumentation frameworks. After having described this theory, the $ASPIC^+$ framework is described. The $ASPIC^+$ framework adopts an intermediate level of abstraction between the fully abstract level of Dung (1995) and concrete instantiating logics. In the end of this chapter, it is discussed how these frameworks can be linked.

2.1 Dung's Abstract Argumentation Framework

Participants of a debate give arguments to convince the opponents of their opinion. These arguments can be conflicting; the *abstract argumentation framework* of Dung (1995) represents the attack relation between arguments. The basic principles of the abstract argumentation framework of Dung (1995) are explained here concisely.

Definition 2.1.1 (Abstract argumentation framework, Dung (1995))**.** An *abstract argumentation framework* (AF) is determined by a pair $(\mathcal{A}, att)$, where $\mathcal{A}$ is a set of arguments and att is a binary relation over $\mathcal{A}$ which represents the attacks between arguments.

The semantics of argumentation is determined by the acceptability of arguments. A set of arguments S attacks an argument A, if some argument of S attacks A. S is said to be *conflict-free* if S does not attack any argument within S, otherwise S is *conflicting*. An argument A is *acceptable* with respect to the set of arguments S, if S attacks each argument attacking A. S is said to be *admissible* if S is conflict-free and it attacks each argument attacking some argument in S.

Extensions are sets of arguments which are considered to be conform to an *acceptance attitude.* For example, a sceptical reasoner will not accept an argument easily, while a credulous reasoner does. The characteristic function F assigns the set of arguments that are acceptable with respect to a certain set. F is monotonic, so it has a smallest set S' such that $F(S') = S'$. Such a set S' is called the least fixed point of F. A *complete extension* is defined as a conflict-free fixed point of the function F. The *grounded extension* is the least fixed point of F. Therefore each grounded extension is also a complete extension. A maximal conflict-free fixed point of F is a *preferred extension.* While there can be more preferred extensions, there exists only one unique grounded extension. Finally, a *stable extension* is a conflict-free set of arguments S that attacks all arguments not belonging to S. Note that each stable extension is a complete and preferred extension, but not the other way around.

2.2 $ASPIC^+$ Framework

The $ASPIC^+$ framework is described according to the definitions given by Modgil and Prakken (2013). In this thesis, only logical languages are used that have a symmetrical negation operator.

An argumentation system together with a knowledge base are used to construct arguments in $ASPIC^+$.

Definition 2.2.1 (Argumentation system, Modgil and Prakken (2013))**.** An *argumentation system* is a quadruple $AS = (\mathcal{L}, \mathcal{R}, n, \preceq)$, where:

- $\mathcal{L}$ is a logical language closed under negation ($\neg$).
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\phi_1, \ldots, \phi_n \rightarrow \phi$ and $\phi_1, \ldots, \phi_n \Rightarrow \phi$ respectively (where ϕ_i, ϕ for $i \in \{1, \ldots, n\}$ are meta-variables ranging over wff in $\mathcal{L}$), and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- n is a partial function such that $n : \mathcal{R}_d \longrightarrow \mathcal{L}$. ($n(r)$ informally means that the defeasible rule r is applicable.)
- $\preceq$ is a preordering on $\mathcal{R}_d$.

From now on, the notation $\psi = -\varphi$ is used just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$ (informally this means that formulas ψ and φ are each other's negation).

Definition 2.2.2 (Strict and defeasible rules, Modgil and Prakken (2013))**.** Let $\varphi_1, \ldots, \varphi_n, \varphi$ be elements of $\mathcal{L}$.

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

This means that instantiating the strict rules only with all valid inferences of a certain logic requires the logic to be monotonic. This can be easily demonstrated by an example. Suppose the logic L is non monotonic, then there must be three sets Γ, Δ and Λ such that $\Delta \vdash_L \Lambda$, but not $\Gamma, \Delta \vdash_L \Lambda$. This would mean that $\Delta \rightarrow \Lambda$, but not $\Gamma, \Delta \rightarrow \Lambda$, so if Δ holds then it is not the case that Λ holds without exception. Therefore, the used logic has to be monotonic.

Definition 2.2.3 (Knowledge base, Modgil and Prakken (2013))**.** A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n, \preceq)$ is a tuple $KB = (\mathcal{K}, \preceq')$, where $\mathcal{K} \subseteq \mathcal{L}$ consists of two disjoint subsets $\mathcal{K}_n$ (the axioms or necessary premises) and $\mathcal{K}_p$ (the ordinary premises) and where $\preceq'$ is a preordering on $\mathcal{K}_p$.

Note that this is slightly different than the definition for knowledge bases in Modgil and Prakken (2013) since there are no *assumptions* in the knowledge base.

Definition 2.2.4 (Argumentation theory, Modgil and Prakken (2013))**.** An *argumentation theory* is a tuple $AT = (AS, KB)$, where AS is an argumentation system and KB is a knowledge base in AS.

The inference rules of an argumentation system are used to construct arguments from a knowledge base. The next definition shows how an argument can be constructed.

Definition 2.2.5 (Argument, Modgil and Prakken (2013))**.** An *argument* A on the basis of a knowledge base $(\mathcal{K}, \preceq)$ in an argumentation system $(\mathcal{L}, \mathcal{R}, n, \preceq')$:

1. φ if $\varphi \in \mathcal{K}$ with
   $\texttt{Prem}(\varphi) = \{\varphi\}$,
   $\texttt{Conc}(\varphi) = \varphi$,
   $\texttt{Sub}(A) = \{\varphi\}$,
   $\texttt{DefRules}(A) = \emptyset$,
   $\texttt{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 $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$ in $\mathcal{R}_s$.

$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
$\texttt{Conc}(A) = \psi$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
$\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$,
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$.

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

From now on, each of these functions $\texttt{Func}$ defined above can also be used on a set of arguments $S = \{A_1, \ldots, A_n\}$ in the following way: $\texttt{Func}(S) = \texttt{Func}(A_1) \cup \ldots \cup \texttt{Func}(A_n)$. Below, an example of an argumentation theory is given. This example will be used several times to clarify more definitions.

Example 2.2.1. Take the following knowledge base $KB = (\mathcal{K}, \preceq)$, where $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$, $\mathcal{K}_n = \{p, q\}$ and $\mathcal{K}_p = \{r, s, \neg t\}$. Take the following ordering $\preceq$ on the elements of $\mathcal{K}_p$: $\neg t \prec s \prec r$. The argumentation system AS is $(\mathcal{L}, \mathcal{R}, n, \preceq')$, where $\mathcal{L}$ is the first-order language, $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ with $\mathcal{R}_s$ is instantiated with all valid inferences in classical logic and $\mathcal{R}_d = \{p, r \Rightarrow t; \neg t \Rightarrow \neg s \vee \neg q\}$. Take the following ordering $\preceq'$ on the defeasible rules: $p, r \Rightarrow t \prec' \neg t \Rightarrow \neg s \vee \neg q$. The argumentation theory $AT = (AS, KB)$ has, among others, the following arguments:

$A_1 : p$	$A_2 : r$	$A_3 : A_1, A_2 \Rightarrow t$	
$B_1 : \neg t$	$B_2 : B_1 \Rightarrow \neg s \vee \neg q$	$B_3 : s$	$B_4 : q$
$C : B_2, B_3 \rightarrow \neg q$	$D : B_2, B_4 \rightarrow \neg s$		

In Figure 2.1 these arguments are visualised. The type of a premise is indicated with a superscript and defeasible inferences are displayed with dotted lines.

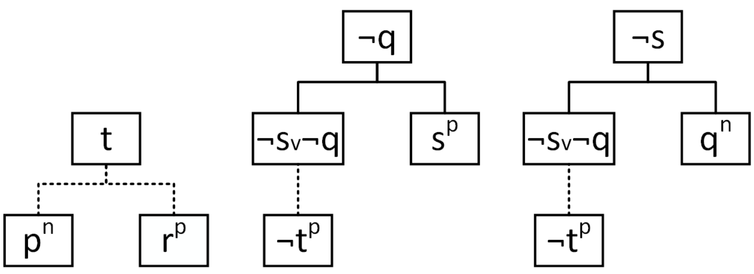

FIGURE 2.1: Arguments of Example 2.2.1

If an argument only uses strict rules, the argument is said to be *strict*. Otherwise it is *defeasible*. If a strict argument exists for φ with all premises in S, then this is noted as $S \vdash \varphi$. Otherwise the argument is defeasible, which is noted as $S \mid\!\sim \varphi$. If an argument only has necessary premises, then the argument is *firm*. Otherwise it is *plausible*. A *basic defeasible argument* is an argument which either has a defeasible top rule or is just an ordinary premise. The set of all basic defeasible arguments of a set of arguments S is denoted with $BD(S)$. The set of all necessary premises are denoted with $NP(S)$.

Definition 2.2.6 (Maximal fallible subarguments, Modgil and Prakken (2013))**.** For any argument A, an argument $A' \in \texttt{Sub}(A)$ is a *maximal fallible subargument* of A if

1. the top rule of A' is defeasible or A' is an ordinary premise;
2. there is no $A'' \in \texttt{Sub}(A)$ such that $A'' \neq A$ and $A' \in \texttt{Sub}(A'')$ and A'' satisfies condition (1).

The set of all maximal fallible subarguments of A are denoted by $M(A)$.

Example 2.2.2. In Example 2.2.1, the maximal fallible subarguments of argument C are B_2 and B_3.The only maximal fallible subargument of argument D is B_2.

Recall that an abstract argumentation framework is determined by a pair $(\mathcal{A}, att)$, where $\mathcal{A}$ is a set of arguments and att is a binary relation over $\mathcal{A}$ which represents the attacks between arguments. Up to now, the arguments of such a framework have been defined. In the following definitions, the attack relation between arguments is described. Note that the input for the attack relation of Dung (1995) is defined in terms of defeat (Definition 2.2.9).

Definition 2.2.7 (Attack, Modgil and Prakken (2013))**.**

- Argument A *undercuts* argument B (on B') if and only if $\texttt{Conc}(A) = -n(r)$ such that the top rule r of B' is defeasible.
- Argument A *rebuts* argument B (on B') if and only if $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.
- Argument A *undermines* B (on $B' = \varphi$) if and only if $\texttt{Conc}(A) = -\varphi$ for some ordinary premise B' of B.

Definition 2.2.8 (Structured argumentation framework, Modgil and Prakken (2013))**.** Let $AT = (AS, KB)$ be an argumentation theory. A *structured argumentation framework* (SAF) is a triple $(\mathcal{A}, att, \preceq)$ where $\mathcal{A}$ is the set of all finite arguments that can be constructed from KB in AS. $(X, Y) \in att$ if and only if X attacks Y and $\preceq$ is an ordering on the set of all arguments that can be constructed from KB in AS.

Argument A *successfully* rebuts or undermines an argument B if A rebuts or undermines B on B' and $A \not\prec B'$. With $A \preceq B'$ it is meant that argument B' is at least as 'good' as A according to the ordering on the arguments; so $A \not\prec B'$ means that B' is not strictly better than A.

Definition 2.2.9 (Defeat, Modgil and Prakken (2013))**.** Argument A *defeats* B if and only if A undercuts or successfully rebuts or successfully undermines B.

The next definitions introduce examples of commonly used argument orderings.

Definition 2.2.10 (Last defeasible rules, Modgil and Prakken (2013))**.** Let A be an argument, then:

- $\texttt{LastDefRules}(A) = \emptyset$ if and only if $\texttt{DefRules}(A) = \emptyset$
- If $A = A_1, \ldots, A_n \rightarrow \varphi$, then $\texttt{LastDefRules}(A) = \{\texttt{Conc}(A1), \ldots, \texttt{Conc}(A_n) \Rightarrow \varphi$, else $\texttt{LastDefRules}(A) = \texttt{LastDefRules}(A1) \cup \ldots \cup \texttt{LastDefRules}(An)$.

In the following definition $\texttt{Prem}_p(A)$ is defined as $\texttt{Prem}(A) \cap \mathcal{K}_p$ and $\lhd$ is a set ordering over sets of defeasible rules and ordinary premises.

Definition 2.2.11 (Last-link principle, Modgil and Prakken (2013))**.** $B \prec A$ under the *last-link principle* if and only if:

- $\texttt{LastDefRules}(A) \lhd \texttt{LastDefRules}(B)$; or
- $\texttt{LastDefRules}(A)$ and $\texttt{LastDefRules}(B)$ are empty and $\texttt{Prem}(A) \lhd \texttt{Prem}(B)$.

Definition 2.2.12 (Weakest-link principle, Modgil and Prakken (2013))**.** $B \prec A$ under the *weakest-link principle* if and only if:

- If both B and A are strict, then $\texttt{Prem}_p(B) \lhd \texttt{Prem}_p(A)$; else
- If both B and A are firm, then $\texttt{DefRules}(B) \lhd \texttt{DefRules}(A)$; else
- $\texttt{Prem}_p(B) \lhd \texttt{Prem}_p(A)$ and $\texttt{DefRules}(B) \lhd \texttt{DefRules}(A)$.

The next example illustrates the definitions given above.

Example 2.2.3. Take a look at the argumentation theory given in Example 2.2.1. Then the corresponding structured argumentation framework is $SAF = (\mathcal{A}, att, \preccurlyeq)$. Then $\{(B_1, A_3), (A_3, B_1), (D, B_3)\} \subseteq att$. Note that for example (B_4, C) and (C, B_4) are not elements of att because C has a strict top rule and B_4 is a necessary premise.
Suppose the weakest-link principle is used for ordering $\preccurlyeq$ the arguments. Then only the following attack succeeds as a defeat: $(A_3, B_1) \in \mathcal{D}$. This is visualised in Figure 2.2. The dotted arrows are the attacks and the other arrows are succeeded attacks, i.e. defeats.

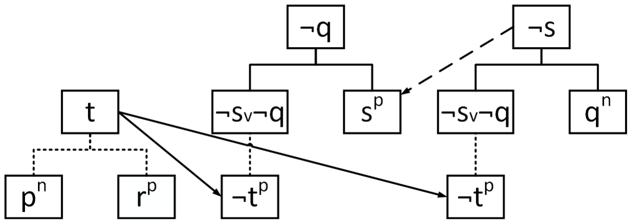

FIGURE 2.2: Attack and defeat according to Example 2.2.1

The next two definitions generalise the notion of closure and consistency.

Definition 2.2.13 (Closure, Modgil and Prakken (2013)). The *closure* for a set $X \in \mathcal{L}$, denoted $Cl_{\mathcal{R}_s}(X)$, with respect to the strict rules $\mathcal{R}_s$ is defined as the smallest set containing X and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in $Cl_{\mathcal{R}_s}(X)$. A set of arguments S is said to be *closed under subarguments* if $\texttt{Conc}(S) = \texttt{Conc}(\texttt{Sub}(S))$. A set of arguments S is said to be *closed under strict rules* if $\texttt{Conc}(S) = Cl_{\mathcal{R}_s}(\texttt{Conc}(S))$.

Definition 2.2.14 (Consistency, Modgil and Prakken (2013)). A set $X \subseteq \mathcal{L}$ is *consistent* if and only if there does not exist a φ such that $\varphi, \neg\varphi \in Cl_{\mathcal{R}_s}(X)$. Otherwise it is *inconsistent*. A set of arguments S is said to be *consistent* if $\texttt{Conc}(S)$ is consistent.

2.3 Linking the Frameworks

In the preceding paragraph, arguments and defeat relations between arguments in an $ASPIC^+$ framework for structured argumentation have been described. This constitutes the input of the abstract argumentation framework of Dung (1995), which results in the instantiation of the abstract argumentation framework with the $ASPIC^+$ framework. Note that the attack relation of the abstract argumentation framework of Dung (1995) corresponds with the defeat relation of the $ASPIC^+$ framework.

Definition 2.3.1 (AF corresponding to a SAF ($AFSAF$)). An *abstract argumentation framework corresponding to a* $SAF = (\mathcal{A}, att, \preceq)$ is a pair $(\mathcal{A}, \mathcal{D})$ such that $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by $(\mathcal{A}, att, \preceq)$.

Below, this definition is illustrated with an example.

Example 2.3.1. Take the $SAF = (\mathcal{A}, att, \preceq)$ of Example 2.2.1, then the abstract argumentation framework corresponding to SAF is $AFSAF = (\mathcal{A}, \mathcal{D})$, where $\mathcal{D} = \{(A_3, B_1)\}$.

Now that arguments and defeat relations have been covered, the problems mentioned in the introduction can be demonstrated by giving examples. This is done in Chapter 3.

Chapter 3

Problem Description

Below, the two main problems of this thesis are described by means of examples. The first problem is about the trivialisation of an argumentation system in case of a single inconsistency. The second problem is about the inefficiency of an argumentation system that has non-minimal arguments.

3.1 Trivialisation Problem

An example of Pollock (1994) is shown here, which will clarify the problem of self-defeat (this problem is also described in Grooters and Prakken (2014)).

Example 3.1.1. Take $\mathcal{R}_s$ such that it consists of all valid inferences in classical logic. Take $\mathcal{R}_d = \{p \Rightarrow q; r \Rightarrow \neg q; t \Rightarrow s\}$ and $\mathcal{K} = \mathcal{K}_p \cup \mathcal{K}_n$ where $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{p, r, t\}$. Then the following arguments can be constructed.

$A_1 : p$	$A_2 : A_1 \Rightarrow q$
$B_1 : r$	$B_2 : B_1 \Rightarrow \neg q$
$C_1 : A_2, B_2 \rightarrow \neg s$	
$D_1 : t$	$D_2 : D_1 \Rightarrow s$

In Figure 3.1, these arguments are visualised. The type of a premise is indicated with a superscript and defeasible inferences are displayed with dotted lines. The dotted boxes are arguments that are defeated.

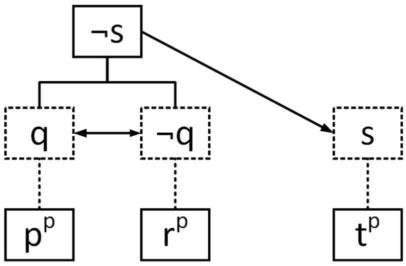

FIGURE 3.1: Problem of ex falso quodlibet

In Figure 3.1, depending on the preferences between arguments, D_2 can be rebutted by C_1. This illustrates that there might be unrelated arguments which are defeated by an argument which has $\perp$ as a premise.

In fact, the only semantics defined by Dung (1995) that has problems with this example is grounded semantics. Since A_2 and B_2 attack each other and at least one of these attacks will (with non-circular argument orderings) succeed as defeat, all preferred or stable extensions contain either A_2 or B_2, but not both. And since both A_2 and B_2 attack C_1 (by directly attacking one of its subarguments), C_1 is for each preferred or stable extension defeated by at least one argument in the extension. Therefore, C_1 is not in any of these extensions so D_2 is in all these extensions. However, if both A_2 and B_2 defeat each other, then neither of them is in the grounded extension, so that extension does not defend D_2 against C_1 and therefore does not contain D_2.

Pollock (1994, 1995) thought that this line of reasoning sufficed to show that his recursive-labelling approach (which was later proved by Jakobovits and Vermeir (1999) to be equivalent to preferred semantics) adequately deals with this problem. However, Caminada (2005) showed that the example can be extended in ways that also cause problems for preferred and stable semantics. Essentially, he replaced the facts p and r with defeasible arguments for p and r and let both these arguments be defeated by a self-defeating argument. Since such self-defeating arguments (if not themselves defeated) cannot be in a preferred extension, their targets A_2 nor B_2 can also not be in a preferred extension, because they are both indirectly defeated by the self-defeating argument. Therefore, D_2 cannot be defended against C_1 and so D_2 cannot be in a preferred extension.

A critic of $ASPIC^+$ or the approach of Pollock (1994, 1995) might argue that the problem is caused by the combination of strict (i.e., deductive) and defeasible inference rules. Indeed, in classical argumentation (Besnard and Hunter (2008); Gorogiannis and Hunter (2011)), the problem can be easily avoided by requiring that the premises of an argument are consistent. However, there are reasons to believe that classical logic is too strong to be able to model all forms of defeasible reasoning; see, for instance, the discussion by Modgil and Prakken (2013). Furthermore, in assumption-based argumentation (ABA)

(Dung, Kowalski, and Toni (2009)), which only has strict inference rules but does not require them to be classical, and which does not require that the premises of arguments are consistent, the problem may or may not arise depending on how it is instantiated. In ABA, which is shown to be a special case of the $ASPIC^+$ framework (Modgil and Prakken (2013)), arguments are built from necessary assumption premises $\mathcal{K}_n$ and strict inference rules $\mathcal{R}_s$. Additionally, an attack in ABA by one argument against another is a deduction by the first argument of the contrary of an assumption supporting the second argument. To show that the same problem arises in an ABA framework, take a look at the following example.

Example 3.1.2. Take $\mathcal{K}_n = \{p, \neg p, s\}$ and let $\mathcal{R}_s$ be derived from classical logic. Then the following arguments can be constructed.

$A : p$
$B : \neg p$
$C : A, B \to \neg s$
$D : s$

This again illustrates that there might be unrelated arguments, which are defeated by an argument that has $\bot$ as a premise.

In case a logic is used in which the inference rule $\{q, \neg q\} \to \neg s$ is invalid, this would prevent the system in Example 3.1.1 from trivialising. This is the main reason to examine the use of paraconsistent logics.

3.2 Efficiency Problem

Another research question is whether it makes a difference if it is required for a strict rule to be minimal. In the $ASPIC^+$ framework, it is not required for arguments to be *minimal*. However, the number of arguments that can be generated will be significantly reduced in case this restriction is introduced. This results in a more efficient system. The following example illustrates the problem.

Example 3.2.1. Suppose the strict rules of an argumentation system are instantiated with classical logic. Take the knowledge base $\mathcal{K}_p = \{p, q, r, s, t, u\}$. Then among other things, argument $A_1 : p, q \to p \wedge q$ can be constructed. Since classical logic is monotonic, all following arguments can also be constructed. Note that even more arguments can be constructed that conclude with $p \wedge q$. For example:

$A_2 : p, q, r \to p \wedge q$ $\quad$ $A_3 : p, q, s \to p \wedge q$
$A_4 : p, q, t \to p \wedge q$ $\quad$ $A_5 : p, q, u \to p \wedge q$
$A_6 : p, q, r, s \to p \wedge q$ $\quad$ $A_7 : p, q, t, u \to p \wedge q$

All arguments in the table of Example 3.2.1 can be considered as redundant since A_1 only uses the really needed information and it has the same conclusion as all other mentioned arguments. Note that this does not make sense for defeasible rules because these are non monotonic, so additional premises can make an important difference.

Dung, Toni, and Mancarella (2010) call these arguments *non redundant arguments* and they proved that, for the ABA framework (Dung et al. (2009)), redundant arguments can be dismissed without affecting the conclusion of extensions. ABA, as formulated in Dung, Mancarella, and Toni (2007), is shown to be a special case of the $ASPIC^+$ framework without preferences and defeasible rules (Prakken (2010)). Therefore, if the same result is found for the $ASPIC^+$ framework, then the results found by Dung et al. (2010) will be generalised.

First the focus lies on the trivialisation problem. Hereafter, this efficiency issue is examined for the $ASPIC^+$ and the later introduced $ASPIC^\star$ framework. The following chapter introduces the four paraconsistent logics.

Chapter 4

Paraconsistent Logics

A paraconsistent logic attempts to deal with a contradiction by tolerating it in a specific way. A logical consequence relation $\vdash$ is said to be paraconsistent if it is not 'explosive', i.e. when $\{A, \neg A\} \vdash B$ does not hold. Three candidate paraconsistent logics, the Logic of Paradox, Da Costa's C_ω system and the logic W, are described on the basis of their semantics and proof system, which will allow for a comparison between the applicability of the logics. For the logic W, the $ASPIC^\star$ framework is introduced in order to enforce that $ASPIC^+$'s strict part behaves according to the logic W.

Another well-known paraconsistent logic is the family of relevant logics. However, this logic is non monotonic, see Read (1988, p. 100). This is a problem since a monotonic logic is needed for the instantiation of the strict rules (as is explained after Definition 2.2.2). Otherwise it can be the case that the antecedents of a strict rule hold, while the conclusion fails and that is not in accordance with the definition of strict rules. For this reason, this logic is not further used.

These three logics are assumed to be consistent. It is shown why these logics are paraconsistent by giving a counterexample of explosion. In each logic, the formal system interprets the formulas of the language of first-order logic. However, the semantics differ for these logics, which is shown in the coming sections.

4.1 Logic of Paradox

The Logic of Paradox (Priest (1979), Priest (1989)) is obtained by relaxing the assumption that a sentence cannot be both true and false. Sentences in the Logic of Paradox (LP) can have two truth values instead of one. The set of possible truth values is $\{\{1\}, \{0\}, \{0, 1\}\}$, where $\{0, 1\}$ is the paradoxical 'true' and 'false'. Next, the semantics and proof theory are given which are sound and complete with respect to each other.

4.1.1 Semantics

The following truth definitions describe the trivalent valuation function (Priest (1979)). D denotes the domain of objects and I maps $p_n(\overrightarrow{x})$ to one of the three truth values $\{\{0\},\{1\}\{0,1\}\}$ with p_n an n-place predicate of the logic $\mathcal{L}$. With $B(d/x)$ for a $d \in D$ it is meant to replace all variables x of B with d.

1. (a) $1 \in v(\neg A) \Leftrightarrow 0 \in v(A)$
   (b) $0 \in v(\neg A) \Leftrightarrow 1 \in v(A)$
2. (a) $1 \in v(A \wedge B) \Leftrightarrow 1 \in v(A)$ and $1 \in v(B)$
   (b) $0 \in v(A \wedge B) \Leftrightarrow 0 \in v(A)$ or $0 \in v(B)$
3. (a) $1 \in v(A \vee B) \Leftrightarrow 1 \in v(A)$ or $1 \in v(B)$
   (b) $0 \in v(A \vee B) \Leftrightarrow 0 \in v(A)$ and $0 \in v(B)$
4. (a) $1 \in v(A \rightarrow B) \Leftrightarrow 0 \in v(A)$ or $1 \in v(B)$
   (b) $0 \in v(A \rightarrow B) \Leftrightarrow 1 \in v(A)$ and $0 \in v(B)$
5. (a) $1 \in v(p_n(\overrightarrow{x})) \Leftrightarrow 1 \in I(p_n(\overrightarrow{x}))$
   (b) $0 \in v(p_n(\overrightarrow{x})) \Leftrightarrow 0 \in I(p_n(\overrightarrow{x}))$
6. (a) $1 \in v(\forall x(B)) \Leftrightarrow$ such that $\forall d \in D$ $1 \in v(B(d/x))$
   (b) $0 \in v(\forall x(B)) \Leftrightarrow$ such that $\exists d \in D$ $0 \in v(B(d/x))$
7. (a) $1 \in v(\exists x(B)) \Leftrightarrow$ such that $\exists d \in D$ $1 \in v(B(d/x))$
   (b) $0 \in v(\exists x(B)) \Leftrightarrow$ such that $\forall d \in D$ $0 \in v(B(d/x))$

From this, it follows that disjunction and conjunction are treated dually: $\neg(A \wedge B) = \neg A \vee \neg B$ and $\neg(A \vee B) = \neg A \wedge \neg B$. $A \rightarrow B$ is defined as $\neg A \vee B$. As can be checked, these truth conditions are sufficient to give all formulas one of the three truth values.

An interpretation is a *model* of a formula f if and only if $1 \in v(f)$ holds in that interpretation. It is a model of a set of formulas if and only if it is a model of every formula in the set. The semantical notion of logical consequence is defined as follows:

- $\Sigma \vDash_{LP} A \;\Leftrightarrow$ for all evaluations v either $1 \in v(A)$ or for some $B \in \Sigma$, $v(B) = \{0\}$

Priest (1979) gives the following valid semantical consequences:

$A \vDash A \vee B$	$A, B \vDash A \wedge B$
$A \rightarrow (B \rightarrow C) \vDash B \rightarrow (A \rightarrow C)$	$A \vDash B \rightarrow A$
$\neg A \rightarrow \neg B \vDash B \rightarrow A$	$\neg(A \vee B) \vDash \neg A$

$\neg\neg A \models A$

$A \wedge B \models A$

$\neg A \models A \rightarrow B$

$A \rightarrow B \models \neg B \rightarrow \neg A$

$\neg(A \rightarrow B) \models A$

$A \rightarrow \neg A \models \neg A$

$\neg A \models \neg(A \wedge B)$

$A, \neg B \models \neg(A \rightarrow B)$

$A \rightarrow (A \rightarrow B) \models A \rightarrow B$

$\neg A, \neg B \models \neg(A \vee B)$

$A \models \neg\neg A$

$A \rightarrow B \models A \wedge C \rightarrow B \wedge C$

$(\forall x)A, (\forall x)B \models (\forall x)(A \wedge B)$

$(\forall x)(A \rightarrow B) \models (\forall x)(\neg B \rightarrow \neg A)$

$(\forall x)(A \rightarrow B) \models (\forall x)A \rightarrow (\forall x)B$

$(\forall x)A, (\forall x)B \models \neg(\exists x)\neg A$

$(\forall x)(A \rightarrow B) \models (\forall x)(\neg A \rightarrow B)$

$(\forall x)A \models (\forall x)(A \vee B)$

$(\forall x)(A \rightarrow B) \models (\exists x)A \rightarrow (\exists x)B$

$(\forall x)(A \rightarrow B) \models (\forall x)(B \rightarrow A)$

$(\forall x)A \models A(x/y)$

However, the following do not hold:

$A \wedge \neg A \models B$

$A, \neg A \vee B \models B$

$A \rightarrow B, \neg B \models \neg A$

$A \rightarrow B, B \rightarrow C \models A \rightarrow C$

$A, A \rightarrow B \models B$

$A \rightarrow B \wedge \neg B \models \neg A$

$(\forall x)A, (\forall x)(A \rightarrow B) \models (\forall x)B$

$(\forall x)(A \rightarrow B), (\forall x)(B \rightarrow C) \models (\forall x)(A \rightarrow C)$

$(\forall x)(A \rightarrow B), (\forall x)\neg B \models (\forall x)\neg A$

4.1.2 Proof Theory

The natural deduction system is almost the same as the natural deduction system for predicate logic (Priest (1989)). Only the negation rules $\neg I$ and $\neg E$ are modified (see below). This natural deduction system is sound and complete with respect to its semantics. Note that a discharged assumption A is 'daggered': $A^\dagger$.

$$\frac{A}{A \vee B}\ \vee I$$

$$\frac{A \vee B \qquad \begin{matrix} A^\dagger \\ \vdots \\ C \end{matrix} \qquad \begin{matrix} B^\dagger \\ \vdots \\ C \end{matrix}}{C}\ \vee E$$

$$\frac{\mathrm{A}}{(\forall x)A}\ \forall I$$

$$\frac{(\forall x)A}{A(a/x)}\ \forall E$$

The $\neg I$ and $\neg E$ rules are deleted:

$$\frac{\begin{array}{c} A^{\dagger} \\ \vdots \\ B \wedge \neg B \end{array}}{\neg A} \neg I$$

$$\frac{\begin{array}{c} \neg A^{\dagger} \\ \vdots \\ B \wedge \neg B \end{array}}{A} \neg E$$

These rules are replaced by:

$$\frac{\begin{array}{cc} A^{\dagger} & \\ \vdots & \\ \neg B & B \end{array}}{\neg A} \text{CON}$$

$$\frac{}{A \vee \neg A} \text{LEM}$$

$$\frac{\neg\neg A}{A} \text{DN}$$

Where in CON, A is the only undischarged assumption and no application of LEM occurs in its subproof. Recall that conjunction is treated dually and that $A \to B$ is defined as $\neg A \vee B$ and $(\exists x)A$ is defined as $\neg(\forall x)\neg A$. Therefore these rules are enough to specify the proof theory.

4.1.3 Counterexample for Explosion

The following sentence is valid in LP:

$A \wedge \neg A \to B$

This sentence would at first sight seem to be incorrect because it is almost like the principle of explosion. However, this sentence cannot be used as a rule of deduction. For example, take an evaluation such that B is strictly false and $A, \neg A$ are undetermined (both true and false), so $A \wedge \neg A$ is undetermined as well. Then $0 \in v(A \wedge \neg A)$, so by the valuation postulates $1 \in v(A \wedge \neg A \to B)$. According to the validity of sentences in models: $\{A \wedge \neg A\} \nvDash_{LP} B$ and also $\{A, \neg A\} \nvDash_{LP} B$. The natural deduction system is sound and complete to this semantics, which implies that $\{A, \neg A\} \vdash_{LP} B$ also does not hold. Therefore, the Logic of Paradox is a paraconsistent logic.

4.2 Da Costa's Basic C-system: C_ω

The basic C system C_ω of Da Costa (1974) (see also Priest (2011), Sylvan (1990)), adds the axioms $\neg\neg A \rightarrow A$ and $A \vee \neg A$ to positive logic, a negation free first-order logic (these added axioms are called the 'Dialectic Double Negation' (DDN) and 'Exclusive Middle' (EM) respectively). C_ω is in certain aspects the dual of intuitionistic logic, since in intuitionistic logic the axiom EM is invalid and the axiom Non-Contradiction (NC, $\neg(A \wedge \neg A)$) is valid. In C_ω, the axiom EM is valid and NC is invalid. Intuitionistic logic tolerates incomplete situations to avoid inconsistency, while the C-systems admit inconsistent situations, but incomplete situations are removed. For example, in C_ω it is possible that all three sentences A, $\neg A$, $\neg\neg A$ are true. However unlike in the Logic of Paradox, sentences can only have one truth value. Next the semantics and the proof theory are given which are sound and complete with respect to each other.

4.2.1 Semantics

C_ω has the following bivalent valuation for formulas built from a logical language $\mathcal{L}$ (Brożek (2012)). Again D denotes the domain of objects. With $A(d/x)$ for a $d \in D$ it is meant to replace all variables x of A with d.

1. $v(A \wedge B) = 1 \Leftrightarrow v(A) = 1$ and $v(B) = 1$
2. $v(A \vee B) = 1 \Leftrightarrow v(A) = 1$ or $v(B) = 1$
3. $v(A \rightarrow B) = 1 \Leftrightarrow v(A) = 0$ or $v(B) = 1$
4. $v(\neg A) = 1 \Leftarrow v(A) = 0$
5. $v(A) = 1 \Leftarrow v(\neg\neg A) = 1$
6. $v((\exists x)A) = 1 \Leftrightarrow$ there exists an d in D such that $v(A(d/x)) = 1$
7. $v((\forall x)A) = 1 \Leftrightarrow$ for every d in D $v(A(d/x)) = 1$ holds

An interpretation of a formula f by its valuation form is a *model* if and only if $v(f) = 1$ in that interpretation. An interpretation is a model of a set of formulas if and only if it is a model of every formula in the set. The semantical logical consequence:

- $\Sigma \vDash_{C_\omega} A \;\Leftrightarrow$ for all evaluations v either $v(A) = 1$ or for some $B \in \Sigma$, $v(B) = 0$

4.2.2 Proof Theory

The following logical axioms specify the proof theory of the system of Da Costa (1974). This proof theory is sound and complete with respect to its semantics.

$A \to (B \to A)$	$(A \to B) \to (A \to (B \to C) \to (A \to C))$
$A \land B \to A$	$A \land B \to B$
$A \to (B \to B \land A)$	$A \to (A \lor B)$
$B \to A \lor B$	$(A \to B) \to (B \to C) \to (A \lor B \to C)$
$A \lor \neg A$	$\neg\neg A \to A$
$A(d/x) \to (\exists x)A$	$(\forall x)A \to A(d/x)$
$B \to A(x) \vdash_{C_\omega} B \to (\forall x)A$	$A(x) \to B \vdash_{C_\omega} (\exists x)A \to B$

Where x does not occur in B.

With the following deduction rule:
$\{A,\ A \to B\} \vdash_{C_\omega} B$
These rules together specify the proof theory.

4.2.3 Counterexample for Explosion

In order to show that $\{A, \neg A\} \vdash_{C_\omega} B$ does not hold, a counter model is given. It is easy to show that explosion fails since the following valuation function can be chosen: $v(A),\ v(\neg A) = 1$ while $v(B) = 0$. Therefore it can be concluded that $\{A, \neg A\} \nvdash_{C_\omega} B$.

4.3 Logic W

Finally, the so-called weak consequence relation originally proposed by Rescher and Manor (1970) is investigated. It is based on the philosophy of preservationists. The idea of preservationism is to reason from inconsistent sets by using consistent subsets. There are different ways to do this; this thesis focusses on reasoning from maximally consistent subsets (since this is a monotonic consequence relation).

Definition 4.3.1 (Weak consequence relation, $\vdash_W$)**.** $\Gamma \vdash_W \alpha$ if and only if there is a maximal consistent subset Δ of Γ such that $\Delta \vdash \alpha$ in classical logic.

Note that the word 'maximal' is not required according to Lindenbaum's Lemma.

Lemma 4.3.1 (Lindenbaum)**.** Every consistent set of formulas can be extended into a maximally consistent one.

4.3.1 Counterexample for Explosion

It is easy to see that $\{a, \neg a\} \vdash_W b$ does not hold, because $\{a, \neg a\}$ is not a maximal consistent subset of $\{a, \neg a\}$. Therefore, this consequence relation is paraconsistent.

4.3.2 Monotonicity and Other Properties

Below, it is shown that two important properties hold. A counterexample demonstrates that the third property does not hold.

[R] $\alpha \in \Gamma$, then $\Gamma \vdash_W \alpha$.
Each $\alpha \in \Gamma$ belongs to some maximally consistent subset Δ of Γ. In classical logic, it holds that if $\alpha \in \Delta$, then $\Delta \vdash \alpha$. Therefore, it obviously holds that $\Gamma \vdash_W \alpha$.

[Mon] $\Gamma \vdash_W \alpha$, then $\Gamma, \Pi \vdash_W \alpha$.
The monotonicity property can be proven as follows. There must be a maximal consistent subset Δ of Γ such that $\Delta \vdash \alpha$. Since $\Delta \subseteq \Gamma \cup \Pi$, there must exist a maximal consistent extension of Δ in $\Gamma \cup \Pi$, Δ', such that $\Delta' \vdash \alpha$. Therefore, $\Gamma, \Pi \vdash_W \alpha$.

[Cut] $\Gamma, \alpha \vdash_W \beta$ and $\Gamma \vdash_W \alpha$, then $\Gamma \vdash_W \beta$.
This rule does not hold, which is shown by a counterexample. Take the set $\Gamma = \{a, \neg a \wedge b\}$. Then $\Gamma \vdash_W b$ and $\Gamma, b \vdash_W a \wedge b$, while it is not the case that $\Gamma \vdash_W a \wedge b$.

Since the **Cut** rule does not hold, a naive instantiation of *ASPIC*$^+$'s strict rules with this logic W would lead to a trivial system, as shown in the following example:

Example 4.3.1. Take the following knowledge base $\mathcal{K}_p = \{p, \neg p, r\}$, $\mathcal{K}_n = \emptyset$, instantiate the strict rules with all valid inferences in the logic W and take $\mathcal{R}_d = \emptyset$. Then the following arguments can be constructed:

$$\begin{array}{ll} A_1 : p & A_2 : A_1 \to p \vee \neg r \\ B : \neg p & C : A_2, B \to \neg r \\ D : r & \end{array}$$

Argument C concludes with $\neg r$ while $\mathcal{K}_p \nvdash_W \neg r$.

The underlying reason for this problem is that the **Cut** rule does not hold for $\vdash_W$. So if we want *ASPIC*$^+$'s strict part to behave according to $\vdash_W$, chaining of strict rules should be excluded.[1] In Example 4.3.1, the argument D is not allowed since A_2 and C both have a strict top rule. The prohibition of the chaining of strict rules will prevent explosion. To this end, the *ASPIC*$^+$'s notion of an argument is redefined, which results in the *ASPIC** framework.

[1]A similar idea was suggested by Martin Caminada in personal communication.

4.3.3 *ASPIC** Framework

In the *ASPIC** framework, it is not allowed to construct arguments that use more than one strict rule after each other. Below, the adapted definition for arguments in an *ASPIC** framework is presented:

Definition 4.3.2 (Argument* in *ASPIC**)**.** An *argument** A on the basis of a knowledge base $KB = (\mathcal{K}, \preceq)$ in an argumentation system $(\mathcal{L}, \mathcal{R}, n, \preceq')$ is:

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

2. $A_1, \ldots, A_n \rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments* with a defeasible top rule or are from $\mathcal{K}$ and such that there exists a strict rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$ in $\mathcal{R}_s$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
   $\texttt{Conc}(A) = \psi$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
   $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$,
   $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$.

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

Arguments* are just a special case of 'normal' arguments. Therefore, all definitions for arguments are the same in case the term argument can be replaced by argument* without problems. *Attack* and *defeat* are just the same for arguments*. A *structured argumentation framework* for the *ASPIC** framework, is just the same except that it only contains arguments*.

A *basic defeasible argument** is an argument* which either has a defeasible top rule or is just an ordinary premise. The set of all basic defeasible arguments* of a set of arguments* S is denoted with $BD^\star(S)$. The set of all necessary premises are denoted with $NP^\star(S)$.

Definition 4.3.3 (Maximal fallible subarguments*)**.** For any argument A, an argument* $A' \in \texttt{Sub}(A)$ is a *maximal fallible subargument** of A if

1. the top rule of A' is defeasible or A' is a non-axiom premise;
2. there is no $A'' \in \texttt{Sub}(A)$ such that $A'' \neq A$ and $A' \in \texttt{Sub}(A'')$ and A'' satisfies condition (1).

The set of all maximal fallible subarguments* of A are denoted by $M^\star(A)$.

Define for any set S of arguments* $S^\#$ as the set of all basic defeasible arguments* together with all necessary premises in S. Then closure and consistency are defined as follows.

Definition 4.3.4 (Closure*)**.** For any $X \subseteq \mathcal{L}$, let the *closure** of X under strict rules, denoted $Cl^\star_{\mathcal{R}_s}(X)$, be the smallest set containing X and the consequent of any strict rule in $\mathcal{R}_s$ whose antecedents are in X. The set of arguments* S is said to be *closed** *under subarguments* if and only if $\texttt{Conc}(S) = \texttt{Conc}(\texttt{Sub}(S))$. The set of arguments* S is said to be *closed** *under strict rules* if and only if $\texttt{Conc}(S) = Cl^\star_{\mathcal{R}_s}(\texttt{Conc}(S^\#))$.

Definition 4.3.5 (Consistency*)**.** A set $X \subseteq \mathcal{L}$ is *consistent** if there is not a $\varphi \in \mathcal{L}$ such that $\varphi, \neg\varphi \in Cl^\star_{\mathcal{R}_s}(X)$. Otherwise it is *inconsistent**. A set of arguments* S is said to be *consistent** if $\texttt{Conc}(S^\#)$ is consistent*.

The rest of the framework remains unchanged.

Now that the framework has been modified and the logics are introduced, we go back to the trivialisation problem to see how these logics perform when it comes to the satisfaction of the rationality postulates.

Chapter 5

Tackling the Trivialisation Problem

This chapter tries to solve the trivialisation problem discussed in Chapter 3. In order to solve this problem, this chapter investigates whether it 'makes sense' to instantiate the strict rules with a monotonic paraconsistent logic. The rationality postulates are relevant requirements on extensions which determine whether the instantiation makes sense. These rationality postulates are first introduced, then some results of Modgil and Prakken (2013); Dung and Thang (2014) are presented which impose conditions that imply these postulates. Finally, these conditions are used to determine whether the postulates are satisfied or not.

5.1 The Rationality Postulates

A logic-based argumentation system should be intuitive and natural when it comes to the logical closure and consistency of an extension. For example, it is counter-intuitive if an argument A is in an extension, while a subargument of A is not; or if two conclusions of arguments A and B are contradictory, while A and B are both in an extension. To avoid these situations, Caminada and Amgoud (2007) defined rationality postulates which impose requirements on any extension of an argumentation framework.

Subargument closure: For every complete extension E, E is closed under subarguments.

Closure under strict rules: For every complete extension E, E is closed under strict rules.

Consistency: For every complete extension E, E is consistent.

As mentioned above, the closure and consistency postulates also have to be checked for the logic W for the $ASPIC^{\star}$ framework. Therefore, the rationality postulates have to be redefined for the $ASPIC^{\star}$ framework.

Subargument* closure*: For every complete extension E, E is closed* under subarguments*.

Closure* under strict rules: For every complete extension E, E is closed* under strict rules.

Consistency*: For every complete extension E, E is consistent*.

Modgil and Prakken (2013) have provided some conditions that imply these rationality postulates for the $ASPIC^{+}$ framework. The next section exposes their results in order to check whether these conditions hold.

5.2 Modgil and Prakken's Conditions

Modgil and Prakken (2013) proved that there are three conditions which imply the satisfaction of the subargument closure, closure under strict rules and the consistency postulates. The first condition is *closure under transposition* or *closure under contraposition.*

Definition 5.2.1 (Closure under transposition, Modgil and Prakken (2013))**.** A set of strict rules $\mathcal{R}_s$ is said to be *closed under transposition* if for each rule $\varphi_1, \ldots, \varphi_n \rightarrow \psi$ in $\mathcal{R}_s$, all the rules of the form $\varphi_1, \ldots, \varphi_{i-1}, \neg\psi, \varphi_{i+1}, \ldots, \varphi_n \rightarrow \neg\varphi_i$ also belong to $\mathcal{R}_s$. An argumentation theory (AS, KB) is *closed under transposition* if the strict rules $\mathcal{R}_s$ of AS are closed under transposition.

Definition 5.2.2 (Closure under contraposition, Modgil and Prakken (2013))**.** An argumentation system is said to be *closed under contraposition* if for all $X \subseteq \mathcal{L}$, all $s \in X$ and all φ it holds that if $X \vdash \varphi$ then $X \backslash \{s\} \cup \{\neg\varphi\} \vdash \neg s$. An argumentation theory (AS, KB) is *closed under contraposition* if the argumentation system AS is closed under contraposition.

The following definition is needed for the second condition.

Definition 5.2.3 (Strict extensions, Modgil and Prakken (2013))**.** The argument A is a *strict extension of a set of arguments* $A_1, \ldots, A_n$ if and only if:

- The ordinary and assumption premises are exactly those in $A_1, \ldots, A_n$;
- The defeasible rules are exactly those in $A_1, \ldots, A_n$;

- The strict rules and axiom premises of A are a superset of the strict rules and axiom premises in $A_1, \ldots, A_n$.

An argument $A_1, \ldots, A_n \to \varphi$ is called a *strict argument* over $\{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n)\}$.

The second condition:

Definition 5.2.4 (Reasonable argument ordering, Modgil and Prakken (2013))**.** $\preceq$ is a *reasonable argument ordering* if and only if:

- $\forall A, B$, if A is strict and firm and B is plausible or defeasible, then $B \prec A$;
- $\forall A, B$, if B is strict and firm, then $B \not\prec A$;
- $\forall A, A', B, C$, such that $C \prec A$, $A \prec B$ and A' is a strict extension of $\{A\}$, then $A' \prec B$, $C \prec A'$;
- Let $\{C_1, \ldots, C_n\}$ be a finite subset of $\mathcal{A}$ and for $i = 1, \ldots, n$ let $C^{+/i}$ be some strict extension of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$. Then it is not the case that $\forall i,\ C^{+/i} \prec C_i$.

Two examples of reasonable argument orderings are the weakest- and last-link principle (Modgil and Prakken (2013)).

The third condition is *axiom consistency.*

Definition 5.2.5 (Axiom consistent, Modgil and Prakken (2013))**.** An argumentation theory is *axiom consistent* if and only if $Cl_{\mathcal{R}_s}(\mathcal{K}_n)$ is consistent.

These three conditions imply the rationality postulates for the *ASPIC*$^+$ framework.

Theorem 5.2.1 (Modgil and Prakken (2013))**.** If an argumentation theory (AS, KB) is closed under transposition or contraposition, is axiom consistent and if the argument preference ordering is reasonable, then the closure and consistency postulates are satisfied.

The results of Modgil and Prakken (2013) require an argumentation theory to be closed under transposition or contraposition, this implies that not all logics are covered by these results. Now, the closure(*) and consistency(*) postulates are checked for these paraconsistent logics. For the *ASPIC** framework, replace the word arguments by arguments*. For the logic W, it is first checked whether these conditions hold before the results are generalised for the *ASPIC** framework.

5.3 The Postulates for the Paraconsistent Logics

In this section, it is tried to prove or disprove the satisfaction of the closure($\star$) and consistency($\star$) postulates for the three paraconsistent logics.

5.3.1 Logic of Paradox

The consistency postulate is not satisfied in case the strict rules are instantiated with all valid inferences in the Logic of Paradox. A counterexample[1] for this property is as follows.

Example 5.3.1. Take an argumentation framework $AF = (\mathcal{A}, att)$. Suppose that $\mathcal{K}$ of the knowledge base is $\mathcal{K}_p \cup \mathcal{K}_n$, with $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{a, \neg a \vee b, \neg a \vee c, \neg b \vee \neg c\}$. Further suppose that $\mathcal{R}_s$ contains all valid inferences in the Logic of Paradox and that there are no defeasible rules ($\mathcal{R}_d = \emptyset$).

It is easily checked that $\mathcal{K}_p$ implies that at least one of a, b or c must be paradoxical. Therefore, there exists an argument $A_1 : a, \neg a \vee b, \neg a \vee c, \neg b \vee \neg c \rightarrow (a \wedge \neg a) \vee (b \wedge \neg b) \vee (c \wedge \neg c)$. Since tautologies are preserved in the Logic of Paradox (Priest (1989)), $\neg(b \wedge \neg b), \neg(a \wedge \neg a)$ and $\neg(c \wedge \neg c)$ are also entailed by $\mathcal{K}$. This implies that there exists an argument $A_2 : a, \neg a \vee b, \neg a \vee c, \neg b \vee \neg c \rightarrow \neg((a \wedge \neg a) \vee (b \wedge \neg b) \vee (c \wedge \neg c))$. These arguments only use strict rules so they can only be attacked on their premises. However, there does not exist an argument built from $\mathcal{K}$ which has a conclusion $\neg d$ for a $d \in \mathcal{K}_p$. To show this, for each $d \in \mathcal{K}_p$ a model has to be found for which $v(d) = \{1\}$ holds.

Model 1: to show that it is not the case that $\neg a$ follows from $\mathcal{K}_p$.
Take the model $v(a) = \{1\}, v(b) = \{1\}$ and $v(c) = \{0, 1\}$. Then it is clear that $v(a) = \{1\}, v(\neg a \vee b) = \{1\}, v(\neg a \vee c) = \{0, 1\}$ and $v(\neg b \vee \neg c) = \{0, 1\}$, but not $1 \in v(\neg a)$.

Model 2: to show that it is not the case that $\neg(\neg a \vee b)$ follows from $\mathcal{K}_p$.
Take again the model $v(a) = \{1\}, v(b) = \{1\}$ and $v(c) = \{0, 1\}$. Then it is clear that $v(a) = \{1\}, v(\neg a \vee b) = \{1\}, v(\neg a \vee c) = \{0, 1\}$ and $v(\neg b \vee \neg c) = \{0, 1\}$, but not $1 \in v(\neg(\neg a \vee b))$.

Model 3: to show that it is not the case that $\neg(\neg a \vee c)$ follows from $\mathcal{K}_p$.
Take the model $v(a) = \{1\}, v(b) = \{0, 1\}$ and $v(c) = \{1\}$. Then it is clear that $v(a) = \{1\}, v(\neg a \vee b) = \{0, 1\}, v(\neg a \vee c) = \{1\}$ and $v(\neg b \vee \neg c) = \{0, 1\}$, but not $1 \in v(\neg(\neg a \vee c))$.

Model 4: to show that it is not the case that $\neg(\neg b \vee \neg c)$ follows from $\mathcal{K}_p$.
Take the model $v(a) = \{0, 1\}, v(b) = \{0\}$ and $v(c) = \{0\}$. Then it is clear that $v(a) = \{0, 1\}, v(\neg a \vee b) = \{0, 1\}, v(\neg a \vee c) = \{0, 1\}$ and $v(\neg b \vee \neg c) = \{1\}$, but not $1 \in v(\neg(\neg b \vee \neg c))$.

[1] This counterexample was brought to my attention by Priest (personal communication).

This means that there are no arguments which attack one of the arguments A_1 and A_2, so A_1 and A_2 are elements of a complete extension E. This means that $(a \vee \neg a) \vee (b \vee \neg b) \vee (c \vee \neg c)$ and $\neg((a \vee \neg a) \vee (b \vee \neg b) \vee (c \vee \neg c))$ are elements of $\texttt{Conc}(E)$. Therefore, this logic-associated argumentation framework does not satisfy the consistency postulate. Note that this is also a counterexample for the satisfaction of the consistency postulate for a SAF by taking an ordering in which all premises are equally strong.

Since the aim is to try to find a paraconsistent logic such that both the closure and consistency postulates are satisfied, this logic is not further inspected.

5.3.2 Da Costa's Basic C-system: C_ω

A counterexample for the consistency postulate is given for Da Costa's C_ω system.

Example 5.3.2. Take an argumentation framework $AF = (\mathcal{A}, att)$. Suppose that $\mathcal{K}$ of the knowledge base is $\mathcal{K} = \mathcal{K}_p \cup \mathcal{K}_n$, with $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{a, a \supset b, a \supset \neg b\}$ with the following valuation: $v(a) = 1$, $v(a \supset b) = 1$ and $v(a \supset \neg b) = 1$. Further suppose that $\mathcal{R}_s$ contains all valid inferences in C_ω and that there are no defeasible rules ($\mathcal{R}_d = \emptyset$). Then the following two arguments exist: $A_1 : a, a \supset b \to b$ and $A_2 : a, a \supset \neg b \to \neg b$. These are shown in Figure 5.1. These two arguments both use a strict rule. This means that these arguments can only be attacked on their premises. However, none of $\neg a$, $\neg(a \supset b)$, $\neg(a \supset \neg b)$ can be derived from $CN(\mathcal{K})$ by using $\mathcal{R}_s$, so there are no arguments which attack A_1 or A_2 on their premises. Therefore, A_1 and A_2 will be elements of a complete extension E, which means that $b, \neg b \in \texttt{Conc}(E)$. It can be concluded that this logic-associated argumentation framework does not satisfy the consistency postulate. Note that this is also a counterexample for the satisfaction of the consistency postulate for a SAF by taking an ordering in which all premises are equally strong.

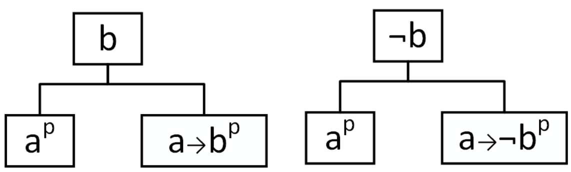

FIGURE 5.1: Elements of a complete extension

Since the consistency postulate is not satisfied, this logic is also not further inspected.

5.3.3 Logic W

Every attempt failed to find a counterexample for the logic W in the $ASPIC^\star$ framework, so it has been tried to use the results of Modgil and Prakken (2013) in order to prove the postulates. However, this is not possible since there is a counterexample for both closure under trans- and contraposition. These counterexamples are given below.

5.3.3.1 Counterexample to Closure under Transposition

The following example shows that closure under transposition does not hold in case the strict rules are instantiated with the logic W.

Example 5.3.3. Since $\{a, b, c\} \vdash a \wedge b$ and because of the monotonicity of the logic W, it holds that $\{a, b, c, \neg a\} \vdash_W a \wedge b$. This means that $a, b, c, \neg a \rightarrow a \wedge b$ is in $\mathcal{R}_s$. However, there is no maximal consistent subset of $\{a, b, \neg(a \wedge b), \neg a\}$ that proves $\neg c$ in classical logic. Therefore $\{a, b, \neg(a \wedge b), \neg a\} \nvdash_W \neg c$ and so $a, b, \neg(a \wedge b), \neg a \rightarrow \neg c \notin \mathcal{R}_s$. This means that if the strict rules $\mathcal{R}_s$ in an argumentation system AS of the argumentation theory $AT = (AS, KB)$ are instantiated with the valid inferences in the logic W, then the argumentation theory AT is not closed under transposition.

5.3.3.2 Counterexample to Closure under Contraposition

The following example shows that closure under contraposition does not hold in case the strict rules are instantiated with the logic W.

Example 5.3.4. Take the following knowledge base $(X, \preceq)$: $X = \mathcal{K}_n \cup \mathcal{K}_p$ with $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{a, b, c, \neg a\}$. Since $\{a, b, c, \neg a\} \vdash_W a \wedge b$ and $\{a, b, \neg(a \wedge b), \neg a\} \nvdash_W \neg c$, it follows that there is a strict argument* $\{a, b, c, \neg a\} \vdash a \wedge b$, but there is not a strict argument* $\{a, b, \neg(a \wedge b), \neg a\} \vdash \neg c$ (because chaining of strict rules is not allowed). Therefore, if the strict rules $\mathcal{R}_s$ in an argumentation system AS of the argumentation theory AT are instantiated with the valid inferences in the logic W, then AT is not closed under contraposition.

The results of Modgil and Prakken (2013) cannot be used to prove the postulates. Dung and Thang (2014) provide weaker conditions that imply these postulates. In the following sections their results are presented, generalised and exploited in order to prove the postulates for the logic W.

5.4 Dung and Thang's Conditions

To be able to use the results of Dung and Thang (2014), some of their terminology has to be introduced first. Subsequently, it is shown that their results can be generalised. In the end it is checked whether the rationality postulates are satisfied.

Closure and consistency are defined as the set of consequences following from the monotonic part of the underlying logic. Amgoud and Besnard (2009) have introduced the idea of using Tarski's abstract logic. Dung and Thang (2014) describe a slightly generalised version of Tarski abstract logics by using a Tarski-like monotonic consequence operator CN to represent the monotonic part. It is defined as follows:

Definition 5.4.1 (CN, Dung and Thang (2014))**.** An operator $CN : 2^{\mathcal{L}} \longrightarrow 2^{\mathcal{L}}$ for a set of formulas $X \subseteq \mathcal{L}$ satisfies:

- $X \subseteq CN(X)$
- $CN(X) = CN(CN(X))$
- $CN(X) = \bigcup\{CN(Y)|Y \subseteq X \text{ and } Y \text{ is finite}\}$
- $CN(\emptyset)$ is consistent

The slightly generalised version of Tarski abstract logics:

Definition 5.4.2 (Abstract logic, Dung and Thang (2014))**.** Given a language $\mathcal{L}$, an *abstract logic* is defined as a pair $(CN, CONTRA)$, where $CONTRA \subseteq 2^{\mathcal{L}}$ is a collection of contradictory sets such that if $S \in CONTRA$, then each superset of S also belongs to $CONTRA$.

Given an abstract logic $(CN, CONTRA)$, $X \subseteq \mathcal{L}$ is *closed* if $X = CN(X)$. X is said to be *contradictory* if $X \in CONTRA$. A set $X \subseteq \mathcal{L}$ is *inconsistent* if its closure is contradictory, otherwise X is *consistent*. This means that an inconsistent set may not be contradictory. X is *minimally inconsistent* if X is inconsistent, but each proper subset is consistent. Dung and Thang (2014) use abstract logics to represent the set of conclusions of arguments.

Definition 5.4.3 (Logic-associated argumentation framework, Dung and Thang (2014))**.** A *logic-associated argumentation framework* (LAF) over a language $\mathcal{L}$ is a quadruple $(AF, \sqsubseteq, AL, Cnl)$, where:

- AF is an abstract argumentation framework.
- $AL = (CN, CONTRA)$ is an abstract logic over $\mathcal{L}$.
- $Cnl : \mathcal{A} \longrightarrow \mathcal{L}$ assigns to each argument A, its conclusion $Cnl(A)$ in $\mathcal{L}$.
- $\sqsubseteq$ is a partial order over $\mathcal{A}$ where $A \sqsubseteq B$ means that A is a subargument of B such that for all arguments $C \in \mathcal{A}$, if C attacks A then C attacks B.

Definition 5.4.4 (Rule-based argumentation system, Dung and Thang (2014))**.** A *rule-based argumentation system* is a pair $AS = (\mathcal{R}_s, \mathcal{R}_d)$ of a set $\mathcal{R}_s$ of strict rules and a set $\mathcal{R}_d$ of defeasible rules such that $CN_{\mathcal{R}_s}(\emptyset)$ is not contradictory.

The next definition identifies the abstract logic underlying a rule-based argumentation system $AS = (\mathcal{R}_s, \mathcal{R}_d)$.

Definition 5.4.5 (AL_{AS}, Dung and Thang (2014))**.** Let $AS = (\mathcal{R}_s, \mathcal{R}_d)$ be a rule-based argumentation system. Define $AL_{AS} = (CN_{AS}, CONTRA_{AS})$, where $CONTRA_{AS}$ is the collection of contradictory sets and $CN_{AS} = Cl_{\mathcal{R}_s}$.

Now, the next lemma follows by definition of AL_{AS}.

Lemma 5.4.1 (Dung and Thang (2014))**.** AL_{AS} is an abstract logic.

Note that this means that $Cl_{\mathcal{R}_s}$ satisfies the conditions of the CN operator, if $\mathcal{R}_s$ is instantiated with inference rules such that $Cl_{\mathcal{R}_s}(\emptyset)$ is not contradictory. Then the next lemma follows from the properties of $Cl_{\mathcal{R}_s}$.

Lemma 5.4.2 (Dung and Thang (2014))**.** Let AF be an abstract argumentation framework and $\sqsubseteq$ be a partial order on the arguments according to the subarguments operator $\texttt{Sub}$, then $(AF, \sqsubseteq, AL_{AS}, \texttt{Conc})$ is a logic-associated argumentation framework.

The following results give conditions that guarantee the closure and consistency postulates described in the previous chapter. The closure and consistency postulates are based on an intuitive idea of a base of an argument and the set of generated arguments from a base.

Definition 5.4.6 (Base of an argument, Dung and Thang (2014))**.** Let A be an argument and BA be a finite set of subarguments of A. BA is a *base* of A ($BA = \texttt{Base}(A)$) if

- $\texttt{Conc}(A) \in Cl_{\mathcal{R}_s}(\texttt{Conc}(B))$;
- For each argument C, C attacks A if and only if C attacks B.

The following example shows the intuitive idea of a base.

Example 5.4.1. Take $\mathcal{K}_n = \emptyset$, $\mathcal{K}_p = \{a, b\}$, $\mathcal{R}_s = \{c \to d\}$ and $\mathcal{R}_d = \{a, b \Rightarrow c\}$. Then the following arguments can be constructed: $A_1 : a$, $A_2 : b$, $A_3 : A_1, A_2 \Rightarrow c$ and $A_4 : A_3 \to d$. See Figure 5.2.

A_4 can only be attacked on its subarguments A_1, A_2, or A_3 because of the strict top rule. Every argument that attacks A_1 or A_2 also attacks A_3, so every argument that

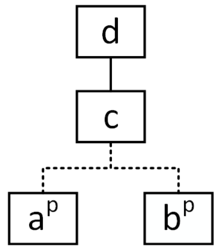

FIGURE 5.2: Arguments of Example 5.4.1

attacks A_4 also attacks A_3. It is easy to see that every argument that attacks A_3 also attacks A_4. $\texttt{Conc}(A_4) \subseteq CN(\texttt{Conc}(A_3))$, so $\{A_3\}$ is a base of A_4. It also holds that $M(A) = \{A_3\}$. The same kind of reasoning applies to the fact that the set $\{A_1, A_2, A_3\}$ is also a base of A_4.
However note that the set $\{A_1, A_2\}$ is not a base of A_4, because A_3 can be rebutted without A_1 or A_2 being attacked.

Note that for any argument A, the set of maximal fallible subarguments $M(A)$ is always a base. This can be easily seen by the fact that $\texttt{Conc}(A) \in Cl_{\mathcal{R}_s}(\texttt{Conc}(M(A)))$. Furthermore, if an argument B attacks a maximal fallible subargument A' of A, then B also attacks A. And if an argument B attacks A, then B must attack an ordinary premise or an argument with a defeasible toprule A''. A'' must be a subargument of a maximal fallible subargument A' of A. Therefore, B attacks $M(A)$. However, the converse does not need to hold: a base is not always the set of maximal fallible subarguments. This is caused by the fact that if BA is a base, then every extension of BA is also a base. More precisely for every base of A it holds that $M(A) \subseteq \texttt{Base}(A)$.

Dung and Thang (2014) have shown that a *compact LAF* implies the closure postulate. Below, the relevant definitions and theorems are given.

Definition 5.4.7 (Generation of arguments, Dung and Thang (2014))**.** An argument A is said to be *generated* by a set of arguments S, if there is a base B of A such that $B \subseteq \texttt{Sub}(S)$. The set of all arguments generated by S is denoted by $GN(S)$.

Lemma 5.4.3. For every set of arguments S, $\texttt{Sub}(S) \subseteq GN(S)$.

Theorem 5.4.4. Let E be a complete extension, then $GN(E) = E$.

Note that Lemma 5.4.3 and Theorem 5.4.4 immediately imply the closure under subarguments postulate since for every complete extension E: $\texttt{Sub}(E) \subseteq GN(E) = E$.

Definition 5.4.8 (Compact, Dung and Thang (2014))**.** A logic-associated argumentation framework is *compact* if for each set of arguments S, $\texttt{Conc}(GN(S))$ is closed under strict rules.

Theorem 5.4.5 (Dung and Thang (2014))**.** Each rule-based system is compact.

Theorem 5.4.6 (Dung and Thang (2014))**.** Each compact logic-associated framework satisfies the closure under strict rules postulate.

Dung and Thang (2014) also showed that a *cohesive LAF* implies the consistency postulate. Below, the relevant definition and theorem are given.

Definition 5.4.9 (Cohesive, Dung and Thang (2014))**.** A logic-associated argumentation framework is *cohesive*, if for each inconsistent set of arguments S, $GN(S)$ is conflicting.

Theorem 5.4.7 (Dung and Thang (2014))**.** A cohesive logic-associated argumentation framework satisfies the consistency postulate.

Dung and Thang (2014) have proven that an abstract logic satisfying the *self-contradiction axiom* also implies the satisfaction of the consistency postulate. For the next results they assume a language $\mathcal{L}$ of literals, where a literal is an atom a or the explicit negation $\neg a$ of atom a.

Definition 5.4.10 ($\neg X$, Dung and Thang (2014))**.** For $X \subseteq \mathcal{L}$, denote $\neg X = \{\neg l | l \in X\}$.

Definition 5.4.11 (Self-contradiction axiom, Dung and Thang (2014))**.** The abstract logic AL_{AS} is said to satisfy the *self-contradiction axiom*, if for each minimal inconsistent set $X \subseteq \mathcal{L}$: $\neg X \subseteq CN_{AS}(X)$.

Theorem 5.4.8 (Dung and Thang (2014))**.** Suppose AL_{AS} satisfies the self-contradiction axiom, then AS is cohesive.

Theorem 5.4.9 (Dung and Thang (2014))**.** Suppose AL_{AS} satisfies the self-contradiction axiom. Then AS satisfies the consistency postulate.

The following theorems show that if rule-based argumentation systems are closed under contra-, or transposition, then the closure and consistency postulates are satisfied.

Theorem 5.4.10 (Dung and Thang (2014))**.** Let $AS = (\mathcal{R}_s, \mathcal{R}_d)$ such that the set of strict rules $\mathcal{R}_s$ is closed under transposition. Then AL_{AS} satisfies the self-contradiction axiom.

Theorem 5.4.11 (Dung and Thang (2014))**.** Let $AT = (AS, KB)$ be an argumentation theory that is closed under contraposition. Then AL_{AS} satisfies the self-contradiction axiom.

It is clear that Theorem 5.4.10 and 5.4.11 generalise the results of Modgil and Prakken (2013). In the following section, the results of Dung and Thang (2014) are generalised.

5.4.1 Generalisations of the Results of Dung and Thang

The results of Dung and Thang (2014) can be generalised in some aspects. The most important ones are treated below.

First, the results of Dung and Thang (2014) can be generalised on the assumed language. From Definition 5.4.10, Dung and Thang (2014) assume a language of literals. It is useful to be able to build arguments from a richer logical language. Therefore the definitions and results of Dung and Thang (2014) have to be generalised.

Definition 5.4.12 ($\neg X$)**.** For $X \subseteq \mathcal{L}$, denote $\neg X = \{\neg f | f \text{ is a formula in } X\}$.

Since the proofs of Dung and Thang (2014) do not use the fact that they only have literals in their language, all other theorems, lemmas and definitions of Dung and Thang (2014) stay the same with this new definition for $\neg X$.

As a second generalisation, the self-contradiction axiom can be weakened without losing its strength. The self-contradiction axiom requires that for each minimal inconsistent set $X \in \mathcal{L}$, the negation of every formula in X is in the set $CN_{AS}(X)$. However to prove that AS is cohesive, Dung and Thang (2014) do not use the fact that the negation of *every* formula has to be in this set. They only use the fact that there must be some $\sigma \in X$ such that $\neg\sigma \in CN_{AS}(X)$. Therefore, this axiom can be weakened as follows while it still implies cohesiveness.

Definition 5.4.13 (Weak self-contradiction axiom)**.** The abstract logic AL_{AS} is said to satisfy the *weak self-contradiction axiom*, if for each minimal inconsistent set $X \subseteq \mathcal{L}$ there is a $\sigma \in X$ such that $\neg\sigma \in CN_{AS}(X)$.

Theorem 5.4.12. Suppose AL_{AS} satisfies the weak self-contradiction axiom, then AS is cohesive.

Furthermore, the $ASPIC^+$ framework is richer than the logic-associated argumentation framework described by Dung and Thang (2014), since they do not take preferences on arguments into account. Therefore the results of Dung and Thang (2014) have to be generalised for preferences between arguments. For this, an adapted definition for the logic-associated argumentation framework, that includes preferences, is needed.

Definition 5.4.14 (Preference based logic-associated argumentation framework (PBLAF))**.** A *preference based logic-associated argumentation framework* over a language $\mathcal{L}$ is a tuple $(AFSAF, SAF, \sqsubseteq, AL, Cnl)$, where:

- $AFSAF = (\mathcal{A}, \mathcal{D})$ is an argumentation framework corresponding to the structured argumentation framework $SAF = (\mathcal{A}, att, \preceq)$.

- $AL = (CN, CONTRA)$ is an abstract logic over $\mathcal{L}$.
- $Cnl : \mathcal{A} \longrightarrow \mathcal{L}$ assigns to each argument A, its conclusion $Cnl(A)$ in $\mathcal{L}$.
- $\sqsubseteq$ is a partial order over $\mathcal{A}$, where $A \sqsubseteq B$ means that A is a subargument of B such that for all arguments $C \in \mathcal{A}$, if $(C, A) \in \mathcal{D}$ then $(C, B) \in \mathcal{D}$.

Since the addition of preferences only influences the defeat relation, the results of Dung and Thang (2014) that use the attack relation have to be generalised. The closure and consistency postulates by adding preferences are checked below.

5.4.1.1 Closure Postulates for a $PBLAF$

The results of Dung and Thang (2014) about the closure postulates are not affected by the preferences between arguments, since it does not have requirements on the attack relation that can be violated by preferences. Therefore, it still holds that each rule-based argumentation system is compact. The proof of theorem 5.4.6 of Dung and Thang (2014) does not use any facts about the attack relation, so this result still holds for a compact $PBLAF$. Also Lemma 5.4.3 and Theorem 5.4.4 still hold for the case with preferences. Next, the consistency postulate is checked for $PBLAF$s.

5.4.1.2 Consistency Postulate for a $PBLAF$

The theorems concerning the consistency postulate, that use the attack relation in their proofs, are Theorems 5.4.8 and 5.4.9. The next theorem generalises these results. It shows that there are some requirements on the preference relation in order to allow it to be cohesive. The following lemma of Dung and Thang (2014) is needed.

Lemma 5.4.13 (Dung and Thang (2014))**.** Let A be an argument and BD be the set of all subarguments of A that are basic defeasible arguments, then $\texttt{Conc}(A) \in CN_{AS}(\texttt{Conc}(BD))$.

Theorem 5.4.14. If the preference relation of a $PBLAF = (AFSAF, SAF, \sqsubseteq, AL_{AS})$ is reasonable, AL_{AS} satisfies the self-contradiction axiom and if AT satisfies axiom consistency, then $PBLAF$ is cohesive.

Proof. Take an inconsistent set of arguments S. Hence $\texttt{Conc}(\texttt{Sub}(S))$ is inconsistent. Define BD to be the set of all arguments of $\texttt{Sub}(S)$, which are basic defeasible arguments. BD cannot be empty because otherwise AT would not be axiom consistent. Lemma 5.4.13 implies that $\texttt{Conc}(\texttt{Sub}(S)) \subseteq CN_{AS}(\texttt{Conc}(BD))$, so $CN_{AS}(\texttt{Conc}(\texttt{Sub}(S))) = CN_{AS}(\texttt{Conc}(BD))$. Therefore BD is inconsistent.

AL_{AS} satisfies the self-contradiction axiom so for all $\sigma \in \texttt{Conc}(BD)$ it holds that $\neg\sigma \in CN_{AS}(\texttt{Conc}(BD))$. Let B the weakest argument of BD and with $\texttt{Conc}(B) = \sigma$. Then, because of the self-contradiction axiom, $\neg\sigma \in CN_{AS}(\texttt{Conc}(BD))$. From $CN_{AS}(\texttt{Conc}(\texttt{Sub}(S))) = CN_{AS}(\texttt{Conc}(BD))$, it follows that $\neg\sigma \in CN_{AS}(\texttt{Conc}(\texttt{Sub}(S)))$. From the compactness of AL_{AS} and $\texttt{Sub}(S) \subseteq GN(S)$, it follows that there is an argument $A \in GN(S)$, such that $\texttt{Conc}(A) = \neg\sigma$. The base of A is S, this means that if A is defeated then S must be under defeat. Therefore, it can be concluded that all subarguments of A, which have a defeasible top rule or are just an ordinary premise, are in BD. The fact that B is the weakest argument in BD in combination with a reasonable argument ordering implies that $A \not\prec B$. This means that A defeats argument B and so $GN(S)$ is conflicting. It can be concluded that the $PBLAF$ is cohesive. □

Observe that in this proof, it is not enough for B to be the weakest maximal fallible subargument since it could be the case that this argument contains a weaker subargument B', which is a basic defeasible argument. In case argument $B' \in \texttt{Sub}(A)$, it could be that $A \prec B$ and then it cannot be shown that $GN(S)$ is conflicting. Therefore, B really has to be the weakest basic defeasible subargument.

In Theorem 5.4.14 some conditions under which a $PBLAF$ is cohesive are described. The results of Dung and Thang (2014) state that whenever AL_{AS} satisfies the self-contradiction axiom, then AS is cohesive. The proof uses the fact that AT is closed under trans-, or contraposition or that AT satisfies the self-contradiction axiom. The weak self-contradiction axiom is much weaker and the next example shows that the weak self-contradiction axiom together with axiom consistency and a reasonable argument ordering does not imply that AS is cohesive. Note that the weak self-contradiction axiom is enough in case preferences are not taken into account.

Example 5.4.2. Take the set $AS = (\mathcal{R}_s, \mathcal{R}_d)$ with $\mathcal{R}_s = \{b \rightarrow \neg c\}$, $\mathcal{R}_d = \{a \Rightarrow b\}$, $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \{a, c\}$. It follows that AT is axiom consistent. The following arguments can be constructed $A_1 :\ a$, $A_2 :\ A_1 \Rightarrow b$, $A_3 :\ A_2 \rightarrow \neg c$, $B :\ c$. Take the following reasonable argument ordering: $A_1, A_2, A_3 \prec B$ with A_1, A_2, A_3 having the same preference. It is easily checked that the axiom consistency is satisfied. The weak self-contradiction axiom also holds, because the only minimally inconsistent set is $X = \{b, c\}$ and also $\neg c \in CN_{AS}(X)$.
Now, it has to be shown that AS is not cohesive. Take $S = \{A_2, B\}$ which is clearly inconsistent. $GN(S) = \{A_2, A_3, B\}$, but $GN(S)$ is not conflicting because $A_2, A_3 \prec B$. This means that AS is not cohesive.

Next, the closure and consistency postulates are checked for the paraconsistent logics.

5.5 Exploiting the Results for the Logic W

The results of Dung and Thang (2014) are used as a guidance for proving the closure* and consistency* postulates for the logic W. At first sight, reusing their results would not seem to be possible since $\vdash_W$ does not satisfy idempotence (as shown by the counterexample to the **Cut** rule) and is therefore not an abstract logic. However, upon closer inspection it turns out that the results on $ASPIC^+$ do not use this property at all. It is then left to verify that their results still hold for the adapted notion of closure and its related adapted notion of consistency.

When necessary the definitions of Dung and Thang (2014) are adapted to the new notions of arguments, closure, and consistency and then their results are verified for the adapted definitions.

Definition 5.5.1 (Base* of an argument*, Dung and Thang (2014))**.** Let A be an argument* and BA be a finite set of subarguments* of A. BA is a *base** of A ($BA = \texttt{Base}^\star(A)$) if

- $\texttt{Conc}(A) \in Cl^\star_{\mathcal{R}_s}(\texttt{Conc}(BA))$;
- For each argument* C, C attacks A if and only if C attacks BA.

Definition 5.5.2 (Generation* of arguments*, Dung and Thang (2014))**.** An argument* A is said to be *generated** by a set of arguments* S, if there is a base* B of A such that $B \subseteq \texttt{Sub}(S)$. The set of all arguments generated* by S is denoted by $GN^\star(S)$.

The following lemma follows by definition of $GN^\star(S)$.

Lemma 5.5.1. For every set of arguments S, $\texttt{Sub}(S) \subseteq GN^\star(S)$.

Theorem 5.5.2. Let E be a complete extension, then $GN^\star(E) = E$.

Proof. First note that according to Definition 5.5.2 for each set S of arguments* $\texttt{Sub}(S) \subseteq GN^\star(S)$, therefore $E \subseteq GN^\star(E)$.
Suppose now that an argument* C defeats an argument* $A \in GN^\star(E)$. Let BA be a base of A such that $BA \subseteq \texttt{Sub}(E)$, then C defeats BA. Hence C defeats $\texttt{Sub}(E)$ and so it defeats E. Since E is a complete extension, every defeat against E is counter defeated by E. A is acceptable with respect to E, so $A \in E$. Therefore $GN^\star(E) \subseteq E$. □

Note that Lemma 5.5.1 and Theorem 5.5.2 immediately imply the closure* under subarguments* postule since for every complete extension E: $\texttt{Sub}(E) \subseteq GN^\star(E) = E$.

Theorem 5.5.3. Each $ASPIC^\star$ AF satisfies the closure* under subarguments* postulate.

Definition 5.5.3 (Compact*)**.** An argumentation framework (AF) is *compact**, if for each set of arguments* S, $GN^\star(S)$ is closed* under strict rules. This is equal to $\texttt{Conc}(GN^\star(S)) = Cl^\star_{\mathcal{R}_s}(\texttt{Conc}(GN^\star(S)^\#))$.

The following two theorems can later be combined for proving closure* under subarguments* postulate.

Theorem 5.5.4. Each compact* AF satisfies the closure* under strict rules postulate.

Proof. Let E be a complete extension. The compactness* implies that $GN^\star(E)$ is closed* under strict rules. From Theorem 5.5.2 E is closed* under strict rules. □

Theorem 5.5.5. If the strict rules of an $ASPIC^\star$ AF are instantiated with all valid inferences of the logic W, then AF is compact*.

Proof.
[$\texttt{Conc}(GN^\star(S)) \supseteq Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$]
Let S be a set of arguments* and $\sigma \in Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$. It needs to be shown that $\sigma \in \texttt{Conc}(GN^\star(S))$. Let X be a minimal subset of $\texttt{Conc}(GN^\star(S)^\#)$ such that $\sigma \in Cl^\star_{\mathcal{R}_s}(X)$. Hence there is a strict argument* A_0 over X with conclusion σ. Further let S_X be a minimal set of arguments* from $GN^\star(S)^\#$ s.t. $\texttt{Conc}(S_X) = X$. Let A be the argument* obtained by replacing each leaf in A_0 (viewed as a proof tree) labelled by a literal α from X by an argument* with conclusion α from S_X. Note that this is possible since all arguments* in S_X are basic defeasible arguments* or are just necessary premises. It is obvious that the conclusion of A is σ. It is shown that S_X is a base of A. Suppose B is an argument* defeating A. Since A_0 is a strict argument* over X, B must defeat a basic defeasible subargument* in S_X. Hence B defeats S_X. Thus $A \in GN^\star(S)$. Hence $\sigma \in \texttt{Conc}(GN^\star(S))$.
[$\texttt{Conc}(GN^\star(S)) \subseteq Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$]
Suppose $\sigma \in \texttt{Conc}(GN^\star(S))$, then it has to be shown that $\sigma \in Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$. $\sigma \in \texttt{Conc}(GN^\star(S))$ means that there is an argument* $A \in GN^\star(S)$ with $\texttt{Conc}(A) = \sigma$.
Suppose A is of the form $\Rightarrow \sigma$ or $\sigma \in \mathcal{K}_p$, then $A \in BD^\star(GN^\star(S))$ and thus $A \in GN^\star(S)^\#$.
Suppose A is of the form $\rightarrow \sigma$ or $\sigma \in \mathcal{K}_n$, then $\sigma \in Cl^\star_{R_s}(\emptyset)$ or $\sigma \in NP^\star(GN^\star(S))$ respectively, so $\sigma \in Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$.
Suppose A is of the form $A_1, \ldots, A_n \Rightarrow \sigma$, then $A \in BD^\star(GN^\star(S))$ and thus $A \in GN^\star(S)^\#$.
Finally, suppose A is of the form $A_1, \ldots, A_n \rightarrow \sigma$, then since $A_1, \ldots, A_n$ are basic defeasible arguments* $A_1, \ldots, A_n \in GN^\star(S)^\#$. Therefore $\sigma \in Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$.
It can be concluded that $\sigma \in Cl^\star_{\mathcal{R}_s}(GN^\star(S)^\#)$.
It is proven that the AF is compact*. □

Definition 5.5.4 (Cohesive*)**.** An AF is *cohesive**, if for each inconsistent* set of arguments* S, $GN^\star(S)$ is conflicting.

Theorem 5.5.6. Each cohesive* AF satisfies the consistency* postulate.

Proof. Let E be a complete extension. Suppose E is inconsistent*. From cohesion, it follows that $GN^\star(E)$ is conflicting. Theorem 5.5.2 states that then E must be conflicting. This is a contradiction so E is consistent*. □

The next two definitions are needed for proving cohesiveness.

Definition 5.5.5 (Self-contradiction axiom*)**.** An AF is said to satisfy the *self-contradiction axiom**, if for each minimal inconsistent set $X \subseteq \mathcal{L}$: $\neg X \subseteq Cl^\star_{\mathcal{R}_s}(X)$.

Definition 5.5.6 (Axiom consistent*)**.** An argumentation theory is *axiom consistent** if and only if $Cl^\star_{\mathcal{R}_s}(\mathcal{K}_n)$ is consistent.

Theorem 5.5.7. If a compact*, axiom consistent* AF has a reasonable argument ordering and satisfies the self-contradiction axiom*, then AF is cohesive*.

Proof. Let S be an inconsistent* set of arguments* and take a minimal inconsistent* subset S' of $\texttt{Sub}(S)$. Definition 4.3.5 combined with axiom consistency* and the minimality of S' causes that $S' \neq \emptyset$ and only contains basic defeasible arguments* or necessary premises. Remark that S' cannot consist of only necessary premises, because of axiom consistency*. Further note that $\texttt{Conc}(S')$ is a minimal inconsistent set. Since AF satisfies the self-contradiction axiom*, for all $\sigma \in \texttt{Conc}(S')$ it holds that $\neg\sigma \in Cl^\star_{\mathcal{R}_s}(\texttt{Conc}(S'))$. Let B be the weakest argument of S' with $\texttt{Conc}(B) = \sigma$. Note that B cannot be a necessary premise because of the reasonable argument ordering and the fact that S' must contain basic defeasible arguments*. By construction of S' it holds that $S' \subseteq GN^\star(S')^\#$. Therefore $\neg\sigma \in Cl^\star_{\mathcal{R}_s}(\texttt{Conc}(GN^\star(S')^\#))$. Because of the compactness of AF it follows that $\neg\sigma \in \texttt{Conc}(GN^\star(S'))$. Therefore, there is an argument* $A \in GN^\star(S')$ such that $\texttt{Conc}(A) = \neg\sigma$. Hence A attacks B. The base of A is S', so it can be concluded that all basic defeasible subarguments* of A are in S'. B is the weakest argument of S', so because of the reasonable argument ordering and the fact that all basic defeasible subarguments* of A are in S' implies that $A \not\prec B$. This means that A defeats B. Since $B \in S' \subseteq GN^\star(S') \subseteq GN^\star(S)$, $GN^\star(S)$ is conflicting. Therefore, AF is cohesive*. □

Theorem 5.5.8. If the strict rules of an *ASPIC** AF are instantiated with all valid inferences of the logic W, then the AF satisfies the self-contradiction axiom*.

Proof. It has to be proved that for every minimal inconsistent* set $X \subseteq \mathcal{L}$ it holds that for each $\sigma \in X$, $\neg\sigma \in Cl^\star_{\mathcal{R}_s}(X)$. Let X be a minimally inconsistent* set and take $S = X \backslash \sigma$. Note that S is a maximal consistent* subset of X and that $S, \sigma \vdash \bot$ (where

$\vdash$ denotes classical entailment). By the deduction theorem for classical logic $S \vdash \sigma \supset \bot$, which implies $S \vdash \neg\sigma$. Since S is a maximal consistent* subset of X, $X \vdash_W \neg\sigma$. This holds for every $\sigma \in X$, so $\neg\sigma \in Cl^\star_{\mathcal{R}_s}(X)$. It can be concluded that AF satisfies the self-contradiction axiom*. □

Combining Theorem 5.5.4, 5.5.5, 5.5.6, 5.5.7 and 5.5.8 results in the following important conclusion, as depicted in Theorem 5.5.9.

Theorem 5.5.9. If the strict rules of an $ASPIC^\star$ AF are instantiated with all valid inferences of the logic W and if AF is axiom consistent* and has a reasonable argument ordering, then AF satisfies the closure* and consistency* postulates.

It can be concluded that the $ASPIC^+$ framework has been combined with the logic W (Rescher and Manor (1970)) in a way that preserves known results on consistency and strict closure while preventing trivialisation in case of rebutting arguments. In order to obtain these results, the $ASPIC^+$ framework had to be adapted by prohibiting chaining of strict rules. This resulted in the $ASPIC^\star$ framework. The next chapter continues with the efficiency problem discussed in Chapter 3.

Chapter 6

Solving the Efficiency Problem

In this chapter, it is investigated if E is an extension of a SAF, all minimal arguments of E also form an extension of the same type of all minimal arguments of the SAF. Furthermore, it is examined if in case only minimal arguments are permitted and E is an extension, E is also a subset of an extension of the same type in case this restriction is lifted. This means that the conclusions that can be drawn from an argumentation framework are not affected in case arguments are required to be minimal. These research questions are researched for the $ASPIC^+$ framework as well as the $ASPIC^\star$ framework. In the last section of this chapter, the results are compared with the results for the efficiency problem for the ABA framework of Dung et al. (2010).

6.1 Minimal Arguments for the $ASPIC^+$ Framework

First, the above described *minimal argument* is formally defined.

Definition 6.1.1 (Minimal argument)**.** A *minimal argument* A on the basis of a knowledge base $KB = (\mathcal{K}, \preceq)$ in an argumentation system $(\mathcal{L}, \mathcal{R}, n, \preceq')$:

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

2. $A_1, \dots, A_n \rightarrow \psi$ if $A_1, \dots, A_n$ are minimal arguments such that there exists a strict rule $\texttt{Conc}(A_1), \dots, \texttt{Conc}(A_n) \rightarrow \psi$ in $\mathcal{R}_s$ and there is not a strict rule $a_1, \dots, a_i \rightarrow \psi$ for $\{a_1, \dots, a_i\} \subset \texttt{Conc}(\{A_1, \dots, A_n\})$.

$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
$\texttt{Conc}(A) = \psi$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
$\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$,
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \to \psi$.

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

Each of these functions defined above can also be used on a set of arguments.

Definition 6.1.2 (A^- and S^-)**.** For an argument A, A^- is a minimal argument corresponding to A. A^- is constructed in the following inductive way. For an argument A, A^- is as follows:

- If $A \in \mathcal{K}$, then $A^- = A$.
- If A is of the form $A_1, \ldots, A_n \to \varphi$, then $A^- = A_i^-, \ldots, A_j^- \to \varphi$ for a minimal subset $\texttt{Conc}(\{A_i, \ldots, A_j\}) \subseteq \texttt{Conc}(\{A_1, \ldots, A_n\})$ such that $\texttt{Conc}(A_i), \ldots, \texttt{Conc}(A_j) \to \varphi \in \mathcal{R}_s$.
- If A is of the form $A_1, \ldots, A_n \Rightarrow \varphi$, then $A^- = A_1^-, \ldots, A_n^- \Rightarrow \varphi$.

For a set of arguments S, define S^- as all minimal arguments of S.

Note that A^- is not guaranteed to be unique. There are exceptions like for argument $A : p \wedge q, q \wedge p \to p$. This argument has to minimal variants, namely $A_1 : p \wedge q \to p$ and $A_2 : q \wedge p \to p$.

Definition 6.1.3 (Minimal SAF, SAF^-)**.** For a $SAF = (\mathcal{A}, att, \preceq)$, let SAF^- be the *minimal* SAF with $SAF^- = (\mathcal{A}^-, att^-, \preceq^-)$. Where att^- is defined as $att \cap (\mathcal{A}^- \times \mathcal{A}^-)$ and $\preceq^- = \preceq \cap (\mathcal{A}^- \times \mathcal{A}^-)$.

Definition 6.1.4 (Extended argument, A^+)**.** Suppose A is an argument, then A^+ is an *extended argument* of A. A^+ is defined as follows:

- $A \in \mathcal{K}$, then $A^+ = A$.

- If A is of the form $A_1, \dots, A_n \rightarrow \varphi$, then $A^+ = A'_1, \dots, A'_m \rightarrow \varphi$ such that there are arguments $A_1^+, \dots, A_n^+$ such that $\{A_1^+, \dots, A_n^+\} \subseteq \{A'_1, \dots, A'_m\}$ and there is a strict rule $\texttt{Conc}(A'_1), \dots, \texttt{Conc}(A'_m) \rightarrow \varphi$.

- If A is of the form $A_1, \dots, A_n \Rightarrow \varphi$, then $A^+ = A'_1, \dots, A'_n \Rightarrow \varphi$ such that there are arguments $A_1^+, \dots, A_n^+$ such that $\{A_1^+, \dots, A_n^+\} = \{A'_1, \dots, A'_n\}$.

Note that A is also an A^+ and that A is an A^{-+}. In general, A^+ is not unique.

The following example clarifies the definitions given above.

Example 6.1.1. Take a SAF with the following arguments:
$A_1 : p$ $\quad A_2 : p \rightarrow q$
$A_3 : r$ $\quad A_4 : p, r \rightarrow q$
Then A_1, A_2 and A_3 are minimal arguments, so SAF^- contains these three arguments. A_2 is the minimal argument corresponding to A_4, so $A_2 = A_4^-$. Now, it is also easy to see that indeed A_4 is an A_2^+. Furthermore A_4 is one example of an A_2^+.

For some results that follow, it is needed that for any argument A, no A^+ can be stronger than A and A^- cannot be weaker than A. Note that this is not implied by the current definition of a reasonable argument ordering. This is illustrated with an example.

Example 6.1.2. Take a look at Example 6.1.1 and take the following argument ordering; A_1 is equally strong as A_2; A_3 and A_4 are also equally strong. Further suppose that $A_2 \prec A_4$. Then the addition of A_3 to A_2 makes it stronger. However, this argument ordering satisfies all properties of a reasonable argument ordering.

Therefore, a different definition is needed instead of the *reasonable argument ordering.*

Definition 6.1.5 (Tolerable argument ordering)**.** $\preceq$ is a *tolerable argument ordering* if and only if:

- For every A^+ of A, $A^+ \preceq A$;

- For any A^- of A, $A^- \succeq A$.

Two examples of tolerable argument orderings are the weakest-link and last-link ordering which we have already seen in Chapter 2 (Definition 2.2.11 and Definition 2.2.12).

The following lemma states that the subarguments of an extended version A^+ are extended versions of the subarguments of A.

Lemma 6.1.1. For any argument A and any extended argument A^+ the following holds: for any $A' \in \texttt{Sub}(A)$ there is an argument $A'' \in \texttt{Sub}(A^+)$ such that $A'' = A'^+$.

Proof. This proof is a proof by induction on the height of argument A (viewed as a proof tree). Suppose A is an element of $\mathcal{K}$ (so the height is 1), then A' and A^+ have to be equal to A. This means that A'' is also equal to A and it is easy to see that $A'' = A'^+$. Suppose that the lemma holds for all arguments of height i for an $i \in \{1, 2, \ldots\}$. Now it has to be proven for arguments of height $i + 1$. Take an arbitrary argument A of height $i + 1$ and take a subargument A' of A. Note that A cannot be an element of $\mathcal{K}$ since the height is greater than 1. Therefore, A has to be of the form $A_1, \ldots, A_n \rightarrow / \Rightarrow \varphi$. Then there are two possibilities: either (i) A' is a subargument of one of the arguments $A_1, \ldots, A_n$, or (ii) A' is equal to A.
(i). There is a $j \in \{1, \ldots, n\}$ such that A' is a subargument of A_j. A_j has a height of i, so there must be an $A'' \in \texttt{Sub}(A_j^+)$ such that $A'' = A'^+$. According to Definition 6.1.4, A'' is a subargument of A^+.
(ii). A' is equal to A. Take A'' to be A^+, then it follows that $A'' = A'^+$ and $A'' \in \texttt{Sub}(A^+)$.
Now it is proved that for every argument the lemma holds. □

This lemma is clarified with the example below.

Example 6.1.3. Take the following arguments:

$A_1 : p \rightarrow p \vee q$	$A_2 : A_1 \Rightarrow s$
$B_1 : p, r \rightarrow p \vee q$	$B_2 : B_1 \Rightarrow s$

Then it is obvious that B_2 is an A_2^+. Lemma 6.1.1 states that for every subargument A' of A_2 there is a subargument B' of A_2^+ such that $B' = A'^+$. For example, take A' to be A_1. Then B_1 is the subargument of B_2 such that $B_1 = A_1^+$.

The preceding result is needed for proving that if argument A attacks/defeats B, then the minimal argument corresponding to A attacks/defeats every extended version of B.

Lemma 6.1.2. If $\preceq$ is a tolerable argument ordering and if argument A defeats/attacks B, then any A^- defeats/attacks every B^+.

Proof.

Attack

(i). Suppose that A undercuts B, so $\texttt{Conc}(A) = \neg n(r)$ for a defeasible top rule r of a $B' \subseteq \texttt{Sub}(B)$. By definition of A^-, the conclusion for every A^- is the same as for A. By Lemma 6.1.1 it holds that for every B^+, there is a subargument $B'' \in \texttt{Sub}(B^+)$ such that $B'' = B'^+$. Note that B' and B'' both have the same defeasible top rule, since by definition of B^+, only the strict rules can mutate. Then it follows that every A^- undercuts B^+ on B''.
(ii). Suppose now that A undermines B, so $\neg\texttt{Conc}(A)$ is the ordinary premise B'. Every B^+ also has this ordinary premise. By definition of A^- (Definition 6.1.2), the conclusion

of every A^- is the same as for A. Therefore any A^- undermines every B^+ on B'.
(iii). Suppose that A rebuts B, so $\neg\texttt{Conc}(A)$ is the conclusion of some basic defeasible argument $B' \in \texttt{Sub}(B)$. The conclusion for every A^- is the same as for A. By Lemma 6.1.1, it holds that for every B^+, there is a subargument $B'' \in \texttt{Sub}(B^+)$ such that $B'' = B'^+$. Note that the conclusions of B' and B'' are the same and that B'' also has a defeasible top rule. Therefore any A^- rebuts B^+ on B''.

Defeat

Suppose argument A defeats B. This means that A attacks B on a subargument B'. Then (i) A undercuts B, or (ii) $A \not\prec B'$.
(i). Suppose that A undercuts B, then it follows that every A^- undercuts B^+ on B'' (see reasoning above for attack (i)). This implies that any A^- defeats every B^+.
(ii). Otherwise it has to be the case that $A \not\prec B'$. Note that, because of the tolerable argument ordering, $A^- \succeq A$. In the proof for attack it is shown that every A^- attacks every B^+ on B'', where $B'' = B'^+$. The tolerable argument ordering causes $B'' \preceq B'$. Therefore for every A^- and B'' it holds that $A^- \not\prec B''$, so every A^- defeats every B^+. □

The following lemma immediately follows from Lemma 6.1.2.

Lemma 6.1.3. If $\preceq$ is a tolerable argument ordering, then for all complete extensions E:

1. If $A \in E$, then $A^- \in E$ for every A^-;
2. If $B \notin E$, then $B^+ \notin E$ for all B^+.

Proof. (1). Suppose $A \in E$ and let B be an argument defeating an A^-. Then by Lemma 6.1.2, B defeats A. Therefore, there must be an argument $C \in E$ such that C defeats B. Hence, A^- is acceptable with respect to E and thus $A^- \in E$.
(2). Suppose $B \notin E$. Then there exists an argument A such that A defeats B and there does not exist a $C \in E$ that defeats A. According to Lemma 6.1.2, A defeats every B^+, so B^+ is not acceptable with respect to E and hence $B^+ \notin E$. □

The following lemma states that the function F is a monotonic bijection from all complete extensions of a SAF^- onto all complete extensions of SAF. This lemma is based on the results of Dung et al. (2010) for the ABA framework. Their results are discussed in Section 6.3.

Lemma 6.1.4. Let $SAF^- = (\mathcal{A}^-, att^-, \preceq^-)$ be the minimal structured argumentation framework corresponding to $SAF = (\mathcal{A}, att, \preceq)$. Let $\preceq$ be a tolerable argument ordering. Also, let $\mathcal{C}$ and $\mathcal{C}^-$ be the sets of complete extensions of SAF and SAF^- respectively. $F(X)$ is a monotonic (with respect to set inclusion) bijection from $\mathcal{C}^-$ onto $\mathcal{C}$ such that:

1. For each $E \in \mathcal{C}^- : F(E)^- = E$.

2. For each $E \in \mathcal{C} : F(E^-) = E$ and $E^- \in \mathcal{C}^-$.

Proof. Take two sets of arguments X and Y such that $X \subseteq Y$. Suppose $F(Y) \subset F(X)$, then there must be an argument A that is acceptable with respect to X but not to Y. Since $X \subseteq Y$, A has to be acceptable with respect to Y. Therefore, the monotonicity of $F(E)$ with respect to set inclusion is obvious.

(1). It will be shown that $F(E)$ is indeed a function from $\mathcal{C}^-$ into $\mathcal{C}$ such that $F(E)^- = E$ by showing that $F(E)$ is a complete extension in SAF, if E is a complete extension in SAF^-. Let E be a complete extension in SAF^-. First it has to be shown that E is an admissable set in SAF.

E is conflict-free in $\mathcal{A}^-$, so it has to be conflict-free in $\mathcal{A}$. It also defeats each minimal argument defeating E and it contains every minimal argument acceptable with respect to E.

For any $A \in E$ and any $B \in \mathcal{A}$ that defeats A, take a B^-. Then $B^- \in \mathcal{A}^-$ defeats A by Lemma 6.1.2, so some $C \in E$ defeats B^-. But then C also defeats B by Lemma 6.1.2 combined with the fact that B is a B^{-+}. This means that $A \in \mathcal{A}$ is acceptable with respect to E. It can be concluded that E defeats every argument that defeats E. It was already shown that E is conflict-free, so $E \subseteq \mathcal{A}$ is an admissible set.

For each argument $A \in F(E)$, any A^- has to be acceptable with respect to E in $\mathcal{A}^-$, so every A^- is in E. Therefore, $(F(E) - E)^- = \emptyset$, thus $F(E)^- = E$.

Now, it has to be shown that $F(E)$ is a complete extension. Let A be acceptable with respect to $F(E)$ and let B defeat A. Hence, there is a $C \in F(E)$ defeating B. $C \in F(E)$ means that C is acceptable with respect to E, thus E defeats every argument defeating C. Suppose an argument D defeats C^-, then D also defeats C by Lemma 6.1.2. Therefore, E must defeat D, so C^- is acceptable with respect to E. Since E is a complete extension of SAF^-, any C^- is in E. C^- defeats B (Lemma 6.1.2). Therefore, E defeats B. Thus, A is acceptable with respect to E, and therefore $A \in F(E)$. As a consequence, $F(E)$ is complete.

(2). Let E be a complete extension in SAF. It will be shown that E^- is complete in SAF^-. First it has to be shown that E^- is an admissible set of $\mathcal{A}^-$. Since E is conflict-free, E^- has to be conflict-free as well.

From Lemma 6.1.3 and the fact that E is complete in SAF, each minimal version of the arguments in E belongs to E. Let $A \in \mathcal{A}^-$ defeat E^-. Hence, there is $B \in E$ defeating A. According to Lemma 6.1.3 $B^- \in E$, so $B^- \in E^-$. Hence, B^- defeats A (Lemma 6.1.2). Thus, E^- is admissible.

Each minimal argument acceptable with respect to E^- is acceptable with respect to E and hence belongs to E and so to E^-. E^- is therefore complete.

Since $E^- \subseteq E$ and E is complete, it is clear that $F(E^-) \subseteq F(E) = E$. It is now shown that each argument acceptable with respect to E is also acceptable with respect to E^-.

Let A be an argument acceptable with respect to E in SAF and let B be an argument defeating A. Hence, there is an argument $C \in E$ defeating B and so each $C^- \in E$ defeats B. Hence E^- defeats B. Thus A is acceptable with respect to E^- in SAF. It can be concluded that $F(E^-) \supseteq F(E) = E$, i.e. $F(E^-) = E$.
[Bijective]
[Injective] Take $X, Y \in \mathcal{C}^-$ such that $F(X) = F(Y)$. It is obvious that $F(X)^- = F(Y)^-$. Then according to the proof for point (1) $F(X)^- = X$ and $F(Y)^- = Y$. Therefore, it follows that $X = Y$.
[Surjective] It has to be shown that for all $Y \in \mathcal{C}$ there is an $X \in \mathcal{C}^-$ such that $F(X) = Y$. Now take X to be Y^-, then the proof for point (2) provides that $X \in \mathcal{C}^-$ and that $F(X) = Y$.
Injectivity and surjectivity provides that F is a bijection from $\mathcal{C}^-$ onto $\mathcal{C}$. □

Below, the main result is given for the conclusions that can be drawn from SAFs and SAF^-s. This theorem is based on the results of Dung et al. (2010) for the ABA framework.

Theorem 6.1.5. Let $SAF^- = (\mathcal{A}^-, att^-, \preceq^-)$ be the minimal structured argumentation framework corresponding to $SAF = (\mathcal{A}, att, \preceq)$. Let $\preceq$ be a tolerable argument ordering. Take $T \in \{$complete, grounded, preferred, stable$\}$, then:

1. Let E be a T extension in SAF, then E^- is a T extension in SAF^-.
2. Let E be a T extension in SAF^-, then $F(E)$ is a T extension in SAF.

Proof.
[$T =$ complete]
Let $\mathcal{C}$ and $\mathcal{C}^-$ be the sets of complete extensions of SAF and SAF^-, respectively. Lemma 6.1.4 states that F is a bijection from $\mathcal{C}^-$ to $\mathcal{C}$. This immediately provides that $F(E)\mathcal{C}$. The second point of Lemma 6.1.4 states that $E^- \in \mathcal{C}^-$.

[$T \in \{$grounded, preferred$\}$]
From Lemma 6.1.4, it follows immediately that for each $E \in \mathcal{C}^-$, $F(E)$ is minimal or maximal with respect to set inclusion in E if and only if E is minimal or maximal respectively in $\mathcal{C}^-$. Hence E is grounded or preferred in SAF^- if and only if $F(E)$ is grounded or preferred in SAF, respectively.

[$T =$ stable]
(1). Take E to be a stable extension in SAF. Suppose for contradiction that E^- is not a stable extension in SAF^-. Then there must be an argument $A \in \mathcal{A}^-$, $A \notin E^-$ such that A is not defeated by any argument in E^-. However, $A \notin E^-$ implies $A \notin E$ since A is minimal. E is stable, thus there must be an argument $B \in E$ such that B defeats

A. But then any $B^- \in E$ (Lemma 6.1.3) defeats A. It is clear that $B \in E^-$, so this is a contradiction with the fact that E^- does not defeat A. Therefore, E^- is a stable extension in SAF^-.
(2). Take E to be a stable extension of SAF^-. Suppose for contradiction that $F(E)$ is not a stable extension of SAF. Then there must be an argument $A \in \mathcal{A}$ such that $A \notin F(E)$ and $F(E)$ does not defeat A. $A \notin F(E)$ means that A is not acceptable with respect to E. Therefore, there must be an argument $B \in \mathcal{A}$ that defeats A such that there is not a $C \in E$ that defeats B. Since E is stable in SAF^-, $B \in E$ so E defeats A. This implies that $F(E)$ defeats A, which is a contradiction with the fact that $F(E)$ does not defeat A. Therefore, $F(E)$ has to be a stable extension of SAF. □

It can be concluded, by combining all these results, that the conclusions that can be drawn from an argumentation framework are not affected in case arguments are required to be minimal.

6.2 Minimal Arguments⋆ for the *ASPIC*⋆ Framework

The same results are discussed for the *ASPIC*⋆ framework. Below, the needed definitions for minimal arguments⋆ are defined.

Definition 6.2.1 (Minimal argument⋆)**.** A *minimal argument*⋆ A on the basis of a knowledge base $KB = (\mathcal{K}, \preceq)$ in an argumentation system $(\mathcal{L}, \mathcal{R}, n, \preceq')$:

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

2. $A_1, \ldots, A_n \rightarrow \psi$ if $A_1, \ldots, A_n$ are minimal arguments⋆ with a defeasible top rule or are from $\mathcal{K}$ and such that there exists a strict rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$ in $\mathcal{R}_s$ and there is not a strict rule $a_1, \ldots, a_i \rightarrow \psi$ for $\{a_1, \ldots, a_i\} \subset \texttt{Conc}(\{A_1, \ldots, A_n\})$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
   $\texttt{Conc}(A) = \psi$,
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
   $\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$,
   $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$.

3. $A_1, \ldots, A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments⋆ such that there exists a defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$ in $\mathcal{R}_d$.

$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
$\texttt{Conc}(A) = \psi$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
$\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi\}$,
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$.

Definition 6.2.2 (A^- and S^-)**.** For an A, A^- is a minimal argument* corresponding to A. A^- is constructed in the following inductive way. For an argument* A, A^- is as follows:

- If $A \in \mathcal{K}$, then $A^- = A$.
- If A is of the form $A_1, \ldots, A_n \rightarrow \varphi$, then $A^- = A_i^-, \ldots, A_j^- \rightarrow \varphi$ for a smallest subset $\texttt{Conc}(\{A_i, \ldots, A_j\}) \subseteq \texttt{Conc}(\{A_1, \ldots, A_n\})$ such that $\texttt{Conc}(A_i), \ldots, \texttt{Conc}(A_j) \rightarrow \varphi \in \mathcal{R}_s$.
- If A is of the form $A_1, \ldots, A_n \Rightarrow \varphi$, then $A^- = A_1^-, \ldots, A_n^- \Rightarrow \varphi$.

For a set of arguments* S, define S^- as all minimal arguments* of S.

Note that A^- is not guaranteed to be unique.

Obviously, the following structured argumentation frameworks are *ASPIC** frameworks that only contain arguments*.

Definition 6.2.3 (Minimal SAF, SAF^-)**.** For a $SAF = (\mathcal{A}, att, \preceq)$, let SAF^- be the *minimal* SAF with $SAF^- = (\mathcal{A}^-, att^-, \preceq^-)$. Where att^- is defined as $att \cap (\mathcal{A}^- \times \mathcal{A}^-)$ and $\preceq^- = \preceq \cap (\mathcal{A}^- \times \mathcal{A}^-)$.

Definition 6.2.4 (Extended argument*, A^+)**.** Suppose A is an argument*, then A^+ is an *extended argument** of A. A^+ is defined as follows:

- $A \in \mathcal{K}$, then $A^+ = A$.
- If A is of the form $A_1, \ldots, A_n \rightarrow \varphi$, then $A^+ = A'_1, \ldots, A'_m \rightarrow \varphi$ such that there are arguments* $A_1^+, \ldots, A_n^+$ such that $\{A_1^+, \ldots, A_n^+\} \subseteq \{A'_1, \ldots, A'_m\}$ and there is a strict rule $\texttt{Conc}(A'_1), \ldots, \texttt{Conc}(A'_m) \rightarrow \varphi$.
- If A is of the form $A_1, \ldots, A_n \Rightarrow \varphi$, then $A^+ = A'_1, \ldots, A'_n \Rightarrow \varphi$ such that there are arguments* $A_1^+, \ldots, A_n^+$ such that $\{A_1^+, \ldots, A_n^+\} = \{A'_1, \ldots, A'_n\}$.

Note that A is also an A^+ and that A is an A^{-+}. In general, A^+ is not unique.

For some results that follow, it is needed to have that for an argument* A, A^+ cannot be stronger than A and that A^- cannot be weaker than A. This is not implied by the current definition of a reasonable argument* ordering. A different definition of a *reasonable argument* ordering* is needed.

Definition 6.2.5 (Tolerable argument* ordering)**.** $\preceq$ is a *tolerable argument* ordering* if and only if:

- For every A^+ of A, $A^+ \preceq A$;
- For any A^- of A, $A^- \succeq A$.

The next lemmas are all needed for proving the equivalence of the conclusions that are drawn from minimal and non-minimal structured argumentation frameworks.

Lemma 6.2.1. For any argument* A and any extended argument* A^+ the following holds: for any $A' \in \texttt{Sub}(A)$ there is an argument* $A'' \in \texttt{Sub}(A^+)$ such that $A'' = A'^+$.

Proof. The proof of Lemma 6.1.1 does not use any properties of arguments that arguments* do not possess. Therefore, the proof of Lemma 6.1.1 also holds for this lemma in case the word argument is replaced by argument*. The used lemmas and definitions are the versions of this section. □

Lemma 6.2.2. If $\preceq$ is a tolerable argument* ordering and argument* A defeats/attacks B, then every A^- defeats/attacks every B^+.

Proof. This proof is exactly the same as for the *ASPIC*$^+$ framework except that the word argument has to be replaced by argument* and that the used lemmas and definitions are the versions of this section. □

Lemma 6.2.3. If $\preceq$ is a tolerable argument* ordering, then for all complete extensions E:

1. If $A \in E$, then $A^- \in E$ for all A^-;
2. If $B \notin E$, then $B^+ \notin E$ for all B^+.

Proof. This proof is also exactly the same as for the *ASPIC*$^+$ framework, except that the word argument has to be replaced by argument* and that the used lemmas and definitions are the versions of this section. □

The following lemma states that the function F is a monotonic bijection from all complete extensions of a SAF^- onto all complete extensions of SAF. This lemma is based on the results of Dung et al. (2010) for the ABA framework. Their results are discussed in the next section.

Lemma 6.2.4. Let $SAF^- = (\mathcal{A}^-, att^-, \preceq^-)$ be the minimal structured argumentation framework corresponding to $SAF = (\mathcal{A}, att, \preceq)$ (for the $ASPIC^\star$ framework). Let $\preceq$ be a tolerable argument* ordering. Also, let $\mathcal{C}$ and $\mathcal{C}^-$ be the sets of complete extensions of SAF and SAF^- respectively. $F(X)$ is a monotonic (with respect to set inclusion) bijection from $\mathcal{C}^-$ onto $\mathcal{C}$ such that:

1. For each $E \in \mathcal{C}^- : F(E)^- = E$.
2. For each $E \in \mathcal{C} : F(E^-) = E$ and $E^- \in \mathcal{C}^-$.

Proof. The proof for Lemma 6.1.4 can be used since no properties of arguments are used that arguments* do not possess. The used lemmas and definitions are the versions of this section. □

Below, the main result is given for the conclusions that can be drawn from SAFs and SAF^-s for the $ASPIC^\star$ framework. This theorem is based on the results of Dung et al. (2010) for the ABA framework.

Theorem 6.2.5. Let $SAF^- = (\mathcal{A}^-, att^-, \preceq^-)$ be the minimal structured argumentation framework corresponding to $SAF = (\mathcal{A}, att, \preceq)$ (for the $ASPIC^\star$ framework). Let $\preceq$ be a tolerable argument* ordering. Take $T \in$ {complete, grounded, preferred, stable}, then:

1. Let E be a T extension in SAF, then E^- is a T extension in SAF^-.
2. Let E be a T extension in SAF^-, then $F(E)$ is a T extension in SAF.

Proof. The proof is exactly the same as for Theorem 6.1.5 except that the word argument has to be replaced by argument* and the used lemmas and definitions are the versions of this section. □

6.3 Comparison to the Results of Dung, Toni and Mancarella

Dung et al. (2010) solved the efficiency problem for the ABA framework. This is a special case of the $ASPIC^+$ framework, without preferences and defeasible rules. Dung et al. (2010) define a *non redundant argument*, which is more general then the *minimal* *argument* defined in this thesis. A non redundant argument is defined in terms of a *less redundant relation*. Below, the relevant definitions of Dung et al. (2010) are given.

Definition 6.3.1 (Less redundant relation $\prec$, Dung et al. (2010))**.** If a relation $\prec$ represents a *less redundant relation* between arguments, then the following properties should be satisfied:

1. $\prec$ is transitive (i.e., for arguments A, B, C, if $A \prec B$ and $B \prec C$, then $A \prec C$);
   reflexive (i.e. for any argument A, $A \prec A$);
   antisymmetric (i.e. for arguments A, B, if $A \prec B$ and $B \prec A$, then $A = B$).
2. given arguments A, B, C, if $A \prec B$ then
   (a) if B attacks C, then A attacks C;
   (b) if C attacks A, then C attacks B.
3. for each argument A there is an argument B such that
   (a) there is no other argument C such that $C \prec B$;
   (b) either $A = B$ or $B \prec A$.

Definition 6.3.2 ((Non) redundant argument, Dung et al. (2010))**.** Given a 'less redundant' relation $\prec$, an argument is *redundant* (with respect to $\prec$), if there exists an argument $B \neq A$ such that $B \prec A$. The set NR of all *non redundant arguments* (with respect to $\prec$) is such that for each argument A there is a $B \in NR$ such that $B \prec A$.

Dung et al. (2010) prove that the same conclusions are obtained in case non redundant (according to a certain less redundant relation) abstract argumentation frameworks are used. The following theorem contains this result.

Definition 6.3.3 ($\prec$-trimmed abstract argumentation framework, Dung et al. (2010))**.** Let $(\mathcal{A}, att)$ be an abstract argumentation framework. Let $\prec$ be a less redundant relation between arguments in $\mathcal{A}$ and $NR \subseteq \mathcal{A}$ be the set of all non redundant arguments with respect to $\prec$. Further, let $att_{NR} = att \cap (NR \times NR)$. Then (NR, att_{NR}) be the $\prec$-*trimmed* version of $(\mathcal{A}, att)$.

Theorem 6.3.1. Let (NR, att_{NR}) be the $\prec$-*trimmed* version of $(\mathcal{A}, att)$. Then

1. Let X be a complete, preferred, or grounded extension in $(\mathcal{A}, att)$. Then $X \cap NR$ is a complete, preferred, or grounded extension, respectively, in (NR, att_{NR}).
2. Let X be a complete, preferred, or grounded extension in (NR, att_{NR}). Then $F(X)$ is a complete, preferred, or grounded extension, respectively, in $(\mathcal{A}, att)$.

Note that Theorem 6.3.1 does not include a proof for stable semantics.

In order to compare the results of Dung et al. (2010) with the results of Section 6.1 and 6.2, one can define a relation according to the concept of extended arguments.

Definition 6.3.4 (Smaller relation $\preccurlyeq$)**.** Let $\preccurlyeq$ be a *smaller relation*, then for argument A, B, argument $A \preccurlyeq B$ if and only if $B = A^+$.

Definition 6.1.4 and Lemma 6.1.2 are a motivation for the following hypothesis.

Hypothesis 6.3.1. The smaller relation is a less redundant relation.

This is a suggestion for future research (see Chapter 8). This hypothesis will generalise the results since one is free te define other less redundant relations.

This thesis generalises the results of Dung et al. (2010) with regard to the $ASPIC^+$ and $ASPIC^\star$ framework, which include preferences and defeasible rules, and it extends the theorem by including stable semantics. The results of Dung et al. (2010) are more general with respect to the less redundant relation; they do not specify the less redundant relation. This relation is more general than the concept of minimal arguments.

The results of this chapter can be generalised by changing the concept of minimal arguments to the less redundant relation. The reason why this can be done, is that the proofs for SAFs only use the properties of a less redundant relation and of the tolerable argument(*) ordering.

The following chapter is about the comparison between the extensions in case conflict-free is defined in terms of attack or defeat.

Chapter 7

Comparison of Defeat and Attack Conflict-Free

Up to now, a conflict-free set of arguments(*) meant that there is no pair of arguments(*) such that one defeats the other. In this chapter, conflict-free is also noted as *defeat conflict-free*. Conflict-free could also be defined such that there is no pair of arguments(*) of which one attacks the other, i.e. *attack conflict-free*. One could question whether it makes sense to say that a set is 'conflict-free' whenever there is a pair of arguments of which one attacks the other. Is it a better idea to use attack conflict-freeness to determine the admissible sets of arguments? Modgil and Prakken (2013) proved for the $ASPIC^+$ framework that the extensions for both conflict-free definitions are the same provided that all arguments are generated. In this chapter, the same is proven for the $ASPIC^\star$ framework, for which the strict rules are instantiated with the logic W, and for minimal structured argumentation frameworks for the $ASPIC^+$ framework.

First some definitions are introduced for the $ASPIC^\star$ framework.

Definition 7.0.5 (Attack conflict-free)**.** A set of arguments(*) S is *attack conflict-free*, if there is not a pair of arguments(*) in S of which one attacks the other.

Definition 7.0.6 (Attack-admissible)**.** A set of arguments(*) S is said to be *attack-admissible*, if S is attack conflict-free and it defeats each argument(*) defeating S.

Definition 7.0.7 (Attack-T extension for $T \in$ {complete, grounded, preferred, stable})**.** An *attack-T extension* E is a T extension in case admissibility is defined in terms of attack-admissible.

The result of Modgil and Prakken (2013):

Theorem 7.0.2 (Modgil and Prakken (2013))**.** If AT is an argumentation theory that is either closed under trans- or contraposition, if all arguments are generated and if the argument ordering of a SAF is reasonable, then E is a T extension of the SAF if and only if E is an attack-T extension of the SAF, for $T \in \{\text{complete, grounded, preferred, stable}\}$.

The following theorem states that the extensions of $ASPIC^\star$ frameworks are the same for both conflict-free definitions in case some conditions are provided.

Theorem 7.0.3. If the strict rules of a minimal $ASPIC^\star$ SAF are instantiated with the logic W and if SAF is axiom consistent* and the argument* ordering is reasonable and all arguments* are generated, then E is a T extension of the SAF if and only if E is an attack-T extension of the SAF, for $T \in \{\text{complete, grounded, preferred, stable}\}$.

Proof. First note that axiom consistency* and the reasonable argument* ordering imply the closure* and consistency* postulates. This fact is later used in this proof.
Since the acceptability of arguments* is defined with respect to the defeat relation in both cases, it suffices to show that E is conflict-free under the attack definition if and only if E is conflict-free under the defeat definition.
From left to right is trivial: if no two arguments* in E attack each other, then no two arguments* in E defeat each other.
From right to left, suppose E is not attack conflict-free. Then it has to be shown that E is not defeat conflict-free.
Suppose $A \in E$ attacks $B \in E$, but A does not defeat B. This implies that A either rebuts or undermines B on a subargument B' and that $A \prec B'$. A rebuts/undermines B', so $\texttt{Conc}(A) = \varphi$ and $\texttt{Conc}(B') = \neg\varphi$ and B' has a defeasible top rule/is an ordinary premise. If A has a defeasible top rule or is an ordinary premise, then B' defeats A and therefore E is not defeat conflict-free.
Else A has a strict top rule or is a necessary premise and we have to take a look at the maximal fallible subarguments* of A. The reasonable argument* ordering states that $A \prec B'$ implies that A cannot be strict and firm, so $M^\star(A) \neq \emptyset$. Take the weakest argument* A' in $M^\star(A)$. A' has a defeasible top rule or it is just an ordinary premise. By definition of arguments* it has to be the case that A is of the form $A', A_1, \ldots, A_n \rightarrow \varphi$ for $n \in \{0, 1, 2, \ldots\}$. Note that $A', A_1, \ldots, A_n$ all have a defeasible top rule or are just a premise. Since the structured argumentation framework is minimal, $X = \texttt{Conc}(\{A', A_1, \ldots, A_n\})$ is a minimal consistent set from which φ can be concluded. Therefore $X \cup \neg\varphi$ is a minimal inconsistent set. The satisfaction of the self-contradiction axiom* gives $\neg\texttt{Conc}(A') \in Cl^\star_{\mathcal{R}_s}(X \cup \neg\varphi)$. Therefore, there is an argument $C : A_1, \ldots, A_n, B' \rightarrow \neg\texttt{Conc}(A')$ (recall the assumption that all arguments* are generated). C is an argument* because $A_1, \ldots, A_n, B'$ either have a defeasible top rule or are

just a premise. These arguments* $A_1, \ldots, A_n, B'$ are also all elements of E, because E is closed* under subarguments*. By construction of C in combination with the reasonable argument* ordering, it is also clear that $A' \preceq C$. Furthermore, C has to be in E because E is closed* under strict rules. Now, it can be concluded that C defeats A'. This means that E is not defeat-conflict free. □

It can be concluded that the extensions for both conflict-free definitions are the same provided the assumptions given in Theorem 7.0.3. The following theorem combines the results of the preceding chapter about minimal structured argumentation frameworks with the result of Modgil and Prakken (2013).

Theorem 7.0.4. Let AT be an argumentation theory that is either closed under trans- or contraposition, take a corresponding $SAF = (\mathcal{A}, att, \preceq)$ and let $\preceq$ be a tolerable and reasonable argument ordering. If all arguments are generated, then E is a T extension of SAF^- if and only if E is an attack-T extension of SAF^-, for $T \in \{$complete, grounded, preferred, stable$\}$.

Proof. Suppose E is a T extension of SAF^-. Theorem 6.1.5 provides that $F(E)$ must be a T extension of SAF. The result of Modgil and Prakken (2013) (Theorem 7.0.2) states that $F(E)$ must be an attack-T extension of SAF. This means that $F(E)$ is attack conflict-free and that it contains each argument acceptable with respect to $F(E)$. Lemma 6.1.4 states that $F(E)^- = E$, so it follows that E must be attack conflict-free. Since E is a T extension of SAF^-, it has to be the case that E contains every argument in $\mathcal{A}^-$ that is acceptable with respect to E. Therefore, E is an attack-T extension of SAF^-.
Suppose now that E is an attack-T extension of SAF^-. Therefore, E is attack conflict-free, which implies that E is defeat conflict-free. Furthermore, E contains every argument acceptable with respect to E. It can be concluded that E is T extension of SAF^-. □

Therefore, the result not only applies to minimal $ASPIC^\star$ SAFs, but also to minimal SAFs. In the next chapter, the conclusion and future research are discussed.

Chapter 8

Conclusion and Future Research

8.1 Conclusion

The main aim of this thesis was to solve the long standing trivialisation problem, also known as the problem of self-defeat, which was first identified by Pollock (1994). This problem may occur when two arguments with contradictory conclusions are taken as a premise for an arbitrary conclusion. Depending on the preferences between arguments, this might cause unrelated arguments to be kept from being justified. This problem arises whenever the strict rules are instantiated with a logic in which everything can be derived in case of an inconsistency. A natural approach to this problem is to instantiate the strict rules with a paraconsistent logic to prevent explosion. In order to test whether such an instantiation makes sense, the goal was to find a paraconsistent logic such that the rationality postulates of Caminada and Amgoud (2007) still hold in case of instantiation.

The paraconsistent C_ω system, the Logic of Paradox and the logic W were discussed. Instantiating the $ASPIC^+$ framework with the logic W for the strict rules still leads to explosion. This is caused by the fact that the **Cut** rule does not hold. In order to enforce that $ASPIC^+$'s strict part behaves according to the logic W, chaining of strict rules had to be excluded. To this end, in Chapter 4 the $ASPIC^\star$ framework was proposed.

The closure and consistency postulates were investigated for the remaining three paraconsistent logics. Counterexamples were found to the consistency postulate for Da Costa's C_ω system and for the Logic of Paradox. Only the logic W remained.

The results of Modgil and Prakken (2013) provide conditions that imply the rationality postulates. These results could not be used for the logic W, since the logic W does not satisfy closure under trans-, or contraposition. The results of Dung and Thang (2014) give weaker conditions that guarantee the closure and consistency postulates. First, some generalisations were given of these results. Then, these results were exploited in order to prove the consistency postulate for the *ASPIC** framework of which the strict rules are instantiated with the logic W. The *ASPIC*$^+$ framework was thus successfully instantiated with a paraconsistent logic, while preserving the rationality postulates. Therefore, the trivialisation problem has been solved.

In answer to the trivialisation problem, Wu and Podlaszewski (2014) introduced the inconsistency-cleaned *ASPIC Lite* system. This system is similar to the argumentation formalism treated by Caminada and Amgoud (2007) and can be seen as a system specified in *ASPIC*$^+$ in which all arguments are equally preferred. Wu and Podlaszewski (2014) define an argument to be *consistent* if the conclusions of all subarguments are consistent. An argumentation framework is *inconsistency-cleaned* if all inconsistent arguments are removed. Wu and Podlaszewski (2014) proved that an inconsistency-cleaned version of the *ASPIC Lite* system avoids trivialisation and satisfies the rationality postulates. However, Leon van der Torre provided a counterexample for the satisfaction of the closure under strict rules postulate for the *ASPIC*$^+$ framework in combination with the last-link principle. This example was originally presented by using the *ASPIC Lite* system, but here it is translated into an *ASPIC*$^+$ framework.

Example 8.1.1 (Found by Leon van der Torre)**.** Given the knowledge base $\mathcal{K} = \emptyset$, $\mathcal{R}_d = \{\Rightarrow p; p \Rightarrow q; \Rightarrow \neg p \vee \neg q\}$ and $\mathcal{R}_s$ is instantiated with all valid inferences in classical logic. Assume that $\Rightarrow p$ has priority 1 (lowest), $\Rightarrow \neg p \vee \neg q$ has priority 2 (middle) and $p \Rightarrow q$ has priority 3 (highest). In that case, we can construct the following arguments with associated (last-link principle) preferences. Table 8.1 depicts arguments that can be generated and Figure 8.1 shows the defeat relation between these arguments.

Argument	**Preference**
$A_1 : \Rightarrow p$	(1)
$A_2 : \Rightarrow \neg p \vee \neg q$	(2)
$A_3 : A_1 \Rightarrow q$	(3)
$A_4 : A_1, A_2 \rightarrow p \wedge \neg q$	(1)
$A_5 : A_1, A_3 \rightarrow p \wedge q$	(1)
$A_6 : A_2, A_3 \rightarrow \neg p \wedge q$	(2)

Table 8.1: Generated arguments

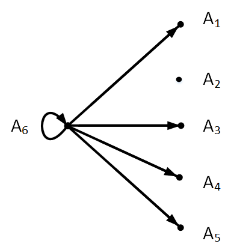

Figure 8.1: Argumentation framework with the defeats

Argument A_6 is an inconsistent argument. So according to the solution proposed by Wu and Podlaszewski (2014) for the case without preferences, A_6 needs to be deleted from the argumentation framework. Figure 8.2 shows the resulting argumentation framework.

There is a complete extension $E = \{A_1, A_2, A_3, A_4, A_5\}$. It does not satisfy closure under strict rules because A_2 and A_3 are in E and A_6 is not in E. Moreover, consistency is also

- A_1
- A_2
- A_3
- A_4
- A_5

FIGURE 8.2: Inconsistency-cleaned version

not satisfied since A_3 and A_4 are both in the complete extension E, but the conclusions q and $p \wedge \neg q$ are not consistent.

A_3 and A_4 have opposite conclusions so without preferences A_4 would defeat A_3. However, with the preference ordering chosen in the counterexample, A_4 is weaker than A_3 so A_4 cannot defeat A_3. A_3 and A_4 are also not being attacked on their subarguments. The fact that these arguments are both in the same complete extension causes the problem. Every argument that concludes with $\neg p$ uses A_1 as a subargument, so this implies that it is an inconsistent argument and it has to be removed from the framework. Therefore, A_1 is not defeated by any argument, which means that A_3 and A_4 are both in a complete extension. It can be concluded that an argument like A_6 is really needed to defeat A_3 and A_4.

Furthermore, it can be observed that if A_6 is not deleted, there are no problems at all. Figure 8.1 shows that in that case there is only one complete extension $E = \{A_2\}$. The consistency and closure postulates are both satisfied that way. Therefore, in this example, it is unwanted that A_6 is removed.

Now, take a look at the solution proposed in this paper: the *ASPIC** framework for which the strict rules are instantiated with the logic W. The same arguments can be constructed as in the original framework. This way, there is only one complete extension $\{A_2\}$ and, as explained above, the rationality postulates are all satisfied. The solution proposed in this thesis is therefore better than the solution of Wu and Podlaszewski (2014), since this solution applies to frameworks that include preferences and defeasible rules.

Another goal was to solve the efficiency problem mentioned in Chapter 3. Since the strict rules are not required to be minimal with regard to its premises, arguments are allowed to use these non-minimal strict rules. It does not make sense for defeasible rules to be required to be minimal, since defeasible rules are usually stronger when more premises are used. It was predicted that these non-minimal arguments are redundant. In that case, it is inefficient to generate these non-minimal arguments. Minimal versions of the *ASPIC*$^+$ and *ASPIC** frameworks (the latter instantiated with the logic W) were introduced in which only minimal strict rules are applied. For these minimal versions, it was proven that the outcomes are not affected under the natural assumption that minimal arguments are not weaker than their non-minimal versions.

This thesis also elucidates the relation between minimal $ASPIC^{\star}$ frameworks and classical argumentation as studied by Gorogiannis and Hunter (2011). They define an argument in the following way; $< \phi, \alpha >$ is an argument if ϕ is minimally consistent such that $\phi \vdash \alpha$, where consistency and the consequence relation $\vdash$ are classically defined. They do not use preferences between arguments, defeasible rules, or ordinary premises. In fact, they work with a version of a minimal $ASPIC^{\star}$ framework instantiated with the logic W without the preferences, defeasible rules, or ordinary premises. Therefore, this thesis' solution to the trivialisation problem combined with this thesis' solution to the efficiency problem is a straight generalisation to the results of Gorogiannis and Hunter (2011).

Modgil and Prakken (2013) suggested that defining conflict-free sets in terms of defeats is conceptually wrong. Nevertheless, they proved for the $ASPIC^{+}$ framework that the extensions are the same in case conflict-free is defined in terms of attack. This thesis ended with proving the same for minimal versions of the $ASPIC^{+}$ and minimal versions of the $ASPIC^{\star}$ frameworks instantiated with the logic W.

8.2 Future Research

In this thesis, it was proven that the rationality postulates are satisfied for $ASPIC^{\star}$ frameworks that are instantiated with the logic W. Furthermore, it has been shown that the extensions obtained from minimal $ASPIC^{\star}$ frameworks instantiated with the logic W have the same conclusions. It seems like that if the strict rules are instantiated with the logic W and if it is required that $\mathcal{R}_s$ only contains minimal strict rules, the strict rules are closed under transposition. Therefore, another research subject could be to provide a formal proof for closure under transposition and to generalise the conditions of Modgil and Prakken (2013) for the $ASPIC^{\star}$ framework. This will result in another proof for the rationality postulates.

It was taken for granted that with any paraconsistent logic, $ASPIC^{+}$ satisfies the postulates of crash-resistance and non-interference of Wu and Podlaszewski (2014). The reason for this is the absence of the Ex Falso principle as a strict inference rule. However, this should be formally proven in the future, since this is not a trivial matter.

Like the research of Wu and Podlaszewski (2014), one might examine restrictions on arguments that imply the rationality postulates. These results should imply the postulates in case preferences are taken into account.

Finally, as is discussed in the end of Chapter 6, the results of that chapter can be generalised. This can be done by changing the concept of minimal arguments to the more general less redundant relation. The reason why this can be done, is that the proofs for SAFs only use the properties of a less redundant relation and of a tolerable argument(*) ordering. In order to retain the results of Chapter 6, it needs to be shown that the smaller relation (Definition 6.3.4) is a less redundant relation.

References

Amgoud, L., & Besnard, P. (2009). Bridging the gap between abstract argumentation systems and logic. In *Scalable Uncertainty Management* (pp. 12–27). Springer.

Amgoud, L., Bodenstaff, L., Caminada, M., McBurney, P., Parsons, S., Prakken, H., ... Vreeswijk, G. (2006). Final review and report on formal argumentation system. *ASPIC Deliverable D2*, *6*.

Bench-Capon, T. J., & Dunne, P. E. (2007). Argumentation in Artificial Intelligence. *Artificial Intelligence*, *171*(10), 619–641.

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

Brożek, B. (2012). The Double Truth Controversy: An Analytical Essay. *Zygon®*, *47*(3), 643–643.

Caminada, M. (2005). Contamination in formal argumentation systems. In *Proceedings of the Seventeenth Belgian-Dutch Conference on Artificial Intelligence (BNAIC-05).* Brussels, Belgium.

Caminada, M., & Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, *171*(5), 286–310.

Da Costa, N. C. A. (1974). On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, *15*(4), 497–510.

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

Dung, P. M., Kowalski, R. A., & Toni, F. (2009). Assumption-based argumentation. In I. Rahwan & G. Simar (Eds.), *Argumentation in Artificial Intelligence* (pp. 199–218). Berlin: Springer.

Dung, P. M., Mancarella, P., & Toni, F. (2007). Computing ideal sceptical argumentation. *Artificial Intelligence*, *171*(10), 642–674.

Dung, P. M., & Thang, P. M. (2014). Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, *49*, 79-109.

Dung, P. M., Toni, F., & Mancarella, P. (2010). Some design guidelines for practical argumentation systems. In P. Baroni, F. Cerutti, M. Giacomin, & G. Simari (Eds.), *Computational Models of Argument. Proceedings of COMMA 2010* (pp. 183–194). Amsterdam etc: IOS Press.

Gorogiannis, N., & Hunter, A. (2011). Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, *175*(9), 1479–1497.

Grooters, D., & Prakken, H. (2014). Combining Paraconsistent Logic with Argumentation. In *Proceedings of COMMA 2014.* (to appear)

Jakobovits, H., & Vermeir, D. (1999). Dialectic semantics for argumentation frameworks. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law* (p. 53-62). New York: ACM Press.

Mares, E. D. (2004). Four-valued semantics for the relevant logic *R*. *Journal of Philosophical Logic*, *33*(3), 327–341.

Modgil, S., & Prakken, H. (2013). A general account of argumentation with preferences. *Artificial Intelligence*, *195*, 361-397.

Payette, G., & Schotch, P. K. (2007). On preserving. *Logica Universalis*, *1*(2), 295–310.

Pollock, J. L. (1994). Justification and defeat. *Artificial Intelligence*, *67*(2), 377–407.

Pollock, J. L. (1995). *Cognitive carpentry: A Blueprint for How to Build a Person.* Mit Press.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, *1*(2), 93–124.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, *8*(1), 219–241.

Priest, G. (1989). Reasoning about truth. *Artificial Intelligence*, *39*(2), 231–244.

Priest, G. (2011). First-Order da Costa Logic. *Studia Logica*, *97*(1), 183–198.

Read, S. (1988). *Relevant logic.* Blackwell, Oxford.

Rescher, N., & Manor, R. (1970). On inference from inconsistent premises. *Journal of Theory and Decision*, *1*, 179-219.

Sylvan, R. (1990). Variations on da Costa C systems and dual-intuitionistic logics I. analyses of C_ω and CC_ω. *Studia Logica*, *49*(1), 47–65.

Wu, Y., & Podlaszewski, M. (2014). Implementing crash-resistance and non-interference in logic-based argumentation. *Journal of Logic and Computation*. doi: 10.1093/logcom/exu017