

5-2019

A Purely Defeasible Argumentation Framework

Zimi Li

The Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

 Part of the [Other Computer Engineering Commons](#)

Recommended Citation

Li, Zimi, "A Purely Defeasible Argumentation Framework" (2019). *CUNY Academic Works*.
https://academicworks.cuny.edu/gc_etds/3078

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.

A Purely Defeasible Argumentation Framework

by

Zimi Li

A dissertation submitted to the Graduate Faculty in Computer Science in
partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York.

2019

©2019

Zimi Li

All Rights Reserved

A Purely Defeasible Argumentation Framework

by

Zimi Li

This manuscript has been read and accepted for the Graduate Faculty in
Computer Science in satisfaction of the proposal requirements for the degree
of Doctor of Philosophy.

Date

Simon Parsons
Chair of Examining Committee

Date

Robert M. Haralick
Executive Officer

Supervisory Committee:

Changhe Yuan, Queens College, The City University of New York

Sergei Artemov, Graduate Center, The City University of New York

Nir Oren, University of Aberdeen

Abstract

A Purely Defeasible Argumentation Framework

by

Zimi Li

Advisor: Simon Parsons

Argumentation theory is concerned with the way that intelligent agents discuss whether some statement holds. It is a claim-based theory¹, that is widely used in many areas, such as law, linguistics and computer science. In the past few years, formal argumentation frameworks have been heavily studied and applications have been proposed in fields such as natural language processing, the semantic web and multi-agent systems. Studying argumentation provides results which help in developing tools and applications in these areas. Argumentation is interesting as a logic-based approach to deal with inconsistent information. Arguments are constructed using a process like logical inference, with inconsistencies giving rise to conflicts between ar-

¹An argument is a claim on our attention and belief, a view that would seem to authorize treating, say, propaganda posters as arguments.[79]

guments. These conflicts can then be handled by well-founded means, giving a consistent set of well-justified arguments and conclusions.

Dung's seminal work [65] tells us how to handle the conflicts between arguments. However, it says nothing about the structure of arguments, or how to construct arguments and attack relationships from a knowledge base. ASPIC⁺ is one of the most widely used systems for structured arguments. However, there are some limitations on ASPIC⁺ if it is to satisfy widely accepted standards of rationality. Since most of these limitations are due to the use of strict rules, it is worth considering using a purely defeasible subset of ASPIC⁺.

The main contribution of this dissertation is the purely defeasible argumentation framework ASPIC_D⁺. There are three research questions related to this topic which are investigated here: (1) Do we lose anything in removing the strict elements? (2) Do purely defeasible version of theories generate the same results as the original theories? (3) What do we gain by removing the strict elements?

I show that using ASPIC_D⁺, it is possible, in a well-defined sense, to capture the same information as using ASPIC⁺ with strict rules. In particular, I prove that under some reasonable assumptions, it is possible to take a well-defined theory in ASPIC⁺, that is one with a consistent set of conclusions,

and translate it into ASPIC_D^+ such that, under the grounded semantics, we obtain the same set of justified conclusions. I also show that, under some additional assumptions, the same is true under any complete-based semantics. Furthermore, I formally characterize the situations in which translating an ASPIC^+ theory that is ill-defined into ASPIC_D^+ will lead to the same sets of justified conclusions. In doing this I deal both with ASPIC^+ theories that are not closed under transposition and theories that are axiom inconsistent. At last, I analyze the two systems in the context of the non-monotonic axioms in [46]. I show that ASPIC^+ and ASPIC_D^+ satisfy exactly same axioms under what I call the “argument construction” interpretation and the “justified conclusions” interpretation under the grounded semantics. Furthermore, because of the lack of strict elements, ASPIC_D^+ satisfies more of the non-monotonic axioms than ASPIC^+ in the “justified conclusions” interpretation under the preferred semantic. This means that ASPIC^+ and ASPIC_D^+ may not have the same justified conclusions under the preferred semantics.

Contents

1	Introduction	1
2	Literature Review	6
2.1	Argumentation Framework	7
2.1.1	Abstract Argumentation Framework	7
2.1.2	Rule-Based Argumentation Framework	19
2.1.3	Other Argumentation Frameworks	51
2.2	Non-monotonic Reasoning	54
2.2.1	Axioms of Non-monotonic System	55
2.2.2	Non-monotonic Systems	59
2.3	Summary	81
3	Motivation	84
4	A Purely Defeasible System	89
4.1	ASPIC _D ⁺ : a purely defeasible system	89

<i>CONTENTS</i>	viii
4.2 Rationality postulates	96
4.3 Unrestricted rebut	100
4.4 The expressiveness of $ASPIC_D^+$	108
4.5 Summary	130
5 Theories that are ill-defined	134
5.1 Theories that are not closed under transposition	138
5.1.1 The problem	138
5.1.2 Under the grounded semantics	143
5.1.3 Under the preferred semantics	164
5.2 Theories that are axiom inconsistent	176
5.3 Summary	197
6 Non-monotonic Properties	202
6.1 Axiomatic Reasoning and $ASPIC^+$	203
6.2 Argument Construction	207
6.3 Justified Conclusions	213
6.3.1 Justified Conclusions in Strict $ASPIC^+$ Theories . . .	215
6.3.2 Justified Conclusions in Regular $ASPIC^+$ Theories . .	217
6.3.3 Justified Conclusions in $ASPIC_D^+$ Theories	226
6.4 Discussion	232

<i>CONTENTS</i>	ix
6.5 Summary	239
7 Conclusions and Future Work	241
7.1 Summary of Contribution	241
7.2 Limitations	246
7.3 Future Work	248
References	250

Chapter 1

Introduction

Argumentation theory is concerned with the way that intelligent agents discuss whether some statements hold. It is a claim-based theory, that is widely used in many areas, such as law, linguistics and computer science. In the past few years, formal argumentation frameworks have been heavily studied and applications have been proposed in fields such as natural language processing, the semantic web and multi-agent systems. Studying argumentation provides results which help in developing tools and applications in these areas. Argumentation is interesting as a logic-based approach to deal with inconsistent information. Arguments are constructed using a process like logical inference, with inconsistencies giving rise to conflicts between arguments. These conflicts can then be handled by well-founded means, giving a consistent set of well-justified arguments and conclusions.

Researchers in AI have done much work to analyze the act of argumen-

tation. Argumentation has been used to provide a proof-theoretic semantics for non-monotonic logic, starting with the most influential work of Dung [65]. Computational argumentation systems have found particular applications in some domains, for example law, where classical logic and decision theory are too abstract to capture the richness of reasoning [66]. There are a number of works which identify the application of argumentation in a specific domain. For example, [14] considers argumentation in legal reasoning, [18] discusses argumentation in machine learning, [72] looks at an interchange format for argumentation and arguments.

Most current formal argumentation frameworks can be linked back to Dung's abstract argumentation framework [65], which gives us an overall idea about the acceptability of the arguments generated by these frameworks. All these argumentation frameworks contain defeat relationships and arguments, allowing us to represent pieces of information and the relationships between them. Dung gives different kinds of extensions allowing us to determine which of the arguments can be accepted and which can not, and the recent labeling approach [13] provides a convenient computational mechanism to determine the different extensions as well as the acceptable arguments.

Dung's abstract argumentation framework gives us an intuitive way to handle conflicts, however, it does not allow us to represent the internal the

structure of arguments. There are a number of structured argumentation frameworks which give a formal way of constructing arguments, such as rule based argumentation [64, 77, 78, 60, 67, 59, 3], logic based argumentation [15], and assumption based argumentation [36]. Because structured argumentation provides a more complete account of the reasoning of an agent, it is the form of argumentation that I study in this thesis.

There is a large family of rule-based argumentation frameworks. ASPIC [60, 67, 59, 3] is one of the most influential one and will be the basis of the work in this thesis. It contains both strict rules and defeasible rules. A strict rule means that without exception, the rule always holds, in contrast, a defeasible rule means that by experience, the rule usually holds. The very first version of ASPIC [3] only contains two attack relations, rebutting and undercutting. Rebutting captures the conflicts between conclusions and the undercutting gives the exception where a defeasible rule fails. However, ASPIC does not satisfy the rationality postulates given by [21], since the language used is not expressive enough to capture all the different kinds of conflicts. By defining restricted rebutting, the problems can be resolved.

Intuitively, if A is proposed as an argument, then one can construct a counter-argument to A , whose final conclusion conflicts with some supporting elements of A . In ASPIC⁺, the authors revised the definition of rebutting

[59]. Furthermore, undermining attack was added in ASPIC⁺ to capture the conflict between conclusions and premises [59]. After that, there have been minor revisions to ASPIC⁺, including preference dependent and preference independent attack [60], which introduces how preferences work in resolving conflicts.

However, there are some limitations on ASPIC⁺ if it is to satisfy widely accepted standards of rationality postulates [21]. Since most of these limitations are due to the use of strict rules, it is worth considering using a purely defeasible subset of ASPIC⁺. The major contribution of this dissertation is determining the limits of purely defeasible reasoning. I will analyze the limitations in two directions. First, I directly compare the difference between ASPIC⁺ theory and its purely defeasible version. Second, I use non-monotonic axioms introduced in [46] to understand the relationship at a more abstract level.

The outline of this dissertation is as follow. Chapter 2 will give a brief literature review for the current research. Chapter 3 summarize the motivation in this dissertation. Chapter 4 and Chapter 5 will introduce ASPIC_D⁺ — an argumentation framework containing only defeasible elements. Chapter 6 will exam the non-monotonic axioms introduced in [46] based on ASPIC⁺ system and ASPIC_D⁺ system. Finally, Chapter 7 concludes this dissertation

and outlines potential future work.

Parts of the results have been published in the following articles:

Z. Li, N. Oren, and S. Parsons. On the links between argumentation-based reasoning and nonmonotonic reasoning. In *International Workshop on Theory and Applications of Formal Argumentation*, pages 67—85. Springer, 2017.

Z. Li and S. Parsons. On argumentation with purely defeasible rules. In *Scalable Uncertainty Management*, pages 330—343. Springer, 2015.

Chapter 2

Literature Review

In this chapter, we will briefly review the existing work in two areas, argumentation frameworks and non-monotonic reasoning. Argumentation is widely studied in AI, and has now been applied for a range of tasks, including: legal reasoning, where classical logic and decision theory are too abstract to capture the richness of reasoning [14, 66]; providing additional information in machine-learning [18]; and as a mechanism to support interaction between autonomous agents [54]. Argumentation also has applications to the semantic web [72]. Given this breadth, I will focus here just on the most related work. From this perspective, there are two important classes of work on argumentation that I will briefly discuss, work on abstract argumentation, and work on rule-based argumentation.

Non-monotonic logic is the study of those ways of inferring additional information from given information that do not satisfy the monotonicity

property satisfied by all methods based on classical logic. [46] proposed a set of axioms which characterize non-monotonic inference in logical systems, and studied the relationships between sets of these axioms. I will use these axioms as the basis to analyze the argumentation frameworks. Many researchers have proposed systems that perform such non-monotonic inferences. The best known are probably: Reiter's default systems [73], Clark's negation as failure [29], circumscription [55], the modal system of [58], autoepistemic logic [62] and inheritance systems [80]. I will give a brief review of the systems and focus on defeasible logic, because it is closely related to argumentation frameworks.

2.1 Argumentation Framework

2.1.1 Abstract Argumentation Framework

All current formal argumentation frameworks can be linked back to Dung's abstract argumentation framework [65], which gives us an overall idea about the acceptability of the arguments generated by these frameworks. All these argumentation frameworks contain defeat relationships and arguments, allowing us to represent pieces of information and the relationships between them. Dung gives different kinds of extensions allowing us to determine which of the arguments can be accepted and which can not. We give a brief

summary here.

Definition 2.1 (Argumentation Framework) *An argumentation framework is a pair $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ where \mathcal{A} is a set of arguments, and \mathcal{R} is a binary relation collecting all pairs of arguments A and B such that A attacks against B , written as $(A, B) \in \mathcal{R}$, i.e. $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$*

A *symmetric* argumentation framework is one for which \mathcal{R} is symmetric, nonempty and irreflexive.

Given a set of arguments containing attacks, one wants to determine which arguments can be accepted and which can not. The answer corresponds to defining an **argument based semantics** or simply **semantics**, which is one of the basic building blocks of argumentation theory. In [65], different semantics for the notion of acceptability have been proposed.

Definition 2.2 (Conflict-free) *A set \mathcal{S} of arguments is said to be **conflict-free** if there are no arguments A, B in \mathcal{S} such that A attacks B .*

Definition 2.3 (Acceptable) *An argument $A \in \mathcal{A}$ is **acceptable** with respect to (w.r.t.) a set \mathcal{S} of arguments iff for each argument $B \in \mathcal{A}$: if B attacks A then B is attacked by \mathcal{S}*

Definition 2.4 (Admissible) *A conflict-free set \mathcal{S} of arguments is **admissible** iff each argument in \mathcal{S} is acceptable w.r.t. \mathcal{S} .*

The process of finding admissible set is monotonic, i.e.,

Lemma 2.1 (Fundamental Lemma) *Let \mathcal{S} be an admissible set of arguments and A, A' be arguments which are acceptable w.r.t. \mathcal{S} . Then*

1. $\mathcal{S}' = \mathcal{S} \cup \{A\}$ is admissible
2. A' is acceptable w.r.t. \mathcal{S}'

The idea of a semantics is to specify some (possibly empty) sets of acceptable arguments for a given argumentation framework. These sets are also called **argument based extensions** or simply **extensions**. In order to specify different extension, we introduce the following function.

Definition 2.5 (Characteristic Function) *The characteristic function F_{AF} of an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ is*

$$F_{AF} : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}},$$

$$F_{AF}(\mathcal{S}) = \{A \mid A \text{ is acceptable w.r.t. } \mathcal{S}\}$$

Using the characteristic function, we can then define extensions.

Definition 2.6 (Semantics) *Let \mathcal{S} be a conflict-free set of arguments, and F be the characteristic function,*

- \mathcal{S} is admissible iff $\mathcal{S} \subseteq F(\mathcal{S})$

- \mathcal{S} is a **complete extension** iff $\mathcal{S} = F(\mathcal{S})$
- \mathcal{S} is a **preferred extension** iff \mathcal{S} is the maximal (w.r.t. set inclusion) complete extension, i.e. largest fixed point of the characteristic function.
- \mathcal{S} is a **grounded extension** iff \mathcal{S} is the minimal (w.r.t. set inclusion) complete extension, i.e. least fixed point of the characteristic function.
- \mathcal{S} is a **stable extension** iff \mathcal{S} is a preferred extension which attacks all arguments in $\mathcal{A} - \mathcal{S}$

These semantics are collectively known as the “Dung semantics”. Later authors have developed additional semantics, such as the the semi-stable semantics [22], the ideal semantics [37] and the stage semantics [81]¹ and sometimes these semantics get described as “Dung semantics” as well.

Now, let’s take Dung’s example [65] to understand the above definitions. This example describes a discussion between two persons I and A , whose countries are at war, about who is responsible for blocking negotiation in their region.

$I(i_1)$ My government can not negotiate with your government because your

¹Clearly [81] was published before [65], but the approach to establishing a consistent set of arguments was only discussed as a form of argumentation semantics more recently.

government does not even recognize my government.

A(a) Your government does not recognize my government either.

I(i_2) But your government is a terrorist government.

Base on the dialogues, we can construct the following argument framework

$\langle \mathcal{A}, \mathcal{R} \rangle$

$$\mathcal{A} = \{i_1, i_2, a\}$$

$$\mathcal{R} = \{(i_1, a), (a, i_1), (i_2, a)\}$$

The conflict-free sets are:

$$\{\}, \{i_1\}, \{i_2\}, \{a\}, \{i_1, i_2\}$$

For each argument, we can tell whether it is acceptable w.r.t. some set.

- i_1 is acceptable w.r.t. $\{i_2\}, \{i_1, i_2\}, \{i_1, i_2, a\}$.
- i_2 is acceptable w.r.t. $\emptyset, \{i_2\}, \{i_1, i_2\}, \{i_1, i_2, a\}$.
- a is not acceptable.

We can easily find out the admissible sets:

$$\{\}, \{i_2\}, \{i_1, i_2\}$$

Next, we can calculate the characteristic functions:

$$\begin{aligned} F_{AF}(\{i_1\}) &= \emptyset & F_{AF}(\{i_2\}) &= \{i_1, i_2\} & F_{AF}(\{a\}) &= \emptyset \\ F_{AF}(\{i_1, a\}) &= \emptyset & F_{AF}(\{i_1, i_2\}) &= \{i_1, i_2\} & F_{AF}(\{i_2, a\}) &= \{i_1, i_2\} \\ F_{AF}(\{i_1, i_2, a\}) &= \{i_1, i_2\} & F_{AF}(\emptyset) &= \{i_2\} & & \end{aligned}$$

Using the characteristic functions, we can determine the different extensions:

- The complete extension set is: $\{i_1, i_2\}$
- The preferred extension set is: $\{i_1, i_2\}$.
- The grounded extension set is: $\{i_1, i_2\}$
- The stable extension set is: $\{i_1, i_2\}$

Labeling Approach

In [13], the author gives a nice summary of the labeling approach to establishing which arguments are acceptable:

Definition 2.7 (Labeling Function) *Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation framework and Λ be a set of labels. A Λ -labeling is a total function*

$$LF : \mathcal{A} \rightarrow \Lambda$$

The issue of semantics may be best understood using the approach of labeling each argument using $\Lambda = \{\text{IN}, \text{OUT}, \text{UNDEC}\}$. In the labeling-based approach, assigning the IN label to an argument A can be explained as the

argument being accepted, while assigning the OUT label to A can be explained as the argument being rejected. The UNDEC label means that we do not know whether the argument should be accepted or not. This is expressed by the following definitions.

Definition 2.8 (Legal Labeling) *Let LF be a labeling of argumentation framework.*

- *An IN-labeled argument is said to be legally IN iff all its attackers are labeled OUT.*
- *An OUT-labeled argument is said to be legally OUT iff it has at least one attacker that is labeled IN.*
- *An UNDEC-labeled argument is said to be legally UNDEC iff not all its attackers are labeled OUT and it does not have an attacker that is labeled IN.*

Given the labeling method, we have the corresponding definitions of Dung's framework.

Definition 2.9 (Admissible Labeling) *An admissible labeling is a labeling LF where each IN-labeled argument is legally IN and each OUT-labeled argument is legally OUT.*

Definition 2.10 (Conflict-free Labeling) *Let LF be a labeling of an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$. LF is conflict-free iff for each $A \in \mathcal{A}$ it holds that:*

- *if A is labeled IN then it does not have an attacker that is labeled IN*
- *if A is labeled OUT then it has at least one attacker that is labeled IN*

Definition 2.11 (Complete Labeling) *A complete labeling is a labeling where every IN-labeled argument is legally IN, every OUT-labeled argument is legally OUT and every UNDEC labeled argument is legally UNDEC.*

From the definition of complete labeling and admissible labeling, we know that every complete labeling is an admissible labeling (but the reverse does not hold in general). An alternative characterization of a complete labeling can be provided:

Proposition 2.1 *Let LF be a labeling of an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$. LF is **complete labeling** iff for each $A \in \mathcal{A}$ it holds that:*

1. *A is labeled IN iff all its attackers are labeled OUT*
2. *A is labeled OUT iff it has at least one attacker that is labeled IN.*

Although Proposition 2.1 does not explicitly mention UNDEC, it follows that each argument that is labeled UNDEC does not have all its attackers OUT (point 1) and it does not have an attacker that is labeled IN (point 2). Therefore, each UNDEC-labeled argument is legally UNDEC.

Then, given a complete labeling LF , we have that

- LF is a **grounded labeling** iff the set of IN arguments is minimal (w.r.t. set inclusion).
- LF is a **preferred labeling** iff the set of IN arguments is maximal (w.r.t. set inclusion).
- LF is a **stable labeling** iff the set of UNDEC arguments is empty.
- LF is a **semi-stable labeling** iff the set of UNDEC arguments is minimal (w.r.t. set inclusion).

Table 2.1 and Figure 2.1 provide an overview of how the admissibility-based semantics can be expressed in terms of complete labelings.

The labeling approach exactly matches the semantics introduced by Dung [65] and others. If LF is a complete labeling, then every x labeled IN by LF is in the complete extension, and so on for grounded, preferred, stable and semi-stable labelings. Note that, as shown in Figure 2.1, all the commonly used

Table 2.1: Describing admissibility based semantics in terms of complete labelings[13]

restriction on complete labeling	resulting semantics
no restrictions	complete semantics
empty UNDEC	stable semantics
maximal IN	preferred semantics
maximal OUT	preferred semantics
maximal UNDEC	grounded semantics
minimal IN	grounded semantics
minimal OUT	grounded semantics
minimal UNDEC	semi-stable semantics

semantics are specializations of the complete semantics. These semantics, and the associated labelings, are therefore collectively known as *complete-based*.

Now, given a set of arguments and attacks, we will find it useful to be able to talk about the subset of arguments that are **IN** since these are the ones that are acceptable according to a particular semantics. Since we may have multiple extensions, and different arguments will be **IN** in different extensions, in general it is not possible to say which arguments are **IN** (accepted) without reference to an extension. If, as we will want to below, we want to talk more generally, we need some additional concepts. We will say that the *status* of an argument is as follows:

- An argument is *credulously accepted* if it is a member of at least one extension.

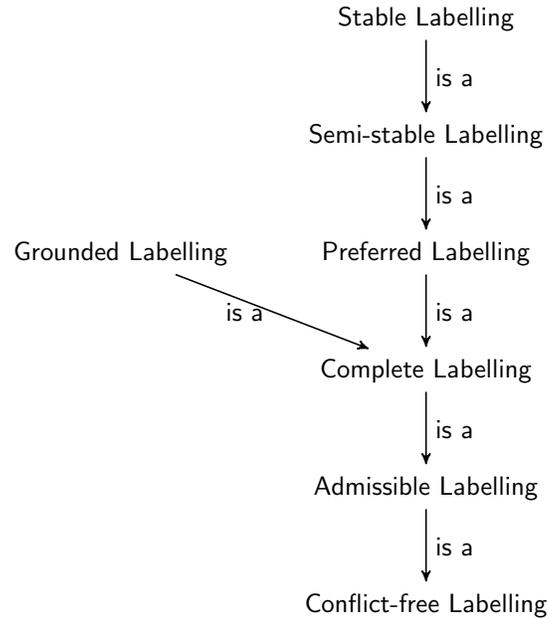


Figure 2.1: Relations among alternative labeling notions

- An argument is *sceptically accepted* if it is a member of every extension.
- If neither of the above hold, then the argument is *rejected*.

Given the relationship between labelings and extensions, this is equivalent to saying that an argument is credulously accepted if it is IN in at least one labeling, that an argument is sceptically accepted if it is IN in every extension, and rejected if it is OUT or UNDEC in every extension.

Under the grounded semantics, where there is one extension (and labeling), an argument will either be accepted or rejected. Under the preferred semantics, we can talk of the *sceptically preferred extension* to mean the set

of all the arguments that are sceptically accepted under the preferred semantics. The sceptically preferred extension is thus the intersection of the preferred extensions. We will call the set of conclusions of the arguments in the sceptically preferred extension the *sceptically justified conclusions*. Thinking about the sceptically preferred extension leads to the notion of an argumentation framework being *relatively grounded*. This will be the case where the grounded extension coincides with the sceptically preferred extension².

Finally, we will need to discuss whether sets of arguments are all accepted under the same conditions. We say that two arguments, A and B have the *same status* iff:

- A is labeled IN iff B is labeled IN; and
- A is labeled UNDEC or OUT iff B is labeled UNDEC or OUT.

Note that we interpret this definition as applying to any semantics. If there is one extension, then A and B are either both accepted or rejected. If there are multiple extensions, then for every extension in which A is IN then B is IN, and for every extension for which A is not IN then B is not IN (and vice versa).

Another line of work on abstract argumentation is to add a notion of

²[30] shows that any symmetric argumentation framework is relatively grounded.

support between arguments to the set of arguments and attacks of [65], and the properties of such “bipolar” systems are explored in, for example [7, 26, 27].

2.1.2 Rule-Based Argumentation Framework

Dung’s abstract argumentation theory gives a nice clue about how to construct argumentation frameworks, however, it says nothing about what arguments actually look like, and how to construct them and the attack relationship from a knowledge base. Rule-based argumentation frameworks, which are one of the largest families of argumentation frameworks, tell us how to do these things. In this section, we will introduce some well-known rule-based argumentation frameworks.

The ASPIC⁺ argumentation framework is one of the most influential rule-based argumentation frameworks. Historically, the ASPIC⁺ framework originates from the European ASPIC (Argumentation Service Platform with Integrated Components) project whose goal was to develop a common framework to underpin services that are emerging as core functions of the argumentation paradigm.

ASPIC⁺ defines two kinds of inference rules: strict rules (denoted \rightarrow), meaning the conclusion is always accepted without any exception, and defea-

sible rules (denoted \Rightarrow), meaning the conclusion is accepted unless there is an exception. The ASPIC⁺ argumentation system, with symmetric negation, is presented below. All the definitions in this section can be found in [60] — we have just made minor changes to the presentation, typically to provide further explanation of aspects that are often found confusing.

Definition 2.12 (ASPIC⁺ Argumentation System) *An argumentation system is a triple $AS = \langle \mathcal{L}, \mathcal{R}, n \rangle$ where:*

- \mathcal{L} is a logical language closed under negation $\bar{\cdot}$ ³.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict (\mathcal{R}_s) and defeasible (\mathcal{R}_d) inference rules of the form $\phi_1, \dots, \phi_n \rightarrow \phi$ and $\phi_1, \dots, \phi_n \Rightarrow \phi$ respectively (where ϕ_i, ϕ are meta-variables ranging over wff in \mathcal{L}), and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $n : \mathcal{R}_d \mapsto \mathcal{L}$ is a naming convention for defeasible rules.

As discussed in [21], it is helpful to think of completing the set of strict rules by considering all the negative connections between propositions mentioned in a strict rule:

Definition 2.13 (Transposition) *A strict rule s is a **transposition** of*

³As presented in [60, page 365], ASPIC⁺ is defined more generally, allowing for both symmetric negation as we have here (through the notion of “contradictory”), but also allowing for an asymmetric form of negation (through the notion of “contrary”).

$\phi_1, \dots, \phi_n \rightarrow \psi$ iff $s = \phi_1, \dots, \phi_{i-1}, \neg\psi, \phi_{i+1}, \dots, \phi_n \rightarrow \neg\phi_i$ for some $1 \leq i \leq n$.

Based on the defined notion of transposition, we now define a closure operator for the set \mathcal{R}_s .

Definition 2.14 (Closure) *Let \mathcal{R}_s be a set of strict rules. $Cl_{tp}(\mathcal{R}_s)$ is a minimal set such that:*

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

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

Definition 2.15 (Closure under Strict Rule) *Let $\mathcal{P} \subseteq \mathcal{L}$, the closure of \mathcal{P} under the set \mathcal{R}_s of strict rules, denoted as $Cl_S(\mathcal{P})$, is the smallest set such that:*

- $\mathcal{P} \subseteq Cl_S(\mathcal{P})$
- if $\phi_1, \dots, \phi_n \rightarrow \psi \in \mathcal{R}_s$ and $\phi_1, \dots, \phi_n \in Cl_S(\mathcal{P})$ then $\psi \in Cl_S(\mathcal{P})$

If $\mathcal{P} = Cl_S(\mathcal{P})$, then \mathcal{P} is said to be closed under the set \mathcal{R}_s . In ASPIC⁺, we think of the rules as the part of the system that allows inferences to be made — it is the machinery that permits reasoning. The information that is subject to reasoning is then contained in a knowledge base:

Definition 2.16 (ASPIC⁺ Knowledge Base) A knowledge base in an argumentation system $\langle \mathcal{L}, \mathcal{R}, n \rangle$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets \mathcal{K}_n (the axioms) and \mathcal{K}_p (the ordinary premises).

We say that a set of propositions in knowledge base is *consistent* iff there do not exist two propositions a and a' such that $a = \overline{a'}$.

The definitions of argumentation system and knowledge base distinguish the premises and the inference rules into two sets, the set of strict elements (\mathcal{R}_s and \mathcal{K}_n) and the set of defeasible elements (\mathcal{R}_d and \mathcal{K}_p). As we will see below, the reason for this distinction is that the defeasible elements are the ones that can be attacked. Combining the notions of argumentation system and knowledge base gives us the notion of an argumentation theory:

Definition 2.17 (ASPIC⁺ Argumentation Theory) An argumentation theory AT is a pair $\langle AS, \mathcal{K} \rangle$ of an argumentation system AS and a knowledge base \mathcal{K} .

Before defining precisely what an argument is, we need to introduce some notions which can be defined by just understanding that an argument is made up of some subset of the knowledge base \mathcal{K} , along with a sequence of rules, that lead to a conclusion. Given this, $\text{Prem}(\cdot)$ returns all the premises, $\text{Rules}(\cdot)$ returns all the inference rules, $\text{Conc}(\cdot)$ returns the conclusion and

$\text{TopRule}(\cdot)$ returns the last rule in the argument. $\text{Sub}(\cdot)$ returns all the sub-arguments of a given argument, that is all the arguments that are a subset of the given argument.

Definition 2.18 (ASPIC⁺ Argument) *An argument A on the basis of an argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$ is:*

1. ϕ if $\phi \in \mathcal{K}$ with: $\text{Prem}(A) = \{\phi\}$; $\text{Rules}(A) = \emptyset$; $\text{Conc}(A) = \{\phi\}$;
 $\text{Sub}(A) = \{A\}$; $\text{TopRule}(A) = \text{undefined}$.
2. $A_1, \dots, A_n \rightarrow \phi$ if A_i are arguments such that there exists a strict rule $\text{Conc}(A_1), \dots, \text{Conc}(A_n) \rightarrow \phi$ in \mathcal{R}_s . $\text{Prem}(A) = \text{Prem}(A_1) \cup \dots \cup \text{Prem}(A_n)$; $\text{Rules}(A) = \text{Rules}(A_1) \cup \dots \cup \text{Rules}(A_n) \cup \{\text{Conc}(A_1), \dots, \text{Conc}(A_n) \rightarrow \phi\}$; $\text{Conc}(A) = \phi$; $\text{Sub}(A) = \text{Sub}(A_1) \cup \dots \cup \text{Sub}(A_n) \cup \{A\}$; $\text{TopRule}(A) = \text{Conc}(A_1), \dots, \text{Conc}(A_n) \rightarrow \phi$.
3. $A_1, \dots, A_n \Rightarrow \phi$ if A_i are arguments such that there exists a defeasible rule $\text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi$ in \mathcal{R}_d . $\text{Prem}(A) = \text{Prem}(A_1) \cup \dots \cup \text{Prem}(A_n)$; $\text{Rules}(A) = \text{Rules}(A_1) \cup \dots \cup \text{Rules}(A_n) \cup \{\text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi\}$; $\text{Conc}(A) = \phi$; $\text{Sub}(A) = \text{Sub}(A_1) \cup \dots \cup \text{Sub}(A_n) \cup \{A\}$; $\text{TopRule}(A) = \text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi$.

We write $\mathcal{A}(AT)$ to denote the set of arguments on the basis of the theory

AT. Sometimes we need to distinguish the kinds of premise and rule used in an argument. Thus we distinguish:

- $\text{Prem}_p(A) = \text{Prem}(A) \cap \mathcal{K}_p$ and $\text{Prem}_n(A) = \text{Prem}(A) \cap \mathcal{K}_n$
- $\text{Rules}_d(A) = \text{Rules}(A) \cap \mathcal{R}_d$ and $\text{Rules}_s(A) = \text{Rules}(A) \cap \mathcal{R}_s$

These distinctions, in turn, allow us to distinguish different classes of argument. We say that an argument A is *consistent* iff $\{\text{Conc}(A') \mid A' \in \text{Sub}(A)\}$ are consistent. We further say that an argument A is *strict* if the only rules that A contains are strict, that is $\text{Rules}_d = \emptyset$; A is *defeasible* if A contains at least one defeasible rule, $\text{Rules}_d \neq \emptyset$; A is *firm* if the only premises that A contains are axioms, $\text{Prem}_p(A) = \emptyset$; A is *plausible* if A contains at least one ordinary premise, $\text{Prem}_p(A) \neq \emptyset$. The definition of strict and defeasible are disjoint; firm and plausible are disjoint. However, a strict argument can contain ordinary premises, a firm argument can contain defeasible rules. Therefore, we need to identify both aspects of an argument to fully characterize it, for example arguments are “strict and firm”.

Given the topics we will be discussing, it is necessary to consider the following relationship between arguments:

Definition 2.19 (Strict Continuation of Arguments) *For any set of arguments $\{A_1, \dots, A_n\}$, the argument A is a strict continuation of $\{A_1, \dots, A_n\}$*

iff:

- *the ordinary premises in A are exactly those in $\{A_1, \dots, A_n\}$;*
- *the defeasible rules in A are exactly those in $\{A_1, \dots, A_n\}$;*
- *the strict rules and axiom premises of A are a superset of the strict rules and axiom premises in $\{A_1, \dots, A_n\}$.*

An argument can be attacked in three ways: on its ordinary premises, on its conclusion, or on its inference rules. These three kinds of attack are called undermining, rebutting and undercutting attacks, respectively.

Definition 2.20 (ASPIC⁺ Attack) *An argument A attacks an argument B iff A undermines, rebuts or undercuts B , where:*

- *A undermines B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' = \phi \in \text{Prem}(B)$ and $\phi \in \mathcal{K}_p$.*
- *A rebuts B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' \in \text{Sub}(B)$ of the form $B'_1, \dots, B'_2 \Rightarrow \phi$.*
- *A undercuts B (on B') iff $\text{Conc}(A) = \overline{n(r)}$ for some $B' \in \text{Sub}(B)$ such that $\text{TopRule}(B)$ is a defeasible rule r of the form $\phi_1, \dots, \phi_n \Rightarrow \phi$.*

We denote “ A attacks B ” by (A, B) .

Note that, in the ASPIC⁺ attack relation, rebutting is *restricted*. That is an argument with a strict `TopRule` can rebut an argument with a defeasible `TopRule`, but not vice versa. ([23] introduce the ASPIC- systems which use unrestricted rebut).

Attacks can be distinguished as to whether they are preference-dependent (rebutting and undermining) or preference-independent (undercutting). The former succeed only when the attacker is preferred. The latter succeed whether or not the attacker is preferred.

Definition 2.21 (Preference Ordering) *A preference ordering \preceq is a binary relation over arguments, i.e., $\preceq \subseteq \mathcal{A} \times \mathcal{A}$, where \mathcal{A} is the set of all arguments constructed from the knowledge base in an argumentation system.*

We say A 's preference level is less than or equal to that of B iff $A \preceq B$.

$A \prec B$ is then defined as usual as $A \preceq B$ and $B \not\preceq A$. [60] shows that a particular class of “reasonable” preference orderings have useful properties.

These orderings are defined as:

Definition 2.22 (Reasonable Argument Orderings) *An argument ordering \preceq is reasonable iff:*

1.
 - $\forall A, B$, if B is strict and firm then $B \not\prec A$;

Thus every strict and firm argument is at least as highly ranked as any other argument.

- $\forall A, B$, if A is strict and firm and B is plausible or defeasible, then $B \prec A$;

Thus any strict and firm argument is more highly ranked than any plausible or defeasible argument.

- $\forall A, A', B$ such that A' is a strict continuation of A , if $A \not\prec B$ then $A' \not\prec B$, and if $B \not\prec A$ then $B \not\prec A'$

Thus applying strict rules and premises to the conclusion of an argument neither weakens nor strengthens that argument.

2. Let $\{C_1, \dots, C_n\}$ be a finite subset of A , and for $i = 1 \dots n$, let $C^{+\setminus i}$ be some strict continuation of $\{C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_n\}$. Then it is not the case that: $\forall i, C^{+\setminus i} \prec C_i$

[60] suggests two ways of establishing a reasonable ordering, the *weakest link principle* and *last link principle*. These two principles define how to combine the two pre-orderings \leq, \leq' over \mathcal{R}_d and \mathcal{K}_p respectively into an ordering over arguments. The weakest link principle considers all the non-strict elements in an argument — defeasible rules and ordinary premises.

Definition 2.23 (Weakest Link Principle) *Let A and B be two arguments. $A \preceq B$ iff:*

- *Both A and B are strict, and $\text{Prem}_p(A) \sqsubseteq \text{Prem}_p(B)$; or*
- *Both A and B are firm, and $\text{Rules}_d(A) \sqsubseteq \text{Rules}_d(B)$; or*
- *$\text{Prem}_p(A) \sqsubseteq \text{Prem}_p(B)$ and $\text{Rules}_d(A) \sqsubseteq \text{Rules}_d(B)$.*

The last link principle considers the last defeasible inference rule used in an argument.

Definition 2.24 (Last Defeasible Rule) *Let A be an argument.*

- *$\text{LastDefRules}(A) = \emptyset$ iff $\text{Rules}_d(A) = \emptyset$.*
- *If $\text{TopRule}(A) = A_1, \dots, A_n \Rightarrow \phi$, then $\text{LastDefRules}(A) = \{A_1, \dots, A_n \Rightarrow \phi\}$; otherwise, $\text{LastDefRules}(A) = \text{LastDefRules}(A_1) \cup \dots \cup \text{LastDefRules}(A_n)$.*

Definition 2.25 (Last Link Principle) *Let A and B be two arguments.*

$A \preceq B$ iff:

- *$\text{LastDefRules}(A) = \text{LastDefRules}(B) = \emptyset$ and $\text{Prem}_p(A) \sqsubseteq \text{Prem}_p(B)$;*
- or*
- *$\text{LastDefRules}(A) \sqsubseteq \text{LastDefRules}(B)$.*

Clearly both these principles hinge on the definition of \preceq in terms of the ordering over rules and premises. There are two ways of defining \preceq , Elitist and Democratic. The Elitist approach compares sets on their minimal elements and Democratic approach compares sets on their maximal elements.

Definition 2.26 (Orderings) *Let Γ and Γ' be finite sets . Then \preceq is defined as follows:*

- *If $\Gamma = \emptyset$ then $\Gamma \not\preceq \Gamma'$.*
- *If $\Gamma = \Gamma' = \emptyset$ then $\Gamma \preceq \Gamma'$.*
- *$\Gamma \preceq_{Eli} \Gamma'$ if $\exists x \in \Gamma, s.t., \forall Y \in \Gamma', X \leq Y$.*
- *$\Gamma \preceq_{Dem} \Gamma'$ if $\forall x \in \Gamma, s.t., \exists Y \in \Gamma', X \leq Y$.*

In other words, $\Gamma \preceq_{Eli} \Gamma'$ if *some* element in Γ is less than *every* element in Γ' , while $\Gamma \preceq_{Dem} \Gamma'$ if *every* element in Γ is less than *some* element of Γ' . When necessary, we will distinguish between the orderings obtained by using the Elitist and Democratic definitions of \preceq , referring to, for example, the “elitist weakest link principle” and the “democratic last link principle”.

By combining the definition of arguments, attack relations and preference ordering, we have the following definitions:

Definition 2.27 (Structured Argumentation Framework) A structured argumentation framework is a triple $\langle \mathcal{A}, att, \preceq \rangle$, where \mathcal{A} is the set of all arguments constructed from the knowledge in the argumentation system, att is the attack relation, \preceq is an preference ordering on \mathcal{A} .

Definition 2.28 (ASPIC⁺ Defeat) A defeats B iff A undercuts B, or if A rebuts/undermines B on B' and B' 's preference level is less than or equal to that of A ($B' \preceq A$).

Then the idea of an argumentation framework follows from Definitions 2.18 and 2.28.

Definition 2.29 (Argumentation Framework) An (abstract) argumentation framework AF corresponding to a structured argumentation framework $SAF = \langle \mathcal{A}, att, \preceq \rangle$ is a pair $\langle \mathcal{A}, Defeats \rangle$ such that $Defeats$ is the defeat relation on \mathcal{A} determined by SAF .

The following provides a concrete example of an ASPIC⁺ argumentation framework, adapted from [60]:

Example 2.1 Consider that we have the argumentation system $AS = \langle \mathcal{L}, \mathcal{R}, n \rangle$ where:

$$\mathcal{L} = \{a, b, c, d, e, f, \bar{a}, \bar{b}, \bar{c}, \bar{d}, \bar{e}, \bar{f}, nd, \overline{nd}\}$$

\mathcal{R} is then:

$$\mathcal{R}_s = \{d, f \rightarrow \bar{b}\}$$

$$\mathcal{R}_d = \{a \Rightarrow b; \bar{c} \Rightarrow d; e \Rightarrow f; a \Rightarrow \overline{nd}\}$$

and $n(\bar{c} \Rightarrow d) = nd$. We then add the knowledge-base \mathcal{K} such that:

$$\mathcal{K}_n = \emptyset$$

$$\mathcal{K}_p = \{a; \bar{c}; e; \bar{e}\}$$

to get the argumentation theory $AT_1 = \langle AS, \mathcal{K} \rangle$. From this we can construct

the arguments:

$$\begin{aligned} A_1 &= [a]; & A_2 &= [A_1 \Rightarrow b]; & A_3 &= [A_1 \Rightarrow \overline{nd}]; \\ B_1 &= [\bar{c}]; & B_2 &= [B_1 \Rightarrow d]; \\ B'_1 &= [e]; & B'_2 &= [B'_1 \Rightarrow f]; & B &= [B_2, B'_2 \rightarrow \bar{b}]; \\ C &= [\bar{e}]; \end{aligned}$$

Let's call this set of arguments \mathbf{A} , so that:

$$\mathbf{A} = \{A_1, A_2, A_3, B_1, B_2, B'_1, B'_2, B, C\}$$

Note that

$$\text{Prem}(B) = \{\bar{c}; e\}$$

$$\text{Sub}(B) = \{B_1; B_2; B'_1; B'_2; B\}$$

$$\text{Conc}(B) = \bar{b}$$

$$\text{TopRule}(B) = d, f \rightarrow \bar{b}$$

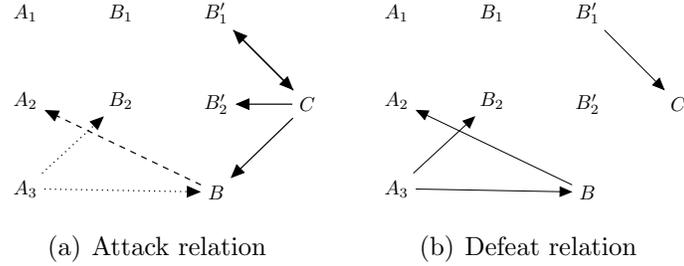


Figure 2.2: The attack and defeat relations from AT_1 in Example 2.1. In (a), a dotted arrow in denotes undercutting, a dashed arrow denotes rebutting, and a solid arrow denotes undermining. In (b) all the arrows denote defeats.

The attacks between these arguments are shown in Figure 2.2(a). These make up the set att , where:

$$att = \{(C, B'_1), (B'_1, C), (C, B'_2), (C, B), (B, A_2), (A_3, B_2), (A_3, B)\}$$

With a preference order \preceq defined by : $A_2 \prec B; C \prec B; C \prec B'_1; C \prec B'_2$, we have the structured argumentation framework $\langle \mathbf{A}, att, \preceq \rangle$. This structured argumentation framework establishes a defeat relation

$$Defeats = \{(B'_1, C), (B, A_2), (A_3, B), (A_3, B_2)\}$$

which is shown in Figure 2.2(b). With this, we can finally write down the argumentation framework $\langle \mathbf{A}, Defeats \rangle$.

Given an abstract argumentation framework constructed from a structured argumentation framework, one can use the labeling approach discussed

above to establish which arguments are acceptable under a given semantics. Since arguments are structured, we may be interested whether to accept the conclusions of an argument. We will extend the ideas above by adopting the terminology of [31] which distinguishes between sceptical, credulous and universal acceptance of the a conclusion of an argument:

Definition 2.30 (Justified Conclusions) *For $AF = \langle \mathcal{A}, Defeats \rangle$ is an argumentation framework. we say that:*

- ϕ is a credulously justified conclusion of AF iff there exists an argument A and an extension E such that $A \in E$ and $\text{Conc}(A) = \phi$.
- ϕ is a sceptical justified conclusion of AF iff for every extension E , there exists an argument $A \in E$ such that $\text{Conc}(A) = \phi$.
- ϕ is a universal justified conclusion of AF iff there exists an argument A for every extension E , such that $A \in E$ and $\text{Conc}(A) = \phi$.

We also extend the notion of status to conclusions. Given two arguments A and B , we say that $\text{Conc}(A)$ and $\text{Conc}(B)$ have the same status iff A and B have the same status.

Clearly under the grounded semantics, all of the notions of justified conclusion in Definition 2.30 coincide, and so if A is acceptable under the

grounded semantics, we just say that the conclusion of A , $\text{Conc}(A)$ is a *justified conclusion*. We will use the notion of the justified conclusions below to compare the sets of conclusions of two argumentation frameworks. If two frameworks have the same justified conclusions then they are, in some sense, equivalent because they allow the same set of conclusions to be drawn. A similar situation can arise in the case of two frameworks that have multiple extensions. If the set of extensions of the two frameworks are the same, then we can pair the sets of extensions $E_{11}, E_{12}, \dots, E_{1n}$ and $E_{21}, E_{22}, \dots, E_{2n}$ so that each E_{1i} is identical to its corresponding E_{2i} and so contains exactly the same arguments. Every argument has the same status in both sets of extensions, and we can again consider the two sets of extensions to allow the same sets of conclusions to be drawn. In such a case, by analogy to the grounded semantics, we say that the sets of extensions have the same justified conclusions.

Modgil and Prakken [60] distinguish several classes of ASPIC⁺ argumentation framework. Of particular interest to me in this dissertation is the class they call *well-defined*, which is the class for which they provide results (frameworks that are ill-defined are problematic for reasons that we will discuss below). For the slightly simpler version of ASPIC⁺ that we are

considering here, the following definition gives captures the essence⁴ of [60]’s notion of “well-defined”⁵, and so we use the same term:

Definition 2.31 (Well-defined) *An argumentation theory $AT = \langle AS, \mathcal{K} \rangle$, where $AS = \langle \mathcal{L}, \mathcal{R}, n \rangle$ is well-defined if and only if it meets the following conditions:*

1. *The strict rules R_s of \mathcal{R} are closed under transposition.*
2. *The strict elements of AT is consistent, so $Cl_S(K_n)$ is consistent. If this condition holds, AT is said to be axiom consistent.*

where $Cl_S(\cdot)$ denotes closure under strict rules.

The original ASPIC argumentation framework was introduced in [3]. In [3], the attacks are only rebutting and undercutting. [20, 21] points out that ASPIC is not expressive enough to capture all the different kinds of conflicts

⁴Any ASPIC⁺ theory, as defined here, that is well-defined in the sense used here will be the basis of a structured argumentation framework that is well-defined in the sense defined in [60].

⁵In [60], a “well-defined” theory needs to be “well-formed”, in addition to the requirements in Definition 2.31. We do not require AT to be well-formed because this property follows from symmetrical negation (for example, see the proof of Proposition 25 in [60]), so any theory we deal with is automatically well-formed. [60] also suggests that rather than being closed under transposition, having a theory be closed under contraposition will suffice to ensure the relevant behavior. A theory is closed under contraposition iff for all $S \subseteq \mathcal{L}$, $s \in S$ and ϕ , if there exists a strict argument A such that $Conc(A) = \phi$, $Prem(A) \subseteq S$, then there exists a strict argument A' such that $Conc(A') = \bar{\phi}$, $Prem(A') \subseteq S \setminus \{s\} \cup \{\bar{\phi}\}$. In other words [61], contraposition means that if there is a strict argument for ϕ which uses some set of premises from S , then replacing any of these premises s with $\bar{\phi}$ gives a strict argument for $\bar{\phi}$. Since closure under transposition is an alternative to closure under contraposition, we see no need to consider both properties here.

that may exist between arguments, and introduced the idea of rationality postulates for structured argumentation. The rationality postulates suggests that all the rule-based argumentation framework should satisfy the following properties.

Definition 2.32 (Rationality postulates) *For an argumentation theory AT and associated argumentation framework AF with a set of sceptical justified conclusions C , and extensions, under a given semantics E_1, \dots, E_n :*

Postulate 1 (Closure) AF satisfies closure, also called closure under strict rules, iff:

1. $Concs(E_i) = Cl_s(Concs(E_i))$, for all $1 \leq i \leq n$
2. $C = Cl_S(C)$

Postulate 2 (Direct consistency) AF satisfies direct consistency iff:

1. $Concs(E_i)$ is consistent for all $1 \leq i \leq n$
2. C is consistent

Postulate 3 (Indirect consistency) AF satisfies indirect consistency iff:

1. $Cl_S(Concs(E_i))$ is consistent for all $1 \leq i \leq n$
2. $Cl_S(C)$ is consistent

where $\text{Concs}(\cdot)$ denotes the conclusions of a set of arguments, and $\text{Cl}_S(\cdot)$ denotes closure under strict rules.

[21] characterises the postulates as follows. Closure ensures that: “the user [can] do [their] own reasoning (take the outcome of the formalism and apply modus ponens using the strict rules) to derive statements that the formalism apparently ‘forgot’ to entail.” Direct consistency ensures that the formalism does not generate “absurdities” in the sense of allowing two conclusions that contradict each other. Indirect consistency ensures that: “one may ... take the outcome of the formalism and apply modus ponens using the strict rules” and be sure that the results will be consistent (and hence sensible). The original ASPIC system, which was studied in [21] did not satisfy closure or indirect consistency, and ASPIC⁺ satisfies it by restricting rebut and requiring theories to be closed under transposition.

In addition to the three rationality postulates of Definition 2.32, a further postulate (sometimes presented as the first of four) is considered by some authors, for example [60]:

Definition 2.33 (Sub-argument Closure) *For an argumentation theory AT and an associated argumentation framework AF with extensions E_1, \dots, E_n , under a given semantics. AF satisfies sub-argument closure iff for any ar-*

argument A in E_i , all sub-arguments of A are in E_i , i.e., $\forall E_i \in \{E_1, \dots, E_n\}$, $\forall A \in E_i$, if $A' \in \text{Sub}(A)$, then $A' \in E_i$.

We do not consider sub-argument closure in any detail here because [21] showed that the original ASPIC system satisfies this property for any complete semantics, and the systems that we consider here are similar enough to ASPIC that the same result holds without modification (other than some slight changes in notation).

The value of the postulates is that they allow us to distinguish different argumentation theories as being rational or not — since rationality can hinge on the content of the theory, some theories expressed in some formalism can be rational, while others are not, and the work in this paper explores areas along exactly this kind of fault line in ASPIC⁺:

Definition 2.34 (Rational) *If an argumentation theory AT meets all the rationality postulates, it is said to be rational. An argumentation theory that is not rational is said to be irrational.*

Because rationality is defined with respect to justified conclusions, it depends on the semantics being applied to obtain the justified conclusions. Any use we make of the term in a formal sense will be in the context of a specific semantics or set of semantics. Informally, we sometimes use the term without

specifying a semantics to either mean for at least one semantics or for any semantics, with the context making clear which. [60] shows that any well-defined ASPIC⁺ structured argumentation framework is rational. However, an ASPIC⁺ framework may be both rational and ill-defined. We can see this in Example 2.1 where $AT_{2.1}$ is ill-defined, but is rational under any semantics because the set of justified conclusions:

$$\{a, b, \bar{c}, e, f, \overline{nd}\}$$

conforms to the rationality postulates, by being closed under strict rules, and both directly and indirectly consistent. The following example is a slight modification of Example 2.1 to highlight a theory which is irrational because its justified conclusions are not consistent:

Example 2.2 Consider a theory $AT_{2.2}$ which has the knowledge base:

$$\mathcal{K}_n = \emptyset$$

$$\mathcal{K}_p = \{b; \bar{c}; e\}$$

and where the remaining elements are the same as in $AT_{2.1}$ from Example 2.1.

Now we can construct the arguments:

$$\begin{aligned} A &= [b]; \\ B_1 &= [\bar{c}]; \quad B_2 = [B_1 \Rightarrow d]; \\ B'_1 &= [e]; \quad B'_2 = [B'_1 \Rightarrow f]; \quad B = [B_2, B'_2 \rightarrow \bar{b}]; \end{aligned}$$

The only attack is (B, A) . However, if we set $B \prec A$, then there are no defeat relations. Therefore, both A and B are justified, and the set of justified conclusions is:

$$\{b, \bar{b}, \bar{c}, d, e, f\}$$

which is not consistent.

Note that, indirect consistency can be derived by direct consistency and closure under strict rules. [67] then introduced the ASPIC⁺ argumentation framework, a revised version of ASPIC. Moreover, it also distinguishes the preference dependent attack and preference independent attack. [2] analyzes the ASPIC⁺ argumentation system, and shows that there are several drawbacks of ASPIC⁺. [69] argues that the criticisms in [2] are not justified, but are caused by a number of misconceptions. Recently, [68] points out that ASPIC⁺ may not be able to capture the lottery paradox [49]. Therefore, [68] introduced fallible arguments, and then revised the rationality postulates.

In general, the ASPIC⁺ argumentation framework does not satisfy all the rationality postulates. However, if the strict rules are closed under transposition, then any ASPIC⁺ argumentation framework satisfies all the rationality postulates. We call these frameworks rational argumentation frameworks.

There are a number of ASPIC⁺ related frameworks published recently.

[23] introduced the ASPIC- argumentation framework, where the rebuttal is unrestricted. [50] focus on purely defeasible argumentation framework, $ASPIC_D^+$, showing that the purely defeasible part of $ASPIC^+$ has the same expressiveness as $ASPIC^+$.

[19] postulates the property of *non-contamination* as being desirable for argumentation system. Non-contamination captures the idea that an argumentation framework should be proof against a localized inconsistency propagating to affect arguments which do not directly depend on the inconsistent elements.

Definition 2.35 (Non-contamination) *It may not be the case that two arguments that rebut each other can be combined into an argument that can keep any arbitrary other argument from becoming justified.*

We can see the value of this postulate by considering a case in which it does not hold. Consider an argumentation framework, based on $ASPIC^+$, in which the *ex falso quodlibet* principle, also called the principle of explosion, is incorporated a strict rule⁶. The principle states that once a contradiction has been asserted, any proposition can be inferred from it. Here is an example using such a framework:

⁶This will be the case for any argumentation system which is a super-set of classical logic since that rule is a (strict) rule in classical logic.

Example 2.3

$$\mathcal{R}_d = \{a \Rightarrow b; c \Rightarrow \neg b; e \Rightarrow d\}$$

$$\mathcal{R}_s = \{b, \neg b \rightarrow \neg d\}$$

$$\mathcal{K}_n = \{a; c; e\}$$

$$\mathcal{K}_p = \emptyset$$

The arguments are:

$$\begin{array}{llll} A_1 = [a] & A_2 = [A_1 \Rightarrow b] & & \\ B_1 = [c] & B_2 = [B_1 \Rightarrow \neg b] & [C'_2] = [A_2, B_2 \rightarrow \neg d] & \\ C_1 = [e] & C_2 = [C_1 \Rightarrow d] & & \end{array}$$

In this example, argument A_2 and argument B_2 rebut each other. That is uncontroversial. However, *ex falso quodlibet* makes it possible to construct argument C'_2 , which defeats C_2 . This is possible whatever formulae e and d represent, so C_2 can be defeated even when d and e are unrelated to a , b and c . (To riff on an example from Martin Caminada, two contradictory arguments about whether or not to put sugar in tea can make it possible to defeat an argument about nuclear disarmament.) Clearly, this does not make any sense — it is not reasonable, in general, for an argument to prevent any other argument from being justified.

Example 2.3 demonstrates that ASPIC⁺ theories do not satisfy the non-interference property. We could complain that rules like $b, \neg b \rightarrow \neg d$ should just not be used — if we want to block the propagation of inconsistency why allow rules that can be used to propagate it — but the problem with doing so is that *ex falso quodlibet* is part of the fabric of classical propositional logic. As a result, any ASPIC⁺-based system that includes classical propositional logic will have this problem. [19] proposes a solution which will work for any argumentation system by redefining arguments to be consistent (as, for example, in [5]), thus ruling out inconsistent arguments (and hence the reason non-interference fails to hold).

The original solution presented by Caminada in [19] is to rule out inconsistent arguments. Caminada does this by constructing arguments as normal, that is as in Definition 2.18, and then filtering out the inconsistent ones before applying Dung’s semantics. As a result of the filtering, no inconsistent arguments can be justified, and since combining two rebutting arguments into a third argument would make that third argument inconsistent, restricting arguments to being consistent ensures non-contamination. Ruling out inconsistent arguments is reasonable — inconsistent arguments are pretty bad arguments after all — and has echoes of [39] where inconsistent arguments are permitted, but are considered weaker than any other kind of argument.

ArgTrust

Similar to ASPIC⁺, [77] uses a graph to represent the argument, which may be more intuitive to a decision maker. The knowledge available to an agent is defined as $\Sigma = P \cup \Delta$ where P is a set of premises, each of which is a logical statement in a language \mathcal{L} . Δ is a set of inference rules, each of which denoted by δ , is of the form:

$$c :- p_1, p_2, \dots, p_n$$

where p_i and c are members of \mathcal{L} . In other words the inference rules link some sets of premises p_i to a conclusion c .

The inference rule can be represented as a graph

Definition 2.36 (Rule Network) *A rule network \mathcal{R} is a directed hypergraph $\langle V^r, E^r \rangle$ where:*

1. *the set of vertices V^r are elements of \mathcal{L}*
2. *the set of edges E^r are inference rules δ*
3. *the initial vertices of an edge $e \in E^r$ are the premises of the corresponding inference rule*
4. *the terminal node of that edge is the corresponding conclusion c .*

Thus a rule network simply connects premises and conclusions of rules. Under certain circumstances, a rule network captures a proof.

Definition 2.37 (Proof Network) *For a given knowledge base $\Sigma = P \cup \Delta$, a rule network $\langle V^r, E^r \rangle$ is a proof network iff every premise of each inference rule $\delta \in E^r$ is either a member of P or the conclusion of some $\delta' \in E^r$*

Some proof networks correspond to arguments:

Definition 2.38 (Argument) *An argument A from a knowledge base $\Sigma = P \cup \Delta$ is a pair $\langle h, H \rangle$ where $H = \langle V^r, E^r \rangle$ is a proof network for h , and h is the only leaf of H .*

H is the support of the argument, and h is the conclusion. $C(H)$ is the set of intermediate conclusions of H , the set of all the conclusions of the $\delta \in E^r$ other than h . $P(H)$ is the set of pure premises of H , the premises of the $\delta \in E^r$ that are not intermediate conclusions of H .

[77] distinguishes a number of ways that a defeat may occur as follows:

Definition 2.39 (Defeats) *An argument $\langle h_1, H_1 \rangle$ defeats an argument $\langle h_2, H_2 \rangle$ iff it rebuts, undermines, intermediate-rebuts, or undercuts it, where:*

1. *An argument $\langle h_1, H_1 \rangle$ rebuts another argument $\langle h_2, H_2 \rangle$ iff $h_1 \equiv \neg h_2$.*

2. An argument $\langle h_1, H_1 \rangle$ premise-undercuts⁷ another argument $\langle h_2, H_2 \rangle$ iff there is a premise $p \in P(H_2)$ such that $h_1 \equiv \neg p$.
3. An argument $\langle h_1, H_1 \rangle$ intermediate-rebuts another argument $\langle h_2, H_2 \rangle$ iff there is an intermediate conclusion $c \in C(H_2)$ such that $c \neq h_2$ and $h_1 \equiv \neg c$.
4. An argument $\langle h_1, H_1 \rangle$ inference-undercuts⁸ another argument $\langle h_2, H_2 \rangle$ iff there is an inference rule $\delta \in \Delta(H_2)$ such that $h_1 \equiv \neg \ulcorner \delta \urcorner$.

where $\ulcorner \cdot \urcorner$ is the naming convention.

Let's look at an example for ArgTrust.

Example 2.4 Suppose that John is trying to decide whether or not he should watch the film *hce*, the only film currently in the knowledge base. Dave, a friend of John, suggests that they should watch the film if the film is an Indian film and directed by Almodovar. Figure 2.3 shows a rule network for this example. The rectangular nodes denote the premises and the conclusion. The oval, which represent a hyper-edge, denotes an inference rule.

⁷Similar to undermines in ASPIC+.

⁸Similar to undercuts in ASPIC+.

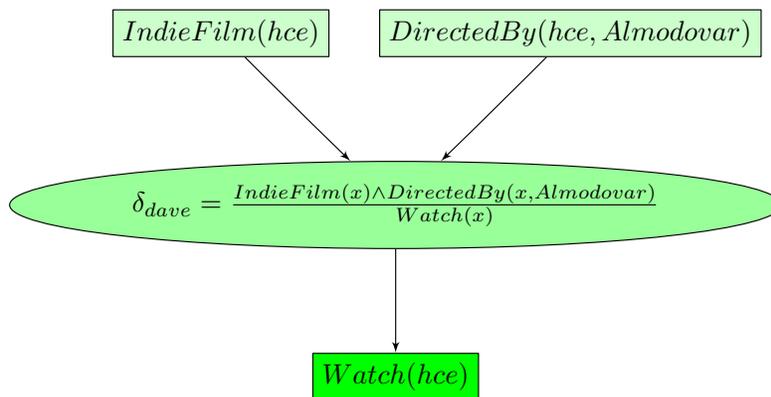


Figure 2.3: Example for Rule Network

Defeasible Logic Programming

There are a number of research works about defeasible logics in argumentation [74, 71, 28]. Among them, [41] is influential. [41] presents an argumentative approach — defeasible logic programming (DeLP). The DeLP language is defined in terms of three disjoint sets: a set of facts, a set of strict rules and a set of defeasible rules.

In the language of DeLP a literal L is a ground atom A or a negated ground atom $\sim A$, where \sim represents strong negation. The following definitions introduce the language of DeLP.

Definition 2.40 (Fact) *A fact is a literal, i.e. a ground atom, or a negated ground atom.*

Definition 2.41 (Strict Rule) *A strict rule is an ordered pair, denoted*

“ $Head \leftarrow Body$ ”, whose first member, *Head*, is a literal, and whose second member, *Body*, is a finite non-empty set of literals. A strict rule with the head L_0 and body $\{L_1, \dots, L_n\}$ can also be written as: $L_0 \leftarrow L_1, \dots, L_n (n > 0)$.

Definition 2.42 (Defeasible Rule) A **defeasible rule** is an ordered pair, denoted “ $Head \prec Body$ ”, whose first member, *Head*, is a literal, and whose second member, *Body*, is a finite non-empty set of literals. A strict rule with the head L_0 and body $\{L_1, \dots, L_n\}$ can also be written as: $L_0 \prec L_1, \dots, L_n (n > 0)$.

Combining the above definitions, we can define a defeasible logic program.

Definition 2.43 (Defeasible Logic Program) A **Defeasible Logic Program** P is a possibly infinite set of facts, strict rules and defeasible rules. In a program P , we will distinguish the subset Π of facts and strict rules, and the subset Δ of defeasible rules. When required, we will denote P as (Π, Δ) .

Next, we will define what constitutes a defeasible derivation and a strict derivation.

Definition 2.44 (Defeasible Derivation) Let $P = (\Pi, \Delta)$ be a DeLP, and L a ground literal. A **defeasible derivation** of L from P , denoted $P \vdash L$, consists of a finite sequence $L_1, L_2, \dots, L_n = L$ of ground literals, and each literal L_i is in the sequence because:

- L_i is a fact in Π , or
- there exists a rule R_i in P (strict or defeasible) with head L_i and body B_1, B_2, \dots, B_k and every literal of the body is an element L_j of the sequence appearing before L_i ($j < i$).

Definition 2.45 (Strict Derivation) Let P be a DeLP and h a literal with a defeasible derivation $L_1, L_2, \dots, L_n = h$. We will say that h has a **strict derivation** from P , denoted $P \vdash L$, if either h is a fact or all the rules used for obtaining the sequence L_1, L_2, \dots, L_n are strict rules.

Next we will introduce the defeasible argumentation formalism.

Definition 2.46 (Argument Structure) Let h be a literal, and $P = (\Pi, \Delta)$ a DeLP. We say that $\langle \mathcal{A}, h \rangle$ is an **argument structure** for h , if \mathcal{A} is a set of defeasible rules of Δ , such that:

1. there exists a defeasible derivation for h from $\Pi \cup \mathcal{A}$,
2. the set $\Pi \cup \mathcal{A}$ is consistent,
3. \mathcal{A} is minimal: there is no proper subset \mathcal{A}' of \mathcal{A} such that \mathcal{A}' satisfies conditions 1 and 2.

In summary, an argument structure $\langle \mathcal{A}, h \rangle$, or simply an argument \mathcal{A} for h , is a minimal consistent set of defeasible rules, obtained from a defeasible

derivation for a given literal h . The literal h will also be called the **conclusion** supported by A . Note that strict rules are not part of an argument structure.

Definition 2.47 (Sub-argument) *An argument structure $\langle \mathcal{A}', h' \rangle$ is a **sub-argument** structure of $\langle \mathcal{A}, h \rangle$ if $\mathcal{A}' \subseteq \mathcal{A}$.*

It is important to note that the union of arguments is not always an argument. That is, given two argument structures $\langle \mathcal{A}, h \rangle$ and $\langle \mathcal{A}', h' \rangle$, the set $\mathcal{A} \cup \mathcal{A}'$ might not be an argument, because $\mathcal{A} \cup \mathcal{A}'$ could be not minimal or $\mathcal{A} \cup \mathcal{A}' \cup \Pi$ could be contradictory.

In DeLP, the attack relationship is defined as follows

Definition 2.48 (Disagree) *Let $P = (\Pi, \Delta)$ be a DeLP. We say that two literals h and h' **disagree**, iff the set $\Pi \cup \{h, h'\}$ is contradictory.*

Definition 2.49 (Attack) *We say that $\langle \mathcal{A}_1, h_1 \rangle$ **attacks** $\langle \mathcal{A}_2, h_2 \rangle$ at literal h , iff there exists a sub-argument $\langle \mathcal{A}, h \rangle$ of $\langle \mathcal{A}_2, h_2 \rangle$ such that h and h_1 disagree.*

Again, we will use an example to introduce defeasible logic programs.

Example 2.5 *Consider the following DeLP.*

$$\Pi = \left\{ \begin{array}{ll} c & d \\ h_1 \leftarrow b & h_2 \leftarrow b \\ p \leftarrow e & \sim p \leftarrow f \\ h \leftarrow h_1, h_2 & \end{array} \right\}, \Delta = \left\{ \begin{array}{ll} b \prec c & b \prec d \\ e \prec c & f \prec d \end{array} \right\}$$

We can construct the following argument structures: $\langle \mathcal{A}_1, h_1 \rangle = \langle \{b \prec c\}, h_1 \rangle$ and $\langle \mathcal{A}_2, h_2 \rangle = \langle \{b \prec d\}, h_2 \rangle$. Consider now the set $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 = \{b \prec c, b \prec d\}$. From $\Pi \cup \mathcal{A}$ there exists a defeasible derivation for h , however, $\langle \mathcal{A}, h \rangle$ is not an argument structure because \mathcal{A} is not the minimal set of defeasible rules that provides an argument structure for h because \mathcal{A}_1 is a proper subset of \mathcal{A} . Therefore the argument should be $\langle \mathcal{A}_1, h \rangle$.

2.1.3 Other Argumentation Frameworks

Unlike the rule-based argumentation framework, there are some other argumentation frameworks, such as logic-based argumentation frameworks. [82] proposed a more general approach to logic-based argumentation. It leaves the logic for deduction as a parameter, and this is developed in assumption-based argumentation [35].

More research about logic-based argumentation include: variants of defeasible logic with annotations for lattice-theoretic truth values [76] and for possibility theory [1], temporal reasoning calculi used with defeasible logic [12] and with classical logic [53], minimal logic [48], and a form of modal logic [38].

Classical Logic-based Argumentation Framework

The classical logic based argumentation framework [25, 4, 15] is an argument iff $\Phi \vdash \alpha$ and there is no $\Phi' \subset \Phi$ such that $\Phi' \vdash \alpha$ and $\Phi \vdash \perp$. [15] explores a framework for argumentation based on classical logic in which an argument is a pair where the first item is a minimal consistent set of formula that proves the second item (which is a formula). The authors assume familiarity with classical logic, and consider a propositional language. We use α, β, \dots to denote formulas and Δ, Φ, \dots to denote sets of formulas, and Δ denotes a database (a finite set of formulas)

Definition 2.50 (Argument) *An argument is a pair $\langle \Phi, \alpha \rangle$ such that*

1. Φ is consistent.
2. $\Phi \vdash \alpha$.
3. Φ is a minimal subset of Δ satisfying (2).

We say that $\langle \Phi, \alpha \rangle$ is an argument for α . We call α the consequent of the argument and Φ the support of the argument.

Conflicts is captured in a concrete form with the notion of defeater.

Definition 2.51 (Defeater) *A defeater for an argument $\langle \Phi, \alpha \rangle$ is an argument $\langle \Psi, \beta \rangle$ such that $\beta \vdash \neg(\phi_1 \wedge \dots \wedge \phi_n)$ for some $\{\phi_1, \dots, \phi_n\} \subseteq \Phi$.*

Some arguments directly oppose the support of others, which is the notion of an undercut.

Definition 2.52 (Undercut) *An undercut⁹ for an argument $\langle \Phi, \alpha \rangle$ is an argument $\langle \Psi, \neg(\phi_1 \wedge \dots \wedge \phi_n) \rangle$ where $\{\phi_1, \dots, \phi_n\} \subseteq \Phi$.*

The another form of conflict is when two arguments have opposite conclusions, which is the notion of a rebuttal.

Definition 2.53 (Rebuttal) *An argument $\langle \Psi, \beta \rangle$ is a rebuttal for an argument $\langle \Phi, \alpha \rangle$ iff $\beta \leftrightarrow \neg\alpha$ is a tautology.*

Let's take a look at an example.

Example 2.6 *Let $\Delta = \{\alpha, \alpha > \beta, \gamma > \neg\beta, \gamma, \neg\alpha\}$. Some arguments are:*

$$\begin{array}{ll} \langle \{\alpha, \alpha > \beta\}, \beta \rangle & \langle \{\gamma, \gamma > \neg\beta\}, \neg\beta \rangle \\ \langle \{\neg\alpha\}, \neg\alpha \rangle & \langle \{\alpha > \beta\}, \neg\alpha \vee \beta \rangle \end{array}$$

Note that, the symbol $>$ denotes material implication in logic.

Unlike some rule-based argumentation frameworks, logic based argumentation does not have any attacks on inference rules, since the only inference rule is general modus ponens which is a strict rule.

⁹same as undermine in ASPIC⁺

Assumption-Based Argumentation Framework

Assumption-Based Argumentation (ABA) [36] was developed as a general-purpose computational framework. Arguments in ABA are defined as backward deductions of a conclusion (using the inference rules of the underlying logic) supported by sets of assumptions. In other words, this can be understood as a strict inference rule with some defeasible assumptions. The assumptions act as the premises of inference rules.

Since ABA is an instance of abstract argument, all the semantic notions for “acceptability” also apply to ABA. In contrast to a number of existing approaches to non-abstract argumentation frameworks, the attack relationship in ABA is reduced to undermining which means the contrary of an assumption. Like ArgTrust, ABA can be represented as a tree as well: the root is the conclusion; for every node, the children of that node are the assumptions supporting the node. All the leaves are either assumptions or τ (empty assumption).

2.2 Non-monotonic Reasoning

Suppose we have a collection of information, there are two methods for extracting more information from the collection: monotonic reasoning and non-monotonic reasoning. The monotonic deductive systems are the familiar log-

ical systems such as classical logic and modal logic. They are called monotonic because the addition of information does not affect the validity of the previous results.

Definition 2.54 (Monotonic System) *Assume we have a collection of information, P , a system is called monotonic iff given addition information P' , $P \vdash \alpha$ then $P \cup P' \vdash \alpha$.*

On the contrary, non-monotonic systems are such that additional information may contradict with previous results. Well-known systems are: Reiter's default systems [73], Clark's negation as failure [29], circumscription [55], the modal system of [58], autoepistemic logic [62] and inheritance systems [80].

Definition 2.55 (Non-monotonic System) *Assume we have a collection of information, P , a system is called non-monotonic iff given addition information P' , $P \vdash \alpha$ then $P \cup P' \not\vdash \alpha$.*

2.2.1 Axioms of Non-monotonic System

Gabbay [40] take the first step to characterize what a non-monotonic system looks like. To be a deductive monotonic logical system, the consequence relation \vdash should satisfy the following conditions:

Reflexivity $\alpha \vdash \alpha$

$$\mathbf{Monotonicity} \quad \frac{\alpha \vdash \gamma}{\alpha \wedge \beta \vdash \gamma}$$

$$\mathbf{Cut} \quad \frac{\alpha \wedge \beta \vdash \gamma \quad \alpha \vdash \beta}{\alpha \vdash \gamma}$$

Gabbay [40] introduces the minimum non-monotonic system, where the consequence relation \vdash should satisfy the following conditions:

$$\mathbf{Reflexivity} \quad \alpha \vdash \alpha$$

$$\mathbf{Restricted Monotonicity} \quad \frac{\alpha \vdash \beta \quad \alpha \vdash \gamma}{\alpha \wedge \beta \vdash \gamma}$$

$$\mathbf{Cut} \quad \frac{\alpha \wedge \beta \vdash \gamma \quad \alpha \vdash \beta}{\alpha \vdash \gamma}$$

Later, Kraus *et al.* [46] summarizes the different axioms among literature, and introduces different systems by combining different axioms.

Kraus *et al.* [46], building on earlier work by Gabbay [40], identified a set of axioms which characterize non-monotonic inference in logical systems, and studied the relationships between sets of these axioms. Their goal was to characterize different kinds of reasoning; to pin down what it means for a logical system to be monotonic or non-monotonic; and — in particular — to be able distinguish between the two. Table 2.2 presents the axioms of [46]. The symbol \vdash encodes a consequence relation, while \models identifies the statements obtainable from the underlying theory. Note that we have

Abbr.	Axiom	Name
Ref	$\alpha \sim \alpha$	Reflexivity
LLE	$\frac{\models \alpha \equiv \beta \quad \alpha \sim \gamma}{\beta \sim \gamma}$	Left Logical Equivalence
RW	$\frac{\models \alpha \leftrightarrow \beta \quad \gamma \sim \alpha}{\gamma \sim \beta}$	Right Weakening
Cut	$\frac{\alpha \wedge \beta \sim \gamma \quad \alpha \sim \beta}{\alpha \sim \gamma}$	Cut
CM	$\frac{\alpha \sim \beta \quad \alpha \sim \gamma}{\alpha \wedge \beta \sim \gamma}$	Cautious Monotonicity ¹⁰
M	$\frac{\models \alpha \leftrightarrow \beta \quad \beta \sim \gamma}{\alpha \sim \gamma}$	Monotonicity
EHD	$\frac{\alpha \sim \beta \leftrightarrow \gamma}{\alpha \wedge \beta \sim \gamma}$	EHD
T	$\frac{\alpha \sim \beta \quad \beta \sim \gamma}{\alpha \sim \gamma}$	Transitivity
CP	$\frac{\alpha \sim \beta}{\bar{\beta} \sim \bar{\alpha}}$	Contraposition
L	$\frac{\alpha_0 \sim \alpha_1, \alpha_1 \sim \alpha_2, \dots, \alpha_{k-1} \sim \alpha_k, \alpha_k \sim \alpha_0}{\alpha_0 \sim \alpha_k}$	Loop
Or	$\frac{\alpha \sim \gamma \quad \beta \sim \gamma}{\alpha \vee \beta \sim \gamma}$	Or

Table 2.2: The axioms from [46] that we will consider.

altered some of the symbols used in [46] to avoid confusion. Equivalence is denoted \equiv (rather than \leftrightarrow), and \leftrightarrow (rather than \rightarrow) denotes the existence of an inference rule.

There are numerous systems studied, we summarize them here:

C A consequence relation is said to be cumulative iff it contains all instances of the Reflexivity axiom and is closed under the inference rules of Left Logical Equivalence, Right Weakening, Cut and Cautious Monotonicity.

CL A consequence relation that satisfies all rules of C and Loop, is said to be loop-cumulative.

P A consequence relation that satisfies all rules of C and Or, is said to be preferential.

CM A consequence relation that satisfies all rules of C and Monotonicity, is said to be cumulative monotonic.

M A consequence relation that satisfies all rules of C and Contraposition, is said to be monotonic.

The first system, C, corresponds to Gabbay's proposal [40]. The second, stronger, system, CL, includes a rule of inference that seems original, and corresponds to models that seem to be more natural. None of those systems assumes, in any essential way the existence of the classical logical connectives, if one allows a finite set of formulas to appear on the left of our symbol \vdash . The systems below assume the classical connectives. The third, stronger,

system, P, has particularly appealing semantics. The fourth system, CM, is stronger than CL but incomparable with P. It provides an example of a monotonic system that is weaker than classical logic. The last one of those systems, M, is stronger than all previous systems and equivalent to classical propositional logic.

Besides the basic axioms, there are a lot of axioms that can be derived from different systems. We will not summarize them here.

2.2.2 Non-monotonic Systems

There are a lot of non-monotonic systems introduced over time. In this section, we will give a brief summary of well-known non-monotonic system.

Clark [29] presents a negation as failure inference rule, whereby $\neg P$ can be inferred if every possible proof of P fails. Moreover, they show that the negation as failure rule only allows to conclude negated facts that could be inferred from the axioms of the completed knowledge base.

Reiter [73] proposes default logic to provide a formal definition of the extensions to an underlying first order theory induced by a set of defaults. They also provide a proof theory by focusing upon a special class of defaults called normal defaults. In addition to providing a proof theory for normal defaults, they obtain the pessimist result that in general the beliefs of a

theory are not recursively enumerable. Finally, they determine conditions under which it is necessary to revise a set of derived beliefs when confronted with some new observations about a world.

McCarthy [55] presents a non-monotonic system, circumscription, to formalize the common sense assumption that things are as expected unless otherwise specified. To implement circumscription in his initial formulation, McCarthy augmented first-order logic to allow the minimization of the extension of some predicates, where the extension of a predicate is the set of tuples of values the predicate is true on. This minimization is similar to the closed world assumption that what is not known to be true is false. Later, [56] use circumscription in an attempt to solve the frame problem [57].

McDermott and Doyle [58] defines a standard language of discourse including the non-monotonic term “consistent”. Moreover, [58] defines the semantics of the language based on models. With the definitions, they justify the soundness and completeness of the theory.

Lin and Shoham [51] first introduce an argument system, contain two kinds of inference rules, namely, monotonic inference rules and non-monotonic inference rules. And they show that most well-known non-monotonic systems, such as default logic, autoepistemic logic, negation as failure and circumscription, can be formulated as special argument systems.

Bondarenko *et al.* [17] continues this line of work, presents an abstract framework for default reasoning which includes Theorist, default logic, logic programming, autoepistemic logic, non-monotonic modal logics, and certain instances of circumscription as special cases.

Billington [16] describes Defeasible Logic, a logic that, as its name implies, differs from classical logic in that it deals with defeasible reasoning. In addition to introducing the logic, [16] shows that defeasible logic satisfies the axioms of reflexivity, cut and cautious monotonicity suggested in [40], thus satisfying what [40] describes as the basic requirements for a non-monotonic system (such a system is equivalent to a cumulative system in [46]).

Governatori *et al.* [43] subsequently established significant links between reasoning in defeasible logic and argumentation-based reasoning. To do this, [43] provides an argumentation system that makes use of defeasible logic as its underlying logic, and shows that the system is compatible with Dung's semantics [65]. [8] presents a framework for defeasible logic, which shows how to tune defeasible logic in order to define variants able to deal with different non-monotonic phenomena. Given Defeasible Logic's close relation to Prolog [63], this line of work is closely related to Defeasible Logic Programming (DeLP) [41], a formalism combining results of Logic Programming and Defeasible Argumentation as we introduced before.

Hunter [44] investigates various consequence relations of deductive argumentation and their satisfaction of various properties. [33] and [34] investigates cumulativity of ASPIC-like structured argumentation frameworks. [32] analyzes cautious monotonicity and cumulative transitivity with respect to Assumption-Based Argumentation.

Prakken and Sartor [70] presents a semantics and proof theory of a system for defeasible argumentation. The semantics of the system is given with a fix-point definition, and the proof theory is stated in dialectical style.

Next, we will give a brief introduction about defeasible logic and its semantics.

A **defeasible theory** D is a pair $(\mathcal{R}, >)$ where \mathcal{R} a finite set of rules, and $>$ a superiority relation on \mathcal{R} . There are three kinds of rules — strict rules, defeasible rules and defeater. Same as the ASPIC⁺ framework, strict rules are rules always true, and defeasible rules are rules that can be defeated by contrary evidence. Defeaters are rules that cannot be used to draw any conclusions, and act as a contrary to the defeasible rules. An example of defeater is “If an animal is heavy then it might not be able to fly”. Formally: $heavy(X) \mapsto \neg flies(X)$. In this case, we do not want to conclude $\neg flies$ if $heavy$, however, we simply want to prevent a conclusion $flies$. The superiority relation is the preference ordering over rules. When $r_1 > r_2$, then

r_1 is called superior to r_2 , and r_2 inferior to r_1 . This expresses that r_1 may override r_2 . The superiority relation is required to be acyclic. Here we only consider essentially propositional rules. Rules containing free variables are interpreted as the set of their ground instances. On the other hand, a defeasible theory D can be defined as $(\mathcal{K}, \mathcal{R}, >)$, where \mathcal{K} is a set of rules with no antecedents.

Before we introduce the details of defeasible theory, we need to define some notions. $A(r)$ denotes the antecedents of a rule r , and $C(r)$ is the conclusion of r . As before, \mathcal{R}_s denotes the set of strict rules, \mathcal{R}_d denotes the set of defeasible rules. In addition, \mathcal{R}_{sd} denotes all the strict rules and defeasible rules. We use the notion $\mathcal{R}[q]$ to represent the set of rules with conclusions q . If q is a literal, $\sim q$ denotes the complementary literal.

Definition 2.56 (Conclusion) *Suppose D is a defeasible theory, a conclusion of D is a tagged literal. In the original defeasible logic there are two tags, ∂ and Δ , that may have positive or negative polarity:*

$+\Delta q$ *which is intended to mean that q is definitely provable in D (i.e., using only strict rules).*

$-\Delta q$ *which is intended to mean that it is proved that q is not definitely provable in D .*

$+∂q$ which is intended to mean that q is defeasibly provable in D .

$-∂q$ which is intended to mean that it is proved that q is not defeasibly provable in D .

Provability is based on the concept of a derivation in D .

Definition 2.57 (Derivation(ambiguity blocked)) *A derivation is a finite sequence $P = (P(1), \dots, P(n))$ of tagged literals satisfying the following conditions.*

$+Δ$ If $P(i + 1) = +Δq$ then

$$\begin{aligned} &\exists r \in \mathcal{R}_s[q] \\ &\forall a \in A(r) : +Δa \in P(1 \dots i). \end{aligned}$$

$-Δ$ If $P(i + 1) = -Δq$ then

$$\begin{aligned} &\forall r \in \mathcal{R}_s[q] \\ &\exists a \in A(r) : -Δa \in P(1 \dots i). \end{aligned}$$

$+∂$ If $P(i + 1) = +∂q$ then either

- (1) $+Δq \in P(1 \dots i)$ or
- (2) (2.1) $\exists r \in \mathcal{R}_{sd}[q] \forall a \in A(r)$
 $+∂a \in P(1 \dots i)$ and
 (2.2) $-Δ \sim q \in P(1 \dots i)$ and

(2.3) $\forall s \in R[\sim q]$ either

(2.3.1) $\exists a \in A(s) : -\partial a \in P(1 \dots i)$ or

(2.3.2) $\exists t \in R_{sd}[q]$ such that

$\forall a \in A(t) : +\partial a \in P(1 \dots i)$ and

$t > s$

$-\partial$ If $P(i+1) = -\partial q$ then

(1) $-\Delta q \in P(1 \dots i)$ and

(2) (2.1) $\forall r \in \mathcal{R}_{sd}[q] \exists a \in A(r)$

$-\partial a \in P(1 \dots i)$ or

(2.2) $+\Delta \sim q \in P(1 \dots i)$ or

(2.3) $\exists s \in R[\sim q]$ such that

(2.3.1) $\forall a \in A(s) : +\partial a \in P(1 \dots i)$ and

(2.3.2) $\forall t \in R_{sd}[q]$ either

$\exists a \in A(t) : -\partial a \in P(1 \dots i)$ or

$t \not> s$

Let's take "flying bird" from [16] as an example to illustrate the defeasible logic.

Example 2.7 Consider the following set of rules:

f_1	$\rightarrow Bird(Tweety)$	<i>Tweety is a bird.</i>
f_2	$\rightarrow Penguin(Tweety)$	<i>Tweety is a penguin.</i>
f_3	$\rightarrow Sick(Tweety)$	<i>Tweety is sick.</i>
r_1	$Penguin(x) \rightarrow Bird(x)$	<i>Penguins are birds.</i>
r_2	$Bird(x) \Rightarrow Flies(x)$	<i>birds usually fly.</i>
r_3	$Penguin(x) \Rightarrow \neg Flies(x)$	<i>Penguins usually do not fly.</i>
r_4	$Bird(x), Sick(x) \mapsto \neg Flies(x)$	<i>Sick birds might not fly.</i>

SUPERIORITY RELATION

$$r_3 > r_2 \text{ and } r_4 > r_2$$

Let $D = (\mathcal{R}, >)$ be the defeasible theory with $\mathcal{R} = \{f_1, r_1, r_2, r_3, r_4\}$. Then $D \vdash +\partial Flies(Tweety)$, which indicates that there is sufficient evidence to defeasibly conclude that Tweety flies. But if Tweety turns out to be a penguin then $(\mathcal{R} \cup \{f_2\}, >) \vdash +\partial \neg Flies(Tweety)$ which indicates that there is sufficient evidence to defeasibly conclude that Tweety does not fly. However if all we know is that Tweety is a sick bird then there is not sufficient evidence to even defeasibly conclude that Tweety flies or does not fly, i.e., $(\mathcal{R} \cup \{f_3\}, >) \vdash -\partial Flies(Tweety)$ and $(\mathcal{R} \cup \{f_3\}, >) \vdash -\partial \neg Flies(Tweety)$.

The above definition is also called the ambiguity blocking approach. In contrast, [8] introduces an ambiguity propagating variant. Let's take an example from [43] to illustrate the difference between the two.

Example 2.8 Consider the following set of rules:

$$\begin{array}{ll} \Rightarrow a & \Rightarrow b \\ \Rightarrow \neg a & a \Rightarrow \neg b \end{array}$$

Here a is ambiguous since we have applicable rules for both a and $\neg a$, and we have no means to decide between them. In a setting where the ambiguity is blocked, b is not ambiguous because we have an applicable rule for b and, at the same time, the rule for $\neg b$ is not applicable since we cannot prove its antecedent. On the other hand, in an ambiguity propagating setting, b is ambiguous because there are rules for both b and $\neg b$, antecedent of the rule for $\neg b$ is ambiguous, and hence the ambiguity is propagated to b . We have proofs in this theory for $-\partial a$, $-\partial \neg a$, $+\partial b$, and $-\partial \neg b$, thus showing the ambiguity blocking behavior of Defeasible Logic.

In the following we introduce an ambiguity propagating variant. The first step is to determine when a literal is “supported” in a defeasible theory. Support for a literal $p(+\Sigma p)$ consists of a monotonic chain of reasoning that would lead us to conclude p in the absence of conflicts. This is defined as:

$+\Sigma$ If $P(i+1) = +\Sigma q$ then

$$\exists r \in \mathcal{R}_{sd}[q]$$

$$\forall a \in A(r) : +\Sigma a \in P(1 \dots i).$$

$-\Sigma$ If $P(i + 1) = -\Sigma q$ then

$$\forall r \in \mathcal{R}_{sd}[q]$$

$$\exists a \in A(r) : -\Sigma a \in P(1 \dots i).$$

A literal that is defeasibly provable is supported, but a literal may be supported even though it is not defeasibly provable. Thus support is a weaker notion than defeasible provability. For example, given two rules $\Rightarrow p$ and $\Rightarrow \neg p$, both p and $\neg p$ are supported, but neither is defeasibly provable. We say that p is ambiguous. In general, a literal is ambiguous if there is a chain of reasoning that supports a conclusion that p is true, and another that supports that $\sim p$ is true.

We can achieve ambiguity propagation behavior by making a minor change to the inference condition for $+\partial$: instead of requiring that every attack on p be inapplicable in the sense of $-\partial$, now we require that the rule for $\sim p$ be inapplicable because one of its antecedents cannot be supported. Here is the formal definition:

$+\partial_{ap}$ If $P(i + 1) = +\partial_{ap} q$ then either

$$(1) +\Delta q \in P(1 \dots i) \text{ or}$$

$$(2) \quad (2.1) \exists r \in \mathcal{R}_{sd}[q] \forall a \in A(r)$$

$$+\partial_{ap} a \in P(1 \dots i) \text{ and}$$

(2.2) $-\Delta \sim q \in P(1 \dots i)$ and

(2.3) $\forall s \in R[\sim q]$ either

(2.3.1) $\exists a \in A(s) : -\Sigma a \in P(1 \dots i)$ or

(2.3.2) $\exists t \in R_{sd}[q]$ such that

$\forall a \in A(t) : +\Sigma a \in P(1 \dots i)$ and

$t > s$

$-\partial_{ap}$ If $P(i+1) = -\partial_{ap}q$ then

(1) $-\Delta q \in P(1 \dots i)$ and

(2) (2.1) $\forall r \in \mathcal{R}_{sd}[q] \exists a \in A(r)$

$-\partial_{ap}a \in P(1 \dots i)$ or

(2.2) $+\Delta \sim q \in P(1 \dots i)$ or

(2.3) $\exists s \in R[\sim q]$ such that

(2.3.1) $\forall a \in A(s) : +\Sigma a \in P(1 \dots i)$ and

(2.3.2) $\forall t \in R_{sd}[q]$ either

$\exists a \in A(t) : -\Sigma a \in P(1 \dots i)$ or

$t \not> s$

Let's use Example 2.8 to say the above notions.

Example 2.9 Consider the defeasible theory of Example 2.8, we have $+\Sigma a$, $+\Sigma -a$, $+\Sigma b$ and $+\Sigma -b$ showing that there are chains of reasoning supporting

a , $\neg a$, b and $\neg b$. Moreover we can derive $-\partial_{ap}a$, $-\partial_{ap}\neg a$, $-\partial_{ap}b$ and $-\partial_{ap}\neg b$ showing that the resulting logic exhibits an ambiguity propagating behavior. In fact b is now ambiguous, and its ambiguity depends on the ambiguity of a .

Next, we will give a formal definition of argumentation framework based on defeasible logic.

Definition 2.58 (Argument) *An argument for a literal p based on a set of rules \mathcal{R} is a (possibly infinite) tree with nodes labeled by literals such that the root is labeled by p and for every node with label h :*

- *If b_1, \dots, b_n label the children of h then there is a rule in \mathcal{R} with body b_1, \dots, b_n and head h .*
- *If this rule is a defeater then h is the root of the argument.*
- *The arcs in a proof tree are labeled by the rules used to obtain them.*

Let's take a look at an example from [43] to illustrate the definition.

Example 2.10 *Consider the following defeasible theory D :*

$$\vdash a$$

$$a \vdash b$$

$$\Rightarrow \neg b$$

Then $\vdash a \vdash b$ is not an argument. The reason is that, as said before, defeaters are only used to prevent conclusions, but do not provide positive evidence. In the example, we have evidence against $\neg a$ (by the first defeater), but no evidence for a . Therefore the second defeater cannot be used since to do so we would need evidence for a . The proof theory of defeasible logic was defined in agreement with this reading, therefore $D \vdash +\partial\neg b$ and $D \vdash +\partial_{ap}\neg b$.

Given a defeasible theory D , the set of arguments that can be generated from D is denoted by $\mathcal{A}rgs_D$. Depending on the rules used, we have different notions of arguments:

- A supportive argument is a finite argument in which no defeater is used.
- A strict argument is an argument in which only strict rules are used.
- An argument that is not strict is called defeasible.

Let's take a look at an example to understand the different notions.

Example 2.11 Consider the following defeasible theory $D = (\mathcal{R}, >)$ with $\mathcal{R} = \{\Rightarrow d; \rightarrow e; \Rightarrow f; a, \neg b \Rightarrow c; e \rightarrow a; f \vdash b; d \Rightarrow \neg b\}$ and the superiority

relation is empty. Now we can construct the following arguments:

$$A = [(\rightarrow e \rightarrow a, \Rightarrow d \Rightarrow \neg b) \Rightarrow c]$$

$$B = [\Rightarrow f \mapsto b]$$

$$C = [\rightarrow e \rightarrow a]$$

Then A is a supportive argument for c , but not a strict argument. B is an argument for b that is not supportive. C is a strict supportive argument for a .

Now we can characterize the definite conclusions of defeasible logic in argumentation-theoretic terms.

Proposition 2.2 *Let D be a defeasible theory and p be a literal.*

- $D \vdash +\Delta p$ iff there is a strict supportive argument for p in Args_D
- $D \vdash -\Delta p$ iff there is no (finite or infinite) strict argument for p in Args_D
- $D \vdash +\Sigma p$ iff there is a supportive argument for p in Args_D
- $D \vdash -\Sigma p$ iff there is no (finite or infinite) argument ending with a supportive rule for p in Args_D

Let's use the above example to explain this:

Example 2.12 *For the theory D in Example 2.11 we have the following:*

$$D \vdash +\Delta a$$

$$D \vdash +\Sigma c$$

$$D \vdash -\Delta f \text{ (there is no strict rule with head } f \text{)}$$

$$D \vdash -\Sigma b \text{ (there is no strict or defeasible rule with head } b \text{)}$$

An argument A attacks a defeasible argument B if a conclusion of A is the complement of a conclusion of B , and that conclusion of B is not part of a strict sub-argument of B . A set of arguments S attacks a defeasible argument B if there is an argument A in S that attacks B . A defeasible argument A is supported by a set of arguments S if every proper sub-argument of A is in S . A defeasible argument A is undercut by a set of arguments S if S supports an argument B attacking a proper non-strict sub-argument of A . That an argument A is undercut by S means that we can show that some premises of A cannot be proved if we accept the arguments in S .

Example 2.13 *For the theory D in Example 2.11, the arguments A and B attacks each other. The argument A is undercut by the set $S = \{\Rightarrow f\}$:*

- S supports the argument B
- B attacks a proper sub-argument $A_{sub} = [\Rightarrow d \Rightarrow \neg b]$ of A .

The main purpose of argumentation semantics is to find out the justified arguments/conclusions. Given the notion of “acceptable”, which we will define later, we can define justified arguments and justified conclusions.

Definition 2.59 (Justified Argument) *Let D be a defeasible theory. We define J_i^D as follows.*

- $J_0^D = \emptyset$
- $J_{i+1}^D = \{a \in \mathcal{A}rgs_D \mid a \text{ is acceptable w.r.t. } J_i^D\}$.

The set of justified arguments in a defeasible theory D is $J\mathcal{A}rgs^D = \cup_{i=1}^{\infty} J_i^D$.

A literal p is justified if it is the conclusion of a supportive argument in $J\mathcal{A}rgs^D$.

That an argument is justified means that it is provable using the underlying logic (+ tag). However, defeasible logic is more expressive, it is able to say when a conclusion is non-provable (− tag). This is captured by the concept rejected. Given the notion of “rejectable”, which we will define later, we can define rejected arguments and rejected conclusions.

Definition 2.60 (Rejected Argument) *Let D be a defeasible theory and T be a set of arguments. We define $R_i^D(T)$ as follows.*

- $R_0^D(T) = \emptyset$

- $R_{i+1}^D(T) = \{a \in \mathcal{A}rgs_D \mid a \text{ is rejectable by } R_i^D(T) \text{ and } T\}$.

The set of rejected arguments in a defeasible theory D w.r.t. T is $R\mathcal{A}rgs^D(T) = \cup_{i=1}^{\infty} R_i^D(T)$. We say that an argument is rejected if it is rejected w.r.t. $J\mathcal{A}rgs^D$. A literal p is rejected by T if there is no argument in $\mathcal{A}rgs_D - R\mathcal{A}rgs^D(T)$, the top rule of which is a strict or defeasible rule with head p . A literal is rejected if it is rejected by $J\mathcal{A}rgs_D$.

Note that it is possible for a literal to be neither justified nor rejected. The situation is similar to defeasible logic, where we may have both $D \not\vdash +\partial p$ and $D \not\vdash -\partial p$.

Now we show how to modify Dung's definition of acceptable argument in order to suit defeasible logic with ambiguity propagation (grounded semantic).

Definition 2.61 (Ambiguity Propagation Acceptable) *An argument A for p is "acceptable" w.r.t a set of arguments S if A is finite, and*

1. A is strict, or
2. every argument attacking A is attacked by S .

A defeasible argument is assessed as valid if those arguments whose counter-arguments have been undermined by arguments that have already been assessed as valid. "Rejectable" is defined as

Definition 2.62 (Ambiguity Propagation Rejectable) *An argument A is “rejectable” by sets of arguments S and T when A is not strict, and either*

1. *a proper sub-argument of A is in S , or*
2. *it is attacked by a finite argument.*

Note that T is not used in this definition. An argument can be rejectable for two reasons:

1. part of the argument has already been rejectable
2. there is a competing argument.

Example 2.14 *For the theory D in Example 2.11, under grounded semantic, the argument A is acceptable w.r.t. $S = \{\Rightarrow d \Rightarrow \neg b\}$ because S attacks B , the only argument attacking A .*

The argument $\Rightarrow d \Rightarrow \neg b$ is rejectable by any sets S and T because it is attacked by the argument B .

There is a match between justified/rejected conclusion and defeasible provability in ambiguity propagating defeasible logic.

Proposition 2.3 *Let D be a defeasible theory, p be a literal, and T be a set of arguments.*

1. $D \vdash +\partial_{ap}p$ iff p is justified under grounded semantics.
2. $D \vdash -\partial_{ap}p$ iff p is rejected by T under grounded semantics.

The following example demonstrate the concepts and the proposition.

Example 2.15 *For the theory D in Example 2.8, under grounded semantic, we have*

$$J_0^D = \emptyset$$

$$J_1^D = J_0^D = J\mathcal{A}rgs^D.$$

We also have

$$R_0^D(T) = \emptyset$$

$$R_1^D(T) = \{\Rightarrow a, \Rightarrow \neg a, \Rightarrow b, \Rightarrow a \Rightarrow \neg b\}$$

$$R_2^D(T) = R_1^D(T) = R\mathcal{A}rgs^D(T).$$

The conclusions a , $\neg a$, b $\neg b$ are rejected, that corresponds to the non-derivability results with the ambiguity propagating $D \vdash -\partial_{ap}a$, $D \vdash -\partial_{ap}\neg a$, $D \vdash -\partial_{ap}b$ and $D \vdash -\partial_{ap}\neg b$.

Next, we show how to modify Dung's definition of acceptable argument in order to suit defeasible logic with ambiguity blocking (defeasible semantic).

Definition 2.63 (Ambiguity Blocking Acceptable) *An argument A for p is “acceptable” w.r.t a set of arguments S if A is finite, and*

1. *A is strict, or*
2. *every argument rebutting A is undercut by S .*

A defeasible argument is assessed as valid if we can show that the premises of all arguments attacking it cannot be proved if we consider valid the arguments in S . The “rejectable” is defined as

Definition 2.64 (Ambiguity Blocking Rejectable) *An argument A is “rejectable” by sets of arguments S and T when A is not strict and*

1. *a proper sub-argument of A is in S , or*
2. *it is attacked by an argument supported by T .*

The simple existence of a competing argument is not enough to state that an argument is rejectable. The attacking argument must be supported by the set of justified arguments.

Now the relation between justified/rejected conclusion and defeasible provability in ambiguity blocking defeasible logic.

Proposition 2.4 *Let D be a defeasible theory, p be a literal.*

1. $D \vdash +\partial p$ iff p is justified under defeasible semantics.
2. $D \vdash -\partial p$ iff p is rejected by $JArgs^D$ under defeasible semantics.

Again, we are using Example 2.8 to demonstrate the concepts and the proposition.

Example 2.16 *For the theory D in Example 2.8, under defeasible semantic, we have*

$$J_0^D = \emptyset$$

$$J_1^D = \{\Rightarrow b\}$$

$$J_2^D = J_1^D = JArgs^D.$$

We also have

$$R_0^D(T) = \emptyset$$

$$R_1^D(T) = \{\Rightarrow a, \Rightarrow \neg a, \Rightarrow a \Rightarrow \neg b\}$$

$$R_2^D(T) = R_1^D(T) = RArgs^D(T).$$

The conclusion b is justified, and a , $\neg a$, $\neg b$ are rejected. That corresponds to the results with the ambiguity blocking $D \vdash -\partial_{ap}a$, $D \vdash -\partial_{ap}\neg a$, $D \vdash +\partial_{ap}b$ and $D \vdash -\partial_{ap}\neg b$.

It is worth noting the differences between defeasible semantics and grounded semantics. In both cases the set of justified arguments is defined by Definition 2.59, but with different notions of acceptability. Under the grounded semantics, any argument attacking an acceptable argument A must be countered by an attack from S . Under the defeasible semantics the kind of counter required is different: the counter-argument must attack a sub-argument, not the conclusion, and the counter-argument need only be supported by S , not be a member of S as in the grounded semantics.

There are similar differences in the definitions of rejected arguments. Under the grounded semantics, an argument is rejected if it is attacked by any finite argument. Under the defeasible semantics, an argument is rejected if it is attacked by a (possibly infinite) argument supported by T .

The defeasible semantics justifies more arguments, but rejects fewer arguments, than the grounded semantics. Thus, although both semantics are fundamentally skeptical, the defeasible semantics can be considered more credulous than the grounded semantics.

Proposition 2.5 *Fix a defeasible theory D . Let A be an argument, and p be a literal.*

1. *If A is justified under the grounded semantics then A is justified under*

the defeasible semantics.

2. *If A is rejected under the defeasible semantics then A is rejected under the grounded semantics.*
3. *If p is justified under the grounded semantics then p is justified under the defeasible semantics.*
4. *If p is rejected under the defeasible semantics then p is rejected under the grounded semantics.*

2.3 Summary

In this chapter, I have reviewed argumentation frameworks and non-monotonic reasoning. Starting with abstract argumentation frameworks, I introduced work on structured argumentation frameworks. In structured argumentation, much of the focus is on the way that arguments are constructed from some logic-based language. Early examples of such work are [47] and [52], and the most influential argumentation structured system to date is ASPIC⁺, as mainly discussed above. Within this broad categorization, one can distinguish three classes of argumentation system. First, there are systems like [5] and [15] which are based on a variant of classical logic. Because the logic defines the set of inference rules, an argument is defined as (just) a pair where

the first item is a minimal consistent set of formula that proves the second item (which is a formula). Attacks between arguments are then the result of conflicts between propositions — for example where one argument asserts that a proposition is true, and another asserts that it is false. The second broad class of structured argumentation systems are those like DeLP [41], ArgTrust [77] and ASPIC⁺ [60] where a logical language is augmented with domain-specific inference rules. In such systems, the rules are considered part of the argument, and in some systems it is possible to attack the application of rules, recognizing another kind of conflict between arguments. The final broad group of structured argumentation systems are assumption-based systems, exemplified by [36], where some formula are explicitly identified as assumptions that may turn out to be incorrect. Conflicts then center around whether assumptions hold or not.

Gabbay [40] take the first step to characterize what a non-monotonic system looks like. Kraus *et al.* [46], building on [40], identified a set of axioms which characterize non-monotonic inference in logical systems, and studied the relationships between sets of these axioms. Besides that, numbers of non-monotonic systems have been introduced. Among those, defeasible logic establishes significant links between reasoning in defeasible logic and argumentation-based reasoning.

In the following chapter, I will continue the line of work started in [23], one in which we study another variation of ASPIC⁺ in which only defeasible elements are present.

Chapter 3

Motivation

There has been increasing interest in formal argumentation in recent years. While much work has concerned abstract argumentation, systems for rule-based (or structured) argumentation are perhaps more interesting from a knowledge representation perspective.

Work on structured argumentation similarly falls into two broad camps, one which distinguishes strict and defeasible components and one which does not. What has come to be known as “logic-based” argumentation, for example [6, 15], builds the mechanism of argumentation on top of a classical logic. Rather than encode defeasibility as in default logic, by stating the exceptions to rules, logic-based argumentation allows any inference to be questioned, for example by deriving the opposite conclusion. In contrast, “rule-based” argumentation, including systems like ASPIC [67, 60] and defeasible logic programming [41] maintain the distinction between strict and

defeasible knowledge, and handle those two kinds of rule differently. For example, in ASPIC only the conclusions of defeasible rules can be challenged.

Amgoud *et al.* [3] was the original description of the ASPIC system, which attempted to generalize existing structured argumentation systems. As a result, most of the then existing systems could be considered to be specializations of ASPIC. However, the initial system was not without flaws. Most importantly, [21] pointed out that ASPIC may lead to some non-intuitive results, suggested that all argumentation frameworks must satisfy three rationality postulates in order to avoid these anomalies, and showed how ASPIC could be modified to satisfy them. [67] then presented an extension of ASPIC, called ASPIC⁺, which satisfies the rationality postulates under a small number of restrictions. [2] and [69] provide further discussion of the approach.

Modgil and Prakken [60] provided a modification of the ASPIC⁺ framework, giving a more general structured framework for argumentation with preferences. It is this version of ASPIC⁺ that is our starting point. [19] and [84] presented some examples where ASPIC-like systems could lead to non-intuitive results and gave solutions. Finally, [23] looked at a new variation of ASPIC⁺ which still satisfies the rationality postulates while loosening the restriction on rebutting attacks that ASPIC⁺ requires to satisfy the

rationality postulates.

This dissertation can be seen as a continuation of the line of work started in [23], one in which I study another variation of ASPIC⁺ in which only defeasible elements are present. I call the defeasible only system as ASPIC_D⁺. There are two, related, reasons for our investigation. The first is to create a simpler system than ASPIC⁺. As mentioned, the original ASPIC system was intended [3] as a generalization of existing argumentation systems, and this makes it complex. This complexity is inherited by ASPIC⁺. I believe that all the complexity has its place, but is not always needed, and I wanted to explore a system that was simpler to use and simpler to implement¹. Removing strict rules is one way to simplify ASPIC⁺, not only reducing the number of different components of the system, but also removing constraints like restricted rebut and the need to complete the knowledge base with the transposition of every strict rule. The second reason is that strict rules are exactly the component of ASPIC and ASPIC⁺ that lead to the anomalies studied in papers like [21] and [23]. As a result, it seems worthwhile to try to establish what exactly what one gains from including strict rules (and thus having to deal with all the anomalies) by studying what one gives up by only having defeasible rules.

¹It is no accident that the system in [75] is exactly the system that I explore here.

There are several research questions related to ASPIC_D^+ argumentation framework:

RQ1 Since ASPIC_D^+ is an subset of ASPIC^+ , do we lose any expressiveness in removing the strict parts of ASPIC^+ argumentation framework?

RQ2 Do the ASPIC^+ and ASPIC_D^+ versions of the same theory have the same justified conclusions?

RQ3 What is the advantage, if any, of removing the strict elements from ASPIC_D^+ ?

Modgil and Prakken [60] introduce the idea of well-defined ASPIC^+ argumentation frameworks. A well-defined ASPIC^+ framework satisfies closure under transposition/contraposition, axiom consistent and well-formed. All well-defined ASPIC^+ frameworks will satisfy the rationality postulates. Furthermore, most research constrains ASPIC^+ to be well-defined. Therefore, I will start with well-defined ASPIC^+ argumentation frameworks. Chapter 4 investigate the relationship between well-defined ASPIC^+ argumentation frameworks and its converted ASPIC_D^+ argumentation framework, answering the three research questions when ASPIC^+ is well-defined. Turning to ill-defined theories, in general, ASPIC^+ argumentation framework can not handle the ill-defined cases. There are three sub-class of ill-defined cases: (1)

not closed under transposition/contraposition; (2) axiom inconsistent; (3) not well-formed. Chapter 5 investigates the first two cases in turn. For each cases, I will give a way to make the $ASPIC^+$ theories well-defined, and compare the justified conclusions with $ASPIC_D^+$ theories, which will answer RQ2. By investigating the relationship between $ASPIC^+$ and $ASPIC_D^+$ theories, I will answer RQ1 and RQ3. The last cases will be discussed in Chapter 7. After that, I answer the above research questions in a different way using the non-monotonic axioms of [46]. As we know, reasoning argumentation frameworks is non-monotonic. So I investigate whether or not the two argumentation frameworks satisfy the non-monotonic axioms introduced in [46]. Chapter 6 considers a family of the consequence relation in the context of argumentation-based reasoning, and describe which axioms are satisfied by $ASPIC^+$ theories and $ASPIC_D^+$ theories under all of these consequence relation.

Chapter 4

A Purely Defeasible System

In this chapter, we introduce a simpler system than ASPIC⁺, which only contains the defeasible elements. By removing strict elements, we also remove the constraints like restricted rebut and the need to complete the knowledge base with the transposition of every strict rule. Intuitively, removing strict elements may mean we lose expressiveness as well. In this chapter, we will answer the question of whether we lose anything, and whether we have the same justified conclusions under grounded semantics and other semantics, by removing strict parts of a well-defined ASPIC⁺ system.

4.1 ASPIC_D⁺: a purely defeasible system

In this section we introduce ASPIC_D⁺, a purely defeasible subset of ASPIC⁺, that is a subset of ASPIC⁺ which has no strict elements. The full definition of ASPIC_D⁺ starts from a variation on the ASPIC⁺ notion of an argumen-

tation system where there are only defeasible elements:

Definition 4.1 (ASPIC_D⁺ Argumentation System) *An argumentation system is a triple $AS_D = \langle \mathcal{L}, \mathcal{R}_d, n \rangle$ where:*

- \mathcal{L} is a logical language closed under negation $\bar{\cdot}$.
- \mathcal{R}_d is a set of defeasible inference rules of the form $\phi_1, \dots, \phi_n \Rightarrow \phi$ (where ϕ_i, ϕ are meta-variables ranging over wff in \mathcal{L}).
- $n : \mathcal{R}_d \mapsto \mathcal{L}$ is a naming convention for defeasible rules.

Definition 4.2 (ASPIC_D⁺ Knowledge Base) *A knowledge base in an argumentation system $\langle \mathcal{L}, \mathcal{R}_d, n \rangle$ is a set \mathcal{K}_p of ordinary premises.*

Definition 4.3 (ASPIC_D⁺ Argumentation Theory) *An argumentation theory AT_D is a pair $\langle AS_D, \mathcal{K}_p \rangle$ of an argumentation system AS_D and a set of ordinary premises \mathcal{K}_p .*

Arguments in ASPIC_D⁺ are then defined as:

Definition 4.4 (ASPIC_D⁺ Argument) *An argument A on the basis of an argumentation theory $\langle \langle \mathcal{L}, \mathcal{R}_d, n \rangle, \mathcal{K} \rangle$ is:*

1. ϕ if $\phi \in \mathcal{K}$ with: $\text{Prem}(A) = \{\phi\}$; $\text{Conc}(A) = \{\phi\}$; $\text{Sub}(A) = \{A\}$;
 $\text{TopRule}(A) = \text{undefined}$.

2. $A_1, \dots, A_n \Rightarrow \phi$ if A_i are arguments such that there exists a defeasible rule $\text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi$ in \mathcal{R}_d . $\text{Prem}(A) = \text{Prem}(A_1) \cup \dots \cup \text{Prem}(A_n)$; $\text{Conc}(A) = \phi$; $\text{Sub}(A) = \text{Sub}(A_1) \cup \dots \cup \text{Sub}(A_n) \cup \{A\}$; $\text{TopRule}(A) = \text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi$.

Note the similarity to Definition 2.18, but also note that there are no strict rules or axioms invoked in the definition, so there are no strict or firm ASPIC_D^+ arguments.

Since any ASPIC_D^+ argumentation theory is an ASPIC^+ argumentation theory with an empty set of strict rules and an empty set of axioms, we have:

Proposition 4.1 *For a given language \mathcal{L} , \mathbf{AT}_D , the set of all possible ASPIC_D^+ argumentation theories, is a subset of \mathbf{AT} , the set of all possible ASPIC^+ argumentation theories.*

Proof *Pick any ASPIC_D^+ theory $AT_D \in \mathbf{AT}_D$. By definition this is a pair $\langle AS_D, \mathcal{K}_p \rangle$ where $AS_D = \langle \mathcal{L}, \mathcal{R}_d, n \rangle$. It is also an ASPIC^+ theory $AT \in \mathbf{AT}$ where $AT = \langle AS, \mathcal{K}_p \rangle$ (an ASPIC^+ theory with no axioms) and $AS = \langle \mathcal{L}, \mathcal{R}_d, n \rangle$ (an ASPIC^+ theory with no strict rules). Having made no specific assumptions about the composition of AT_D , the result holds for all possible ASPIC_D^+ theories. \square*

However, despite the fact that the set of all possible ASPIC_D^+ theories is

a subset of all possible ASPIC⁺ theories, we can translate any specific ASPIC⁺ theory into a specific ASPIC_D⁺ theory. We demonstrate this by defining a translation:

Definition 4.5 (Defeasible version) ASPIC_D⁺ theory AT_D is the defeasible version of ASPIC⁺ theory $AT = \langle AS, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ where $AS = \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_p, n \rangle$ iff:

- $AS_D = \langle \mathcal{L}, \mathcal{R}_d \cup \mathcal{R}_{d'}, n' \rangle$, where $\mathcal{R}_{d'} = \{ \phi_1, \dots, \phi_n \Rightarrow \phi \mid \phi_1, \dots, \phi_n \rightarrow \phi \in \mathcal{R}_s \}$ and n' is n extended to name all the rules in $\mathcal{R}_{d'}$.
- $AT_D = \langle AS_D, \mathcal{K}_p \cup \mathcal{K}_{p'} \rangle$, where $\mathcal{K}_{p'} = \{ \phi \mid \phi \in \mathcal{K}_n \}$

If AT_D is the defeasible version of AT , we call AS_D the defeasible version of AS and write $AT_D = \text{def}(AT)$ and $AS_D = \text{def}(AS)$. We call the set of rules $\mathcal{R}_{d'}$ that were strict in AT the set of converted rules, and the set of premises $\mathcal{K}_{p'}$ that were axioms in AT are the set of converted premises. The defeasible version of an argument $A \in \mathcal{A}(AT)$ is an argument $A_D \in \mathcal{A}(AT_D)$ such that every axiom in A is replaced by the corresponding converted premise, and every strict rule in A is replaced by the corresponding converted rule.

In other words, AT_D is the defeasible version of AT , if every axiom of AT becomes an ordinary premise of AT_D , and every strict rule in AT becomes a defeasible rule of AT , while all other components of AT are unchanged.

Note that we will also use the term “translated rule” to mean “converted rule”, as defined here, and the term “translated premise” to mean “converted premise”. Note also that while the translation process allows for the naming of the converted strict rules, so that they are treated exactly like any other defeasible rule, no ASPIC⁺ theory AT will contain undercutters for such rules, because they are strict, and so no ASPIC _{D} ⁺ theory AT_D that is the result of translating AT will make use of the names.

Now, given a preference order \preceq over the elements of an ASPIC⁺ theory AT , we will need to specify the preference order \preceq_D over the defeasible version of the theory. One way to specify \preceq_D is as follows in terms of the pre-orderings over the rules and premises of AT_D .

Definition 4.6 (Strict-first preference ordering) *Given an ASPIC⁺ theory $AT = \langle \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, n \rangle, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ and preference orders \leq and \leq' over the defeasible rules and premises of that theory, the strict-first preference orderings \leq_{sf} and \leq'_{sf} over the rules and premises of the defeasible version of AT , $AT_D = \langle \langle \mathcal{L}, \mathcal{R}_d \cup \mathcal{R}_{d'}, n' \rangle, \mathcal{K}_p \cup \mathcal{K}_{p'} \rangle$ are such that:*

- *For every $r, r' \in \mathcal{R}_d$, $r \leq_{sf} r'$ iff $r \leq r'$, and for every $k, k' \in \mathcal{K}_p$, $k \leq'_{sf} k'$ iff $k \leq' k'$.*
- *For any $r \in \mathcal{R}_d$ and any $r' \in \mathcal{R}_{d'}$, $r <_{sf} r'$, and for every $r', r'' \in \mathcal{R}_{d'}$,*

$$r' =_{sf} r''.$$

- For any $k \in \mathcal{K}_p$ and any $k' \in \mathcal{K}_{p'}$, $k <_{sf}' k'$, and for every $k', k'' \in \mathcal{K}'_p$,

$$k' =_{sf}' k''.$$

where $r =_{sf} r'$ if $r \leq_{sf} r'$ and $r' \leq_{sf} r$, $r <_{sf} r'$ if $r \leq_{sf} r'$ and $r' \not\leq_{sf} r$,
 $k =_{sf}' k'$ if $k \leq_{sf}' k'$ and $k' \leq_{sf}' k$, and $k <_{sf}' k'$ if $k \leq_{sf}' k'$ and $k' \not\leq_{sf}' k$

In other words, all the elements of AT_D that were defeasible in AT have the same preference order as in AT , and all elements that were strict in AT are strictly higher in the preference order than any element that was defeasible in AT .

The notion of attack in $ASPIC_D^+$ differs from that in $ASPIC^+$ in that there is no restriction on rebut, and any rule can be undercut:

Definition 4.7 (ASPIC_D⁺ Attack) *An argument A attacks an argument B iff A undermines, rebuts or undercuts B , where:*

- A undermines B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' = \phi \in \text{Prem}(B)$.
- A rebuts B (on B') iff $\text{Conc}(A) = \bar{\phi}$ for some $B' \in \text{Sub}(B)$.
- A undercuts B (on B') iff $\text{Conc}(A) = \overline{n(r)}$ for some $B' \in \text{Sub}(B)$.

Of course, since there are no strict rules in $ASPIC_D^+$, were we to define rebut as being restricted, it would make no difference to the attacks that were gen-

erated. With these definitions, we can once again combine the definition of arguments, attack relations and the preference ordering from Definition 2.21 to get notions of a structured argumentation framework and defeat that are the same as for ASPIC⁺.

Now, much of this chapter is concerned with the relationship between ASPIC_D⁺ and ASPIC⁺. To begin to understand this relationship, consider this version of Example 2.1:

Example 4.1 *Consider the ASPIC_D⁺ argumentation system AS_D which is the defeasible version of the system AS in Example 2.1. We have the theory $AT_{4.1}$:*

$$\mathcal{R}_d = \{a \Rightarrow b; \bar{c} \Rightarrow d; e \Rightarrow f; a \Rightarrow \overline{nd}; d, f \Rightarrow \bar{b}\}$$

$$\mathcal{K}_p = \{a; \bar{c}; e; \bar{e}\}$$

and $n(\bar{c} \Rightarrow d) = nd$. We can construct the arguments:

$$\begin{aligned} A_1 &= [a]; & A_2 &= [A_1 \Rightarrow b]; & A_3 &= [A_1 \Rightarrow \overline{nd}]; \\ B_1 &= [\bar{c}]; & B_2 &= [B_1 \Rightarrow d]; \\ B'_1 &= [e]; & B'_2 &= [B'_1 \Rightarrow f]; & B &= [B_2, B'_2 \Rightarrow \bar{b}]; \\ C &= [\bar{e}]; \end{aligned}$$

Compared with the attacks in Example 2.1, there is an additional attack here: A_2 rebuts B . With the same preference ordering \preceq over arguments as in Example 2.1 ($A_2 \prec B; C \prec B; C \prec B'_1; C \prec B'_2$), the defeat relation remains the same.

Having introduced ASPIC_D^+ and given some examples of its use, we now consider some of the basic properties of the system.

4.2 Rationality postulates

Recall the rationality postulates introduced by Caminada and Amgoud [21].

The postulates are as follows:

Definition 4.8 (Rationality postulates) *For an argumentation theory AT and associated argumentation framework AF with a set of sceptical justified conclusions C , and extensions, under a given semantics E_1, \dots, E_n :*

Postulate 1 (Closure) AF satisfies closure, also called closure under strict rules, iff:

1. $\text{Concs}(E_i) = \text{Cl}_s(\text{Concs}(E_i))$, for all $1 \leq i \leq n$
2. $C = \text{Cl}_s(C)$

Postulate 2 (Direct consistency) AF satisfies direct consistency iff:

1. $\text{Concs}(E_i)$ is consistent for all $1 \leq i \leq n$
2. C is consistent

Postulate 3 (Indirect consistency) AF satisfies indirect consistency iff:

1. $Cl_S(\text{Concs}(E_i))$ is consistent for all $1 \leq i \leq n$
2. $Cl_S(C)$ is consistent

where $\text{Concs}(\cdot)$ denotes the conclusions of a set of arguments, and $Cl_S(\cdot)$ denotes closure under strict rules.

Now we consider how ASPIC_D^+ behaves with respect to the rationality postulates. Without strict rules, two of the postulates follow immediately.

Proposition 4.2 (Closure under Strict Rules) *The conclusions of any extensions of an ASPIC_D^+ theory are closed under strict rules.*

Proof *With no strict rules, the conclusion follows immediately. \square*

Proposition 4.3 (Direct Consistency) *The conclusions of any extension of an ASPIC_D^+ theory are consistent.*

Proof *Suppose the conclusions of one of the extensions E are inconsistent, i.e., there exist two arguments $A, A' \in E$ such that $\text{Conc}(A) = \overline{\text{Conc}(A')}$. If $\text{Conc}(A) \in \mathcal{K}$, by Definition 4.7, then A' undermines A . On the other hand, if $\text{Conc}(A) \notin \mathcal{K}$, by Definition 4.7, then A' rebuts A . Either way, A' attacks A . Similarly, A attacks A' . According to Definition 2.28, at least one of the attack relations is a defeat relation. Therefore, E is not conflict-free and thus*

E is not an extension under Dung's semantics. The contradiction defeats the assumption of inconsistency and the result holds. \square

Proposition 4.4 (Indirect Consistency) *The closure under strict rules of the conclusions of any extension of an ASPIC_D^+ theory is consistent.*

Proof *With no strict rules, this follows immediately from Proposition 4.3.*

\square

Despite the triviality of two of the results, it is worth noting that there are no restrictions on the semantics for which these results hold — they hold for all the complete-based semantics. Thus ASPIC_D^+ satisfies all three rationality postulates without any limitation on the knowledge or rules in a given theory. In contrast, while ASPIC^+ satisfies all three rationality postulates, this is under the condition that any argumentation theory is well defined — that is it is axiom consistent and any strict rules that it contains are closed under transposition. The requirement that strict rules are closed under transposition means that, in order to ensure that the rationality postulates are satisfied, anyone writing an argumentation theory in ASPIC^+ has to carefully craft the set of strict rules to include all the rules implied by closure under transposition¹. There is no such restriction if ASPIC_D^+ is

¹Or must automatically complete the set of rules using some computer programs.

used. Furthermore, since there is a defeasible version of any ASPIC⁺ theory, translating that theory into ASPIC_D⁺ provides another route to ensuring compliance with the rationality postulates. Thus in the sense of compliance with the rationality postulates, we can argue that ASPIC_D⁺ extends what is possible in ASPIC⁺².

ASPIC_D⁺ has an additional advantage over ASPIC⁺— since we don't have to modify ASPIC_D⁺ theories to make them conform to the rationality postulates, we can avoid problems where ensuring rationality in ASPIC⁺ can change the conclusions of the original theory. Such a case is shown in the following example:

Example 4.2 *We start with theory AT₄:*

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}; \bar{b} \Rightarrow d\} & \mathcal{R}_s &= \{a' \rightarrow b'; b \rightarrow c\} \\ \mathcal{K}_n &= \{a; a'\} & \mathcal{K}_p &= \{\bar{d}\} \end{aligned}$$

We convert this theory into a rational one by making the strict rules closed under transposition.

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}; \bar{b} \Rightarrow d\} & \mathcal{R}_s &= \{a' \rightarrow b'; b \rightarrow c; \bar{b}' \rightarrow \bar{a}'; \bar{c} \rightarrow \bar{b}\} \\ \mathcal{K}_n &= \{a; a'\} & \mathcal{K}_p &= \{\bar{d}\} \end{aligned}$$

²Of course, the validity of this extension hinges on the extent to which the ASPIC_D⁺ version of an ASPIC⁺ theory generates the same conclusions, and we address this question at length below.

The arguments are:

$$\begin{aligned} A_1 &= [a] & A_2 &= [A_1 \Rightarrow b] & A_3 &= [A_2 \rightarrow c] \\ B_1 &= [a'] & B_2 &= [B_1 \rightarrow b'] & B_3 &= [B_2 \Rightarrow \bar{c}] & B_4 &= [B_3 \rightarrow \bar{b}] \\ C &= [\bar{d}]; & C' &= [B_4 \Rightarrow d] \end{aligned}$$

In the original theory, \bar{d} was always a justified conclusion. However, in the converted theory, d is a justified conclusion only if $C \prec C'$.

The point that this example illustrates is that making an arbitrary ASPIC⁺ theory rational — in the sense of making sure that it conforms to the rationality postulates — may require the addition of some strict rules. These rules, in turn, may lead to conclusions that might be considered unintuitive in the sense that they overturn conclusions of the unmodified theory. In the example above, the addition of the strict rules means that under some circumstances \bar{d} ceases to be a conclusion. We argue that even if ASPIC⁺ theories are automatically completed to make them rational, this completion will entail additional knowledge engineering to establish many conclusions and to check that they make sense. Such work is not required when ASPIC_D⁺ is used as the representation language.

4.3 Unrestricted rebut

Another issue to consider is the effect of moving to a system with only unrestricted rebut. The original ASPIC framework [3] used unrestricted rebut

— an argument could be rebutted by any argument no matter what the `TopRule` was. However, [21] pointed out that unrestricted rebut may lead to non-intuitive results — exactly those characterized by the rationality postulates — and suggested using restricted rebut in order to make it possible to apply complete-based semantics while maintaining rationality. Our results show that in ASPIC_D^+ one can use the complete-based semantics with unrestricted rebut, but before we examine how this is the case, let's look at an example that illustrates the problems with restricted rebut — this example is taken from from [2].

Example 4.3 *Consider the theory $AT_{4.3}$ which makes use of the same language \mathcal{L} as before:*

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}\} & \mathcal{R}_s &= \{a' \rightarrow b'; b \rightarrow c\} \\ \mathcal{K}_n &= \{a; a'\} & \mathcal{K}_p &= \emptyset \end{aligned}$$

Note that this example is of an ASPIC^+ theory that is ill-defined since it is not closed under transposition. The arguments are:

$$\begin{aligned} A_1 &= [a] & A_2 &= [A_1 \Rightarrow b] & A_3 &= [A_2 \rightarrow c] \\ B_1 &= [a'] & B_2 &= [B_1 \rightarrow b'] & B_3 &= [B_2 \Rightarrow \bar{c}] \end{aligned}$$

Because of the use of restricted rebut, A_3 rebuts B_3 , but not vice versa, and we get the attack and defeat in Figures 4.1(a) and 4.1(b) respectively.

One can argue, as in [2], that the behavior that one sees in this example is unintuitive since with one strict and one defeasible rule each, both arguments might be considered to be in some sense equally strong, in which case preferring one over the other at the purely syntactic level is questionable³. The next example shows how this problem goes away when using unrestricted rebut in ASPIC_D^+ :

Example 4.4 *Now, consider the ASPIC_D^+ version of $AT_{4.3}$ from the previous example which we will call $AT_{4.4}$. This is:*

$$\begin{array}{ll} \mathcal{R}_d = \{a \Rightarrow b; b' \Rightarrow \bar{c}\} & \mathcal{R}'_d = \{a' \Rightarrow b'; b \Rightarrow c\} \\ \mathcal{K}'_p = \{a; a'\} & \mathcal{K}_p = \emptyset \end{array}$$

The arguments are the defeasible versions of the arguments of $AT_{4.3}$ (note that we give them the same names as their strict counterparts to simplify the following discussion):

$$\begin{array}{lll} A_1 = [a] & A_2 = [A_1 \Rightarrow b] & A_3 = [A_2 \Rightarrow c] \\ B_1 = [a'] & B_2 = [B_1 \Rightarrow b'] & B_3 = [B_2 \Rightarrow \bar{c}] \end{array}$$

The attacks and defeats between these arguments depend on the preference

³The idea that the two arguments are equally strong assumes that the strength of an argument is affected by the strength of all of its components — for example by summing/averaging the strength of the rules, or using the weakest link approach — while preferring the argument with the strict last rule is closer to a last link approach in having a single rule determine the strength of the argument. Of course for strength to be a determining factor, we have to be discussing which of two attacking arguments defeat one another, and ASPIC^+ does not even recognize an attack between A_3 and B_3 .

order over the arguments. Using a strict-first preference ordering and the weakest link principle, the defeat relation will depend on the relative preference of the two rules $a \Rightarrow b$ and $b' \Rightarrow \bar{c}$. There are three possible scenarios: (i) $(a \Rightarrow b) = (b' \Rightarrow \bar{c})$ ⁴, (ii) $(a \Rightarrow b) < (b' \Rightarrow \bar{c})$, and (iii) $(b' \Rightarrow \bar{c}) < (a \Rightarrow b)$. The results in all three cases are shown in Figure 4.1, attack relation in Figure 4.1(c), then the defeats in cases (i), (ii) and (iii) in Figures 4.1(d), 4.1(e) and 4.1(f) respectively.

The point we want to make here is not that ASPIC⁺ is wrong in the way it handles the example, but that the requirement to use restricted rebut (which is required to meet the rationality postulates) forces a particular outcome to the conflict between A_3 and B_3 , and therefore restricts what can be represented. ASPIC_D⁺, since it adopts unrestricted rebut, is more flexible and allows for a wider range of scenarios to be represented. Adjusting the preference order over the rules makes it possible for either of A_3 and B_3 to not defeat the other. Thus we can model situations in which c , \bar{c} or neither is a justified conclusion.

Next consider another example, from [23], of the consequences of using restricted rebut.

Example 4.5 *John*: “Bob will attend conferences A and I this year, as he

⁴If r and r' are two rules, $r = r'$ is defined as $r \leq r'$ and $r' \leq r$

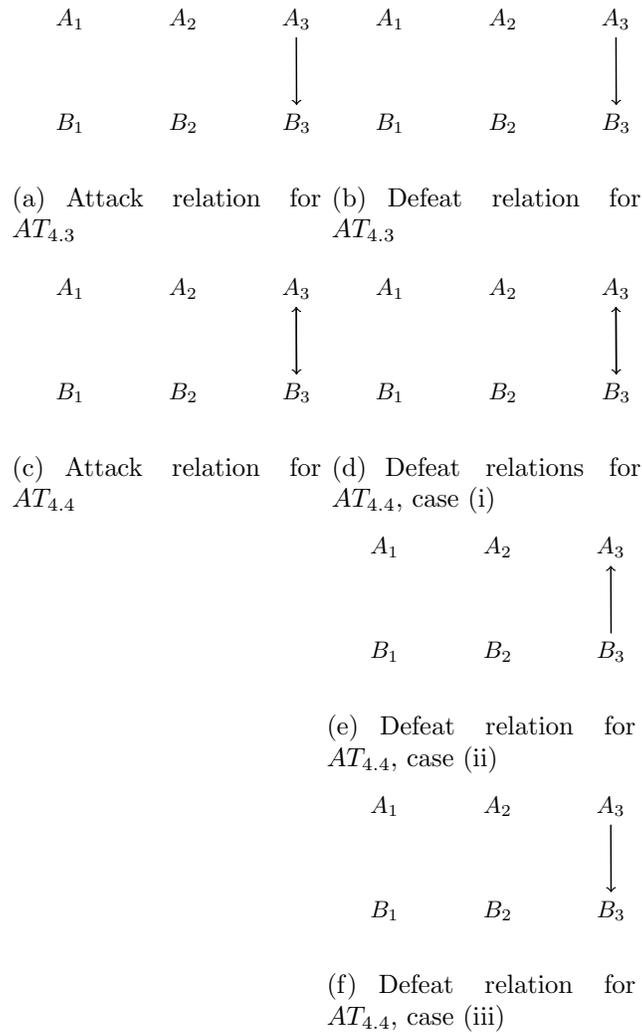


Figure 4.1: Attack and defeat relations for $AT_{4.3}$ and $AT_{4.4}$ from Examples 4.3 and 4.4.

has papers accepted at both.” Mary: “That won’t be possible, as his budget of \$1000 only allows for one foreign trip.” Formally, this discussion could be

modeled using the argumentation theory $AT_{4.4}$:

$$\mathcal{R}_d = \{accA \Rightarrow attA; accI \Rightarrow attI; budget \Rightarrow \overline{both}\} \quad \mathcal{R}_s = \{attA, attI \rightarrow both\}$$

$$\mathcal{K}_n = \{accA; accI; budget\} \quad \mathcal{K}_p = \emptyset$$

The statements translate into

$$John : [accA; accA \Rightarrow attA; accI; accI \Rightarrow attI; attA, attI \rightarrow both]$$

$$Mary : [budget; budget \Rightarrow \overline{both}]$$

In $ASPIC^+$ or any other formalism based on restricted rebut, Mary's argument does not attack John's argument, since the conclusion Mary wants to attack, *both*, is the consequent of a strict rule. As [23], points out, in $ASPIC^+$, if Mary wants to attack John's argument, she can only do so by attacking the consequent of a defeasible rule. That is, she would be forced to choose to attack either *attA* or *attI*, meaning she essentially has to utter one of the following two statements:

1. Mary': "Bob will attend *I*, so can't attend *A*; his budget doesn't allow him to attend both."
2. Mary'': "Bob will attend *A*, so can't attend *I*; his budget doesn't allow him to attend both."

The respective formal counter arguments are as follows:

$$\text{Mary}' : [\text{budget}; \text{budget} \Rightarrow \overline{\text{both}}; \text{accI}; \text{accI} \Rightarrow \text{attI}; \overline{\text{both}}, \text{attI} \rightarrow \overline{\text{attA}}]$$

$$\text{Mary}'' : [\text{accA}; \text{accA} \Rightarrow \text{attA}; \text{budget}; \text{budget} \Rightarrow \overline{\text{both}}; \overline{\text{both}}, \text{attA} \rightarrow \overline{\text{attI}}]$$

Critically [23], Mary does not know which of the two conferences Bob will attend, yet using restricted rebut forces her to make concrete statements about this. From the perspective of commitment in dialogue [83], this is unnatural — one should not be forced to commit to things one has insufficient reasons to believe in. This problem would go away if rebut was allowed to be unrestricted since Mary's argument:

$$\text{budget}; \text{budget} \Rightarrow \overline{\text{both}}$$

would attack John's argument.

As the provenance of the previous examples shows, the issue of restricted versus unrestricted rebut has been studied for a while. [24] raised three open questions related to the use of unrestricted rebut:

1. Are there any non-admissibility based semantics whose entailment satisfies the rationality postulates?
2. Are there any abstract argumentation semantics, apart from grounded,

that satisfy the rationality postulates when applying unrestricted rebutting?

3. In which way can the approach of value-based argumentation be repaired in order to yield consistent conclusions?

At the time [24] was published, none of these questions could be answered. The work that has subsequently come closest to answering them is [23]. [23] presented a system called ASPIC- which allows unrestricted rebut and can satisfy all three rationality postulates. However, ASPIC- only satisfies the rationality postulates when using the grounded semantics — using other semantics causes the rationality conditions to be violated.

In contrast, ASPIC_D^+ has no problems with satisfying the rationality postulates with unrestricted rebut. Since rebutting is unrestricted natively in ASPIC_D^+ , Propositions 4.2–4.4 give us the following:

Proposition 4.5 (Unrestricted Rebut) *Any ASPIC_D^+ theory satisfies the rationality postulates for all complete-based semantics when using unrestricted rebut.*

Proof *Since, as discussed above, all rebutting in ASPIC_D^+ is unrestricted, Propositions 4.2–4.4 demonstrate that the rationality postulates hold for all complete-based semantics when using unrestricted rebut. \square*

Thus ASPIC_D^+ goes further than ASPIC^- in extending the scope of reasoning possible with unrestricted rebut.

Of course, the results in Propositions 4.2–4.5 are achieved by giving up strict rules, and it is natural to ask what the consequence is for what can be represented in an ASPIC_D^+ theory. We have already seen in Example 4.4 that the use of unrestricted rebut in ASPIC_D^+ gives some additional flexibility in terms of what can be represented when compared with ASPIC^+ . However, it is equally important to ask whether using ASPIC_D^+ , and hence being limited to defeasible rules, means any restriction on what can be represented? This is the subject of the next section.

4.4 The expressiveness of ASPIC_D^+

The results in this section begin to investigate the relationship between ASPIC_D^+ and ASPIC^+ with regard to what conclusions can be drawn from theories expressed in both formalisms. In particular, we identify the conditions under which we can take an arbitrary ASPIC^+ theory and construct an ASPIC_D^+ theory with the same justified conclusions. This section begins the work by limiting consideration to ASPIC^+ theories that are well-defined and hence rational.

We start, however, with results that hold for any ASPIC^+ theory, ob-

serving that:

Proposition 4.6 *For a given language \mathcal{L} , there is a defeasible version AT_D of any ASPIC⁺ argumentation theory AT .*

Proof *Consider the clauses of Definition 4.5 as a series of rewrite rules. Any AT can be converted into its defeasible version by turning every axiom into an ordinary premise and every strict rule into a defeasible rule. \square*

This means that whatever information we have in an ASPIC⁺ theory, we can capture it in an ASPIC_D⁺ theory — we don't lose the ability to represent information about the world by using ASPIC_D⁺ rather than ASPIC⁺. However, it is not just representing information that is important. The set of arguments that can be constructed from a theory, and, in particular, the justified conclusions of a theory are also important. We have:

Proposition 4.7 *Given an ASPIC⁺ theory AT and its defeasible version AT_D , $|\mathcal{A}(AT)| = |\mathcal{A}(AT_D)|$ and for every $A \in AT$ there is exactly one $A_D \in \mathcal{A}(AT_D)$ such that A_D is the defeasible version of A .*

Proof *We show there is a 1-to-1 map between $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$. For each argument that is just a premise or an axiom $A = [\phi]$, we have $A_D = [\phi]$ that is just a premise; for each argument $A = [A_1, \dots, A_n \Rightarrow \phi]$, we have*

$A_D = [A_{1_D}, \dots, A_{n_D} \Rightarrow \phi]$; for each argument $A = [A_1, \dots, A_n \rightarrow \phi]$, we have $A_D = [A_{1_D}, \dots, A_{n_D} \Rightarrow \phi]$. \square

Thus any ASPIC⁺ theory can be turned into an ASPIC_D⁺ theory, and we can generate the same number of arguments, but arguments that had strict components will now only have defeasible components. Furthermore, there are preference orderings such that the same preferences exist between ASPIC_D⁺ arguments as between the corresponding ASPIC⁺ arguments:

Proposition 4.8 *Consider an ASPIC⁺ theory AT and its defeasible version AT_D where the preference ordering over AT_D is the strict-first version of the ordering over AT . Using the elitist weakest link principle, for any $A, B \in \mathcal{A}(AT)$, and $A_D, B_D \in \mathcal{A}(AT_D)$, where A_D, B_D are the defeasible versions of A and B , $A_D \preceq_D B_D$ iff $A \preceq B$.*

Proof Let $AT = \langle AS, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ and $AT_D = \langle AS_D, \mathcal{K}_{p'} \cup \mathcal{K}_p \rangle$. Consider the preference order \leq over rules in AT , and the preference order \leq' over premises. Let $\langle \leq_D, \leq'_D \rangle$ contain all the relations in $\langle \leq, \leq' \rangle$. Since AF_D has more defeasible elements than AF , we need to determine where these elements fit in the ordering. With a strict-first ordering, the translated strict rules/axioms have the highest preference ordering, and so the weakest links in $\mathcal{A}(AT_D)$ are not the translated strict rules/axioms. Furthermore, all the

remaining rules/premises in AT_D have the same preference ordering as in AT . Therefore, under the elitist weakest link principle, AT_D and AT have the same preference ordering over arguments. \square

Note that under the democratic weakest link principle, the defeasible versions of arguments that contain strict rules are all equally preferred and are strictly preferred to any argument that does not contain strict rules.

What the above result tells us is that using the elitist weakest link principle, we can take a set of ASPIC⁺ arguments, create the defeasible versions of those arguments, and still have the same preference ordering as over the original set of arguments. This allows us to work towards our first important result: that if we adopt the grounded semantics we can construct a defeasible version of a given well-defined ASPIC⁺ framework such that the justified conclusions of both theories are the same. Note the focus on well-defined ASPIC⁺ theories — for the rest of this section, well-defined ASPIC⁺ theories will be our starting point:

Lemma 4.1 *Consider a well-defined ASPIC⁺ theory AT and its defeasible version AT_D . If the preference order over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ is the same, then any extension of AT is an extension of AT_D under the same semantics.*

Proof *From Proposition 4.7 we know that there is a one to one mapping*

between $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ such that for every argument in $\mathcal{A}(AT)$ its defeasible version is in $\mathcal{A}(AT)$. In addition, we are told that the preference order \preceq over $\mathcal{A}(AT_D)$ is the same as the preference order \preceq_D over $\mathcal{A}(AT_D)$. Now, consider the attack relations att and att_D over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$. If $(A, A') \in att$, then $(A_D, A'_D) \in att_D$ and there is an attack between the defeasible versions of the arguments A_D and A'_D . However, att_D can contains more attacks. (A_D, A'_D) can be in att_D when $(A, A') \notin att$ iff (1) A' is (just) an axiom in AT or (2) A' has a strict **TopRule** and the attack is not permitted by restricted rebut. We now show, in turn, that these additional attacks do not affect the extensions.

First, if A' is an axiom, then A'_D , which as alone premise that is the defeasible version of an axiom, has the highest possible preference. Thus it can only be defeated by an A_D that has the highest level of preference. Such an argument is the defeasible version of a strict and firm argument. However, if A was strict and firm, AT would be ill-defined (it would have two strict elements in conflict). Therefore, we have the same defeat relations over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ and hence the same extensions for AT and AT_D .

Second, if $\text{TopRule}(A')$ is strict, there are two sub-cases that concern us. (a) If $A_D \prec A'_D$, the attack does not become a defeat. Thus AF_D and AF have the same defeat relation, therefore they have the same extensions.

(b) If $A'_D \preceq A_D$, then there is one more defeat relation over $\mathcal{A}(AT_D)$ than over $\mathcal{A}(AT)$. We will show that this additional defeat relation has no effect. Consider applying all the defeat relations except this additional one — there are three possibilities for the labelling of A'_D (that will be mirrored by the labelling of A' which does not have to contend with this additional defeat) and for each of these, we have to consider all three possibilities for the status of A_D .

1. A'_D is labeled **IN**. If A_D is labeled **IN**, then AT has two **IN** arguments, A and A' , and the conclusions of these arguments are in the set of justified conclusions. However, since A_D and A'_D rebut one another, the conclusions of A and A' are contradictory, violating direct consistency. Thus A and A' cannot both be **IN**, and so neither can A'_D and A_D before the application of the new defeat. If A_D is labeled **OUT** then adding the defeat relation (A_D, A'_D) has no effect. If A_D is labeled **IN**, the situation is more complicated. We start by noting that A will also be **UNDEC**, and then consider how this can be the case. A has a strict top rule, so $A' = [A'_1, \dots, A'_n \rightarrow a]$ where the top rule is $p_1, \dots, p_n \rightarrow a$. Similarly, $A = [A_1, \dots, A_n \Rightarrow \bar{a}]$ with a top rule $q_1, \dots, q_n \Rightarrow \bar{a}$. By closure under transposition, there exists a strict rule $p_1, \dots, p_{i-1}, \bar{a}, p_{i+1}, \dots, p_n \rightarrow \bar{p}_i$ in AT . Since $A'_D \preceq A_D$, it is not possible for A' to be strict, so A' has

at least one defeasible sub-argument, and hence a sub-argument with a defeasible top rule. Lets assume that this is one of the A'_1, \dots, A'_n that combine with the strict top rule, and call it A'_i . Using the strict rule from the transposition of the top rule, we get an argument $B = [A'_1, \dots, A'_{i-1}, A, A'_{i+1}, \dots, A'_n \rightarrow \bar{p}_i]$ which rebuts A' . B is A plus the A'_j , $j \neq i$, and the transposed strict rule. If A'_1, \dots, A'_n do not have defeasible top rules, then we chain the corresponding transposed strict top rule(s) to B to build an argument that attacks A' further down the argument tree until we get an argument, call it B' , which rebuts A' on its defeasible sub-argument. Now, $A' \preceq B'$ since $A' \preceq A$ and B' is A plus some sub-arguments of A' and a sequence of strict (transposed) rules. Therefore B' defeats A' . Moreover, any defeater of B' must be a defeater of A or A' . Next we consider the labeling. Since A' is labeled IN, all the defeaters of A' are labeled OUT. Since A is labeled UNDEC, all the defeaters of A are labeled OUT or UNDEC. Therefore, the defeaters of B , which are the defeaters of A or A' , are labelled OUT or UNDEC. Thus B is labeled IN or UNDEC. Since B defeats A' , A' can not be labeled IN, contradicting what we started with.

2. A'_D is labeled OUT. Adding one more defeat relation (A_D, A'_D) has no

effect.

3. A'_D is labeled UNDEC. If A_D is labeled OUT or UNDEC, then adding the defeat relation (A_D, A'_D) has no effect. However, if A_D is labeled IN, then applying the last defeat relation means that A'_D will now be labeled OUT while A' , which does not have to contend with (A, A') , will be UNDEC.

So A'_D cannot be initially labeled IN. If it is labelled OUT, the status of A'_D cannot change as a result of the additional defeat. If A'_D is initially labeled UNDEC, the status of A'_D can change. However, by showing that A'_D does not defeat any other arguments we can show that this change does not affect the justified conclusions. Consider an argument $B \in \mathcal{A}(AT_D)$ that is attacked by A'_D . A'_D cannot undercut B since the conclusion of A'_D is not a “rule” (if it were a rule, there would be no rebut between A_D and A'_D and there would be no new defeat relation to consider). A'_D can not undermine B since the conclusion of A'_D is not a premise because we know that A' and hence A'_D has a **TopRule**. So we can only be dealing with a rebut, and since we already know that A_D rebuts A_D , B has to be an argument of which A_D is a sub-argument. Since $A'_D \prec A_D$, A'_D does not defeat B .

Thus, in all of these three sub-cases of (b), the additional defeat (A_D, A'_D) has no effect on the status of the arguments in $\mathcal{A}(AT_D)$, again there is no

difference between the extensions of AT and AT_D , and the result holds. \square

What this tells us is that if we translate an ASPIC theory AT into its defeasible version AT_D and then compute extensions, we will get all of the extensions of the original theory. However, Lemma 4.1 leaves open the possibility that, in general, we will get some additional extensions beyond those that we would obtain from AT . If we restrict ourselves to the grounded extension, the extension of AT_D (since it is the grounded semantics there is exactly one) is also the extension of AT :

Proposition 4.9 *Consider a well-defined ASPIC⁺ theory AT and its defeasible version AT_D . If the preference order over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ is the same, then the grounded extension of AT_D is the grounded extension of AT .*

Proof Lemma 4.1 tells us that any extension of AT is an extension of AT_D . Since AT_D can have at most one grounded extension, this must be the same as the grounded extension of AT . \square

We can also show this result directly. Moreover, the proof explains why the extensions only coincide under the grounded semantics:

Proposition 4.10 *Consider a well-defined ASPIC⁺ theory AT and its defeasible version AT_D . If the preference order over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ is the same, then the grounded extension of AT_D is the grounded extension of AT .*

Proof This is similar to the proof of Lemma 4.1. Consider the attack relations att and att_D over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$. att_D may contain more relations than att , and these relations have the following properties:

- $(A, A') \in att$ and $(A_D, A'_D) \in att_D$
- $\text{TopRule}(A)$ is a strict rule
- $(A'_D, A_D) \in att$, but $(A', A) \notin att$

As the proof of Lemma 4.1 shows, if $A'_D < A_D$, then the additional attack relation (A'_D, A_D) is not a defeat relation. Therefore, any extension of AT_D is an extension of AT . If $A'_D \geq A_D$, AT_D has more defeat relations than AT , namely, (A'_D, A_D) , and we have to show that taking out this defeat has no effect under the grounded semantics. Let $A_D = [A_1, \dots, A_n \Rightarrow \phi]$, $\text{Conc}(A'_D) = \bar{\phi}$. Note that $\text{TopRule}(A)$ is a strict rule, we can assume $B_i = [A_1, \dots, A_{i-1}, A'_D, A_{i+1}, \dots, A_n \Rightarrow \overline{\text{Conc}(A_i)}]$. So every two arguments in $\{A_D, B_1, \dots, B_m\}$ attack each other, and $B_i \preceq A_D$ (since A_D is a sub-argument of B_i). There are three possibilities for the status of A_D , we are considering the three possibilities one by one:

1. A_D is labeled IN. Then taking out the defeat (A'_D, A_D) has no effect.
2. A_D is labeled UNDEC. If A'_D is labeled IN or OUT, then taking out the

defeat (A'_D, A_D) has no effect. If A'_D is labeled UNDEC, and at least one of the other defeaters of A_D is labeled UNDEC, then taking out the defeat (A'_D, A_D) has no effect. If A'_D is labeled UNDEC, and all the other defeaters of A_D are labeled OUT, then it is impossible for A_D to be labeled UNDEC, and B_i to be labeled OUT (since $B_i \preceq A_D$ and A_D defeats B_i).

3. A_D is labeled OUT. It is impossible for A'_D labeled UNDEC. If A'_D is labeled OUT, taking out the defeat (A'_D, A_D) has no effect. If A'_D is labeled IN, then it may be possible that B_i are all labeled OUT under preferred semantics. And in this case, all A_D and A'_D are equally preferred, so under the grounded semantics, they are all labeled UNDEC.

Overall, any extension of AT_D is the extension of AT under the grounded semantics. □

Let's take a look at an example to see why AT_D may contain more extensions than AT under, for example, the preferred semantics. The following example is based on Caminada's tandem example [13], but is simplified to the case of two, rather than three, conflicting arguments. (So, rather than being about the impossibility of three people riding on a tandem, it could be about two people riding on a bicycle with one seat.) The advantage of the simpler example here is that it makes it clear than any conflict can lead to this kind

of problem when translating into from ASPIC^+ into ASPIC_D^+ , it does not have to be a three-way interaction.

Example 4.6 Consider an ASPIC^+ theory $AT_{4.6}$:

$$\begin{array}{ll} \mathcal{K}_n = \{wa, wb\} & \mathcal{K}_p = \emptyset \\ \mathcal{R}_s = \{a \rightarrow \bar{b}; b \rightarrow \bar{a}\} & \mathcal{R}_d = \{wa \Rightarrow a; wa \Rightarrow b\} \end{array}$$

where all the defeasible rules have the same preference ordering. Now we can construct arguments:

$$\begin{array}{lll} A_1 = [wa]; & A_2 = [A_1 \Rightarrow a]; & A_3 = [A_2 \rightarrow \bar{b}]; \\ B_1 = [wb]; & B_2 = [B_1 \Rightarrow b]; & B_3 = [B_2 \rightarrow \bar{a}]; \end{array}$$

And all the attacks are defeats. The defeat relations are shown in Figure 4.2(a). Now, let's translate $AT_{4.6}$ to the ASPIC_D^+ theory $AT_{4.6'}$:

$$\begin{array}{ll} \mathcal{K}'_p = \{wa, wb\} & \mathcal{K}_p = \emptyset \\ \mathcal{R}'_d = \{a \Rightarrow \bar{b}; b \Rightarrow \bar{a}\} & \mathcal{R}_d = \{wa \Rightarrow a; wa \Rightarrow b\} \end{array}$$

Now we can construct arguments:

$$\begin{array}{lll} A_1 = [wa]; & A_2 = [A_1 \Rightarrow a]; & A_3 = [A_2 \Rightarrow \bar{b}]; \\ B_1 = [wb]; & B_2 = [B_1 \Rightarrow b]; & B_3 = [B_2 \Rightarrow \bar{a}]; \end{array}$$

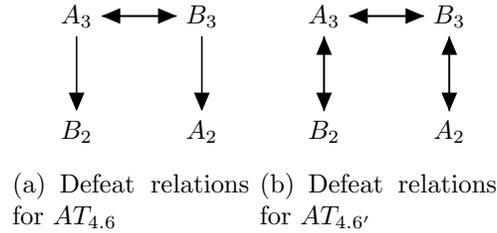


Figure 4.2: Defeat relations for $ASPIC^+$ and $ASPIC_D^+$ versions of the same theory, as in Example 4.6.

With strict-first preference ordering, there are more defeats in $AT_{4.6'}$ than $AT_{4.6}$ — the defeat relations for Both theories are shown in Figure 4.2(b). As a result, there is a preferred extension $\{A_2, B_2\}$ for $AT_{4.6'}$ which is not an extension for $AT_{4.6}$.

Together Propositions 4.8, 4.9 and 4.10 give us:

Proposition 4.11 *Consider a well-defined $ASPIC^+$ theory AT and its defeasible version AT_D where the preference ordering over AT_D is the strict-first version of the ordering over AT . Using the elitist weakest link principle and the grounded semantics, AT and AT_D have exactly the same set of justified conclusions.*

Proof *Proposition 4.8 tells us that the preference orders over the arguments $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ will be the same, and then 4.9 and 4.10 tell us that AT and AT_D will have the same grounded extension. With the same extension,*

the justified conclusions of the two theories will be the same. \square

In other words, there is a way of creating a defeasible version AT_D of any ASPIC⁺ theory AT so that the grounded extension of AT will coincide with the grounded extension of AT_D and the two theories will have exactly the same set of justified conclusions. This situation is illustrated by the following example:

Example 4.7 Consider that we start with the following theory $AT_{4.7}$ which uses the same language as before. This is $AT_{4.3}$ closed under transposition and hence well-defined:

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}\} & \mathcal{R}_s &= \{a' \rightarrow b'; b \rightarrow c; \bar{b}' \rightarrow \bar{a}'; \bar{c} \rightarrow \bar{b}\} \\ \mathcal{K}_n &= \{a; a'\} & \mathcal{K}_p &= \emptyset \end{aligned}$$

Then we can construct the following arguments:

$$\begin{aligned} A_1 &= [a] & A_2 &= [A_1 \Rightarrow b] & A_3 &= [A_2 \rightarrow c] \\ B_1 &= [a'] & B_2 &= [B_1 \rightarrow b'] & B_3 &= [B_2 \Rightarrow \bar{c}] & B_4 &= [B_3 \rightarrow \bar{b}] \end{aligned}$$

The attack relation is shown in Figure 4.3(a). Now we translate this framework to the ASPIC_D⁺ theory $AT_{4.7}'$:

$$\begin{aligned} \mathcal{R}_d &= \{a \Rightarrow b; b' \Rightarrow \bar{c}\} & \mathcal{R}'_d &= \{a' \Rightarrow b'; b \Rightarrow c; \bar{b}' \Rightarrow \bar{a}'; \bar{c} \Rightarrow \bar{b}\} \\ \mathcal{K}'_n &= \{a; a'\} & \mathcal{K}_p &= \emptyset \end{aligned}$$

Then we can construct the following arguments:

$$\begin{array}{l} A_1 = [a] \quad A_2 = [A_1 \Rightarrow b] \quad A_3 = [A_2 \Rightarrow c] \\ B_1 = [a'] \quad B_2 = [B_1 \Rightarrow b'] \quad B_3 = [B_2 \Rightarrow \bar{c}] \quad B_4 = [B_3 \Rightarrow \bar{b}] \end{array}$$

The attack relations are shown in Figure 4.4(a). Now let's consider the different possible preference orderings over rules:

- i) $a \Rightarrow b = b' \Rightarrow \bar{c}$ ⁵. Under the elitist weakest link principle, all the attack relations are defeat relations, see Figures 4.3(a) and 4.3(b) and Figures 4.4(a) and 4.4(b). Here $AT_{4.7'}$ has additional defeat relations, but they are directed at arguments that are already defeated. Under the grounded semantics, the set of arguments in the extension of both theories is $\{A_1, B_1, B_2\}$.
- ii) $a \Rightarrow b < b' \Rightarrow \bar{c}$. Under the elitist weakest link principle, the defeat relations are shown in Figure 4.3(c) and 4.4(c). Again $AT_{4.7'}$ has an additional defeat relation, but again it has no effect on the extensions. Under the grounded semantics, the set of arguments in the extension of both theories is $\{A_1, B_1, B_2, B_3, B_4\}$.
- iii) $a \Rightarrow b > b' \Rightarrow \bar{c}$. Under the elitist weakest link principle, the defeat relations are shown in Figure 4.3(d) and 4.4(d). As before $AT_{4.7'}$

⁵As before, $A = B$ is defined as $A \leq B$ and $B \leq A$

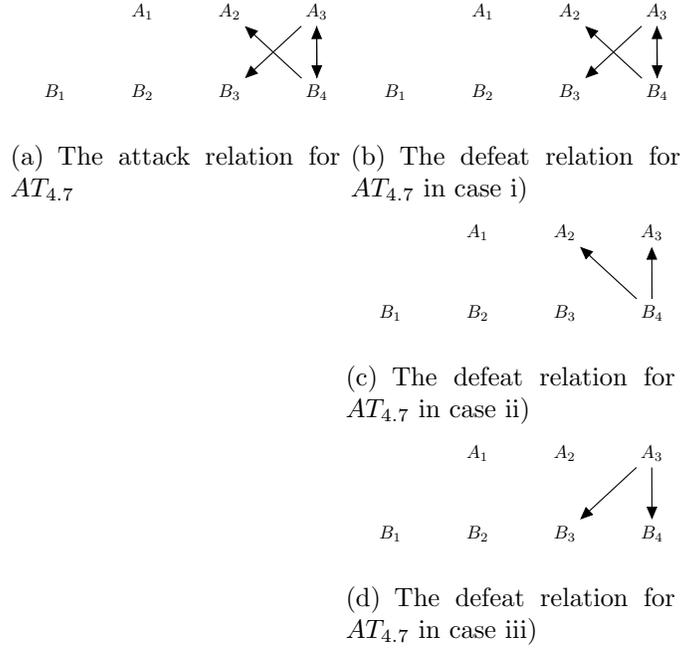


Figure 4.3: Attack and defeat relations for $AT_{4.7}$ in Example 4.7

has additional defeats, but they have no effect. Under the grounded semantics, the set of arguments in the extension of both theories is $\{A_1, A_2, A_3, B_1, B_2\}$.

For all the cases above, the justified conclusions of $AT_{4.7}$ and $AT_{4.7'}$ are exactly the same.

The example shows exactly what Propositions 4.9 and 4.10 prove for the general case. If we take an ASPIC⁺ theory AT , then we can convert it — under the assumption of a strict-first ordering — into an ASPIC_D⁺ theory AT_D which — under the elitist weakest link principle — will have

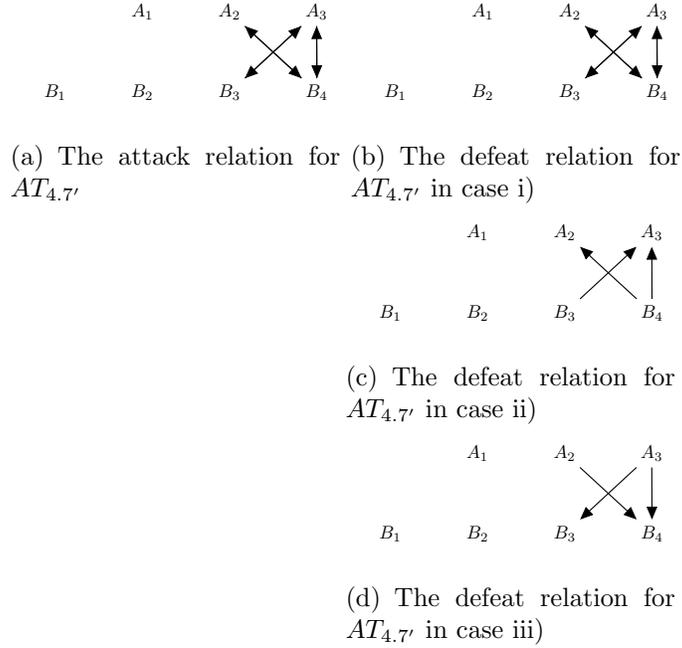


Figure 4.4: Attack and defeat relations for $AT_{4.7'}$ in Example 4.7

exactly the same grounded extension as AT . Hence AT_D will have exactly the same justified conclusions as AT under these conditions. As a result, we can reasonably claim that $ASPIC_D^+$ is capable of capturing exactly the same information as $ASPIC^+$ — we can represent knowledge in such a way that it gives the same conclusions

This equivalence does not hold for the preferred semantics for all $ASPIC^+$ theories (it only holds for a subset of all possible theories, for example those with a single extension). A simple corollary of Proposition 4.10 is that:

Corollary 4.1 Consider a well-defined $ASPIC^+$ theory AT and its defea-

sible version AT_D where the preference ordering over AT_D is the strict-first version of the ordering over AT . Using the elitist weakest link principle and the preferred semantics, AT and AT_D have the exactly same extensions iff there is no argument A such that:

- $\text{TopRule}(A)$ is a strict rule $\phi_1, \dots, \phi_n \rightarrow \phi$; and
- $\exists \mathbf{A} = \{A_1, \dots, A_n\}$, where $A_i \in \mathbf{A}$ and $\text{Conc}(A_i) = \phi_i$; and
- $\forall i, j$ such that $A_i, A_j \in \mathbf{A}$, $A_i \preceq A_j$ and $A_j \preceq A_i$.

This follows because the conditions in the Corollary are exactly those that Proposition 4.10 identifies as the conditions under which AT_D will not have more extensions than AT .

What Corollary 4.1 states is that there is a way of creating a defeasible version AT_D of some ASPIC⁺ theories AT so that the preferred extensions of AT will be exactly the preferred extensions of AT_D , and vice versa. These will be the theories AT without arguments with strict top rules, or, theories with arguments with strict top rules that have subarguments that are not all equally preferred. However, the equivalence between an ASPIC⁺ theory and its defeasible version that we established for grounded semantics does not hold for all ASPIC⁺ theories under the preferred semantics⁶. So, it is

⁶Note that this is not what we argued in [50], though it does hold for some theories.

natural to wonder if we can recover this equivalence by altering some aspect of the translation between an ASPIC⁺ theory and its defeasible counterpart. Given the role of preference orders in the previous results, one way to do this is by changing how we define the preference order over the resulting defeasible theory.

The following definitions are extensions of the weakest link and last link principles:

Definition 4.9 (Translated Weakest-Link Principle) *Consider an ASPIC⁺ theory $AT = \langle \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, n \rangle, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ and its defeasible version ASPIC_D⁺ theory $AT_D = \langle \langle \mathcal{L}, \mathcal{R}_{d'} \cup \mathcal{R}_d, n' \rangle, \mathcal{K}_{p'} \cup \mathcal{K}_p \rangle$. Let A and B be two arguments in $\mathcal{A}(AT_D)$. $A \preceq B$ iff:*

- $\text{TopRule}(A) \notin \mathcal{R}_{d'}$; and
- $\text{Prem}_p(A) \preceq \text{Prem}_p(B)$ and $\text{Rules}_d(A) \preceq \text{Rules}_d(B)$

Definition 4.10 (Translated Last-Link Principle) *Consider an ASPIC⁺ theory $AT = \langle \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, n \rangle, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ and its defeasible version ASPIC_D⁺ theory $AT_D = \langle \langle \mathcal{L}, \mathcal{R}_{d'} \cup \mathcal{R}_d, n' \rangle, \mathcal{K}_{p'} \cup \mathcal{K}_p \rangle$. Let A and B be two arguments in $\mathcal{A}(AT_D)$. $A \preceq B$ iff:*

In that paper we gave the result that taking an ASPIC⁺ theory AT , and converting it, under the assumption of a strict-first ordering, into an ASPIC_D⁺ theory AT_D would mean that AT would have exactly the same justified conclusions as AT_D under all semantics. That result was wrong.

- $\text{TopRule}(A) \notin \mathcal{R}_d$; and
- *Either*
 - $\text{LastDefRules}(A) = \text{LastDefRules}(B) = \emptyset$ and $\text{Prem}(A) \leq \text{Prem}(B)$;
 - or*
 - $\text{LastDefRules}(A) \leq \text{LastDefRules}(B)$.

In contrast to Definitions 2.23 and 2.25, which apply to any pair of arguments, the above definitions are applicable to a theory which is the defeasible version of an ASPIC⁺ theory. What Definitions 4.9 and 4.10 add is a check that the final rule in A did not get translated from a strict rule⁷. This is not a particularly elegant way to ensure that the justified conclusions of AT_D are identical to those of AT , but it is enough:

Proposition 4.12 *Consider a well-defined ASPIC⁺ theory AT and its defeasible version AT_D . Under either the translated weakest-link principle or the translated last-link principle, and any complete-based semantics, AT and AT_D have the exactly same set of extensions.*

⁷The way we have defined the translated weakest and last link principles are one way to do this, there are others, but they all rely on being able to tell, somehow, that a given rule is the translation of a strict rule so that any argument that ends with that rule can be defeated.

Proof From Proposition 4.7 we know that there is a one to one mapping between $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$ such that for every argument in $\mathcal{A}(AT)$ its defeasible version is in $\mathcal{A}(AT)$. Now, consider the attack relations att and att_D over $\mathcal{A}(AT)$ and $\mathcal{A}(AT_D)$. If $(A, A') \in att$, then $(A_D, A'_D) \in att_D$ and there is an attack between the defeasible versions of the arguments A_D and A'_D . However, att_D can contain more attacks. (A_D, A'_D) can be in att_D when $(A, A') \notin att$ iff (1) A' is (just) an axiom in AT or (2) A' has a strict **TopRule** and the attack is not permitted by restricted rebut. We now show, in turn, that these additional attacks will not be defeats.

1. If A' is an axiom in AT , then $A \prec A'$, otherwise AT is ill-defined, since the strict part is not consistent. Therefore, under the strict-first version of the ordering over AT , $A_D \prec A'_D$. Thus the additional attack is not a defeat.
2. If **TopRule**(A') is a strict rule, then the additional attacks are all rebuttals — the result of turning strict rules into defeasible rules — and so are preference dependent attacks. So we have to consider the preferences of the arguments being attacked.
 - if $A \prec A'$, then $A'_D \not\prec A_D$ under either the translated weakest-link principle or the translated last-link principle, since the second

condition of those principles are not met. Thus the additional attacks will not become defeats.

- if $A' \preceq A$, then $A'_D \not\preceq A_D$ under either the translated weakest-link principle or the translated last-link principle, since the first conditions are not met. Thus the additional attacks will not become defeats.

Overall, the additional attacks relations will not be defeats. Therefore, AT and AT_D have the exactly same defeat relations, and hence the set of extensions of AT and AT_D will be the same. \square

Because AT and AT_D have the exactly same set of defeats, they will have the same set of extensions under any semantics if we use the translated weakest link principle or the translated last link principle, and hence the same justified conclusions:

Corollary 4.2 *Consider a well-defined ASPIC⁺ theory AT and its defeasible version AT_D where the preference ordering over AT_D is the strict-first version of the ordering over AT . Under either the translated weakest link principle or the translated last link principle, and any complete-based semantics AT_D and AT have the same justified conclusions.*

4.5 Summary

Now, what we have done in this chapter, is to introduce the purely defeasible system ASPIC_D^+ and investigate the relations between ASPIC^+ and ASPIC_D^+ .

First, Propositions 4.6–4.8 show, in a straightforward way, that any ASPIC^+ theory can be converted an ASPIC_D^+ theory so that there is an equivalent set of arguments, and this can be done in such a way that the same preference ordering exists over the two sets of arguments. This answers the research question RQ1.

Then, I investigate the relations between the conclusions that can be drawn from a well-defined ASPIC^+ theory AT and the ASPIC_D^+ theory AT_D that is its defeasible version. We have

- For the grounded semantics, under the elitist weakest link principle and strict-first preference ordering, AT and AT_D have exactly the same extension and hence the same set of justified conclusions. This is formally stated in Propositions 4.9, 4.10 and 4.11.
- For any complete-based semantics, under the elitist weakest link principle and strict-first preference ordering, any extension of AT is an extension of AT_D . This is shown in Lemma 4.1. However, AT_D may

have more extensions than AT . An example of this situation is given in Example 4.6.

- For any complete-based semantics, under either translated weakest link principle or the translated last link principle, AT and AT_D have the exactly same set of extensions and hence the same set of justified conclusions. This is formally stated in Proposition 4.12 and Corollary 4.2.

Of these three results, I consider the first two the most important. The first says that under the grounded semantics and the (I think very reasonable) notion of a strict-first preference ordering, the justified conclusions of an ASPIC⁺ theory and its defeasible counterpart are the same. The second says that under other semantics, while the defeasible version of a theory will obtain all the extensions of the ASPIC⁺ original, it may also have additional extensions. The, weaker, third result, shows how the extensions of any well-defined ASPIC⁺ theory and its defeasible counterpart can be made to coincide by placing additional conditions on the translation from original to defeasible. This, then, means that the justified conclusions of the theories will be the same. I consider this result to be weaker because, as noted above, the last of these three results requires us to adopt a very specific preference order, one that recalls which rules in AT_D were strict rules in AT , effectively

recapturing the notion of strict rules whilst also making them defeasible⁸. This answers the research question RQ2.

There are two points that should be borne in mind when considering this last result. First, there is no need for such a preference order if we adopt the grounded semantics — under the grounded semantics AT and AT_D will have the same set of justified conclusions with a simple strict-first preference order. Second, if we deal with semantics other than grounded, what we are doing is replacing the special handling of strict rules that $ASPIC^+$ requires when identifying attacks (using restricted rebut) with the special handling of some instances of attack when mapping them to defeat by manipulating the preference order. Both are somewhat awkward. We would argue that neither is more awkward than the other. And, of course, $ASPIC^+$ needs to use restricted rebut even when adopting the grounded semantics. Finally, we argued that one of the advantages that $ASPIC_D^+$ has over $ASPIC^-$ is that $ASPIC_D^+$ obeys the rationality postulates for all semantics while permitting unrestricted rebut whereas $ASPIC^-$ only obeys the rationality postulates with unrestricted rebut for the grounded semantics. This advantage is miti-

⁸Note that we do not show that this restriction is necessary to ensure that the justified conclusions of any well-defined $ASPIC^+$ theory and its defeasible counterpart are the same, merely that this is a sufficient condition. We could not find a way of making the justified conclusions of the two theories to coincide without this condition, but that does not mean for sure that a weaker condition can't be found.

gated somewhat by the fact that ASPIC_D^+ only provides the same justified conclusions as ASPIC^+ for the grounded semantics unless one imposes the special handling of once-strict rules which, in effect, re-restrict unrestricted rebut. This answers the research question RQ3.

I will deal with case where the theory is ill-defined in the next chapter.

Chapter 5

Theories that are ill-defined

The previous chapter investigated the relationship between the conclusions obtained from a theory expressed in ASPIC^+ and that from the same theory expressed in ASPIC_D^+ . The results depend on the initial theory being well-defined. We studied well-defined ASPIC^+ theories because those are the ones that “make sense” in ASPIC^+ , in terms of being guaranteed to conform to the rationality postulates. However, in restricting ourselves to only consider ASPIC_D^+ theories that correspond to well-defined ASPIC^+ theories sells ASPIC_D^+ short. After all, all ASPIC_D^+ theories are “well-defined” in the sense that they satisfy rationality postulates. So, in this chapter, we study ASPIC_D^+ theories that correspond to ASPIC^+ theories that are ill-defined. Since we already know the most important thing about these ASPIC_D^+ theories — that they are rational — we will continue to analyse their performance against the corresponding ASPIC^+ theories, especially in

the case of ASPIC^+ theories that are not closed under transposition, by considering versions of those ASPIC^+ theories have been made well-defined.

Given an ASPIC^+ theory AT that is ill-defined, we will investigate the relationship between the conclusions of a version of AT that has been made well-defined, and the defeasible version of AT obtained by converting it into ASPIC_D^+ (which, as we also showed above, is always rational). Note that in looking at the results of these two approaches to handling a theory that is ill-defined theory, we are asking a lot of ASPIC_D^+ — for the two approaches to have the same conclusions, ASPIC_D^+ must somehow obviate the effect of the transposed strict rules. Looking at the differences, though, helps expose a lot about what it means to make an arbitrary ASPIC^+ theory well-defined. Note also, that there is a simple way to ensure that an ASPIC^+ theory that is ill-defined does not have different conclusions from a corresponding ASPIC_D^+ theory. That is to first make the ASPIC^+ theory well defined (for example by adding transposed strict rules), and then converting it into ASPIC_D^+ . In such a case, as Proposition 4.11 showed, under the grounded semantics the justified conclusions of the two theories will be the same.

Given the definition of “well-defined”, there are two classes of ASPIC^+ theory that are ill-defined, and which we study here; those that are not axiom-consistent, and those which do not have a full set of transposed strict

rules. Before doing that we note, in passing, that there are ASPIC⁺ theories that are ill-defined but are rational, as in the following example.

Example 5.1 Consider an ASPIC⁺ theory $AT_{5.1}$, which is ill-defined:

$$\begin{array}{ll} \mathcal{K}_n = \emptyset & \mathcal{K}_p = \{a, \bar{b}\} \\ \mathcal{R}_s = \{a \rightarrow b\} & \mathcal{R}_d = \emptyset \end{array}$$

where the two premises have the same preference. Now we can construct arguments:

$$A = [a]; \quad B = [A \rightarrow b]; \quad C = [\bar{b}]$$

Because of restricted rebut, the only defeat is (B, C) , which is shown in Figure 5.1(a). Therefore, the justified conclusions are $\{a, b\}$ under any semantics. This demonstrates that the theory is directly consistent, and, since all the strict rules have been applied in arriving at these conclusions, the theory is also indirectly consistent and closed under strict rules. Let's compare these results to those of the well-defined theory $AT'_{5.1}$ created by closing $AT_{5.1}$ under transposition:

$$\begin{array}{ll} \mathcal{K}_n = \emptyset & \mathcal{K}_p = \{a, \bar{b}\} \\ \mathcal{R}_s = \{a \rightarrow b; \bar{b} \rightarrow \bar{a}\} & \mathcal{R}_d = \emptyset \end{array}$$

Now we can construct arguments:

$$A = [a]; \quad B = [A \rightarrow b]; \quad C = [\bar{b}]; \quad D = [C \rightarrow \bar{a}]$$

The defeat relation is shown in Figure 5.1(b). The justified conclusion is \emptyset under the grounded semantics, and $\{\{a, b\}, \{\bar{a}, \bar{b}\}\}$ under the preferred semantics. Finally, let's consider the ASPIC_D^+ version of $AT_{5.1}''$:

$$\begin{array}{ll} \mathcal{K}_n = \emptyset & \mathcal{K}_p = \{a, \bar{b}\} \\ \mathcal{R}_s = \{a \Rightarrow b\} & \mathcal{R}_d = \emptyset \end{array}$$

Now we can construct arguments:

$$A = [a]; \quad B = [A \Rightarrow b]; \quad C = [\bar{b}]$$

The defeat relation is shown in Figure 5.1(c). The justified conclusion is $\{a\}$ under the grounded semantics, and $\{\{a, b\}, \{a, \bar{b}\}\}$ under the preferred semantics.

Despite the fact that theories like the one above can be rational while not being well-defined, we will not consider this to be a separate class of theories for the purposes of our work. Rather, they will be considered along with irrational theories that are ill-defined.

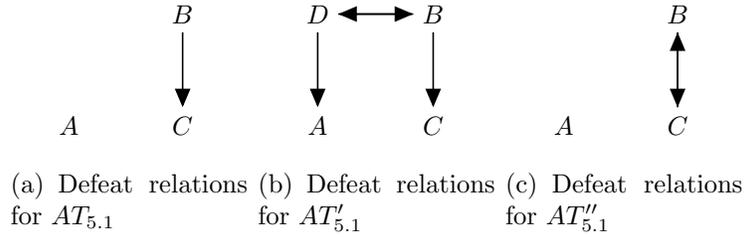


Figure 5.1: Defeat relations for various versions of the same theory, as in Example 5.1.

5.1 Theories that are not closed under transposition

Example 4.7 contains an $ASPIC^+$ theory that is closed under transposition, and so there is a guarantee that there is no inconsistency in the conclusions. This is one situation in which the results in the previous section will hold, and a situation in which an $ASPIC^+$ theory and its $ASPIC_D^+$ counterpart have the same justified conclusions. This section looks deeper into the relationship between $ASPIC^+$ and $ASPIC_D^+$ investigating what can be said about this relationship for $ASPIC^+$ theories that are closed under transposition.

5.1.1 The problem

We start with an example of a theory that is not closed under transposition:

Example 5.2 *Again we consider the example in [60] which gives us the fol-*

lowing ASPIC⁺ theory. Here we will call this $AT_{5.2}$:

$$\mathcal{R}_s = \{t, q \rightarrow \bar{p}\}$$

$$\mathcal{R}_d = \{\bar{s} \Rightarrow t; r \Rightarrow q; a \Rightarrow p\}$$

$$\mathcal{K}_n = \emptyset$$

$$\mathcal{K}_p = \{a; r; \bar{r}; \bar{s}\}$$

then we can construct the arguments:

$$A' = [a]; \quad A = [A' \Rightarrow p];$$

$$B_1 = [\bar{s}]; \quad B'_1 = [B_1 \Rightarrow t];$$

$$B_2 = [r]; \quad B'_2 = [B_2 \Rightarrow q];$$

$$B = [B'_1, B'_2 \rightarrow \bar{p}];$$

$$C = [\bar{r}];$$

The attack relations are: C attacks B_2 , C attacks B'_2 , C attacks B ; B_2 attacks C ; B attacks A . Now, let's assume the preference over defeasible rules and premises as follow:

$$r \Rightarrow q < a \Rightarrow p; \quad \bar{r} <' r$$

$$\bar{a} <' r; \quad \bar{s} <' \bar{r}$$

By applying the weakest link principle, the defeat relations reduce to: B_2

defeats C . The justified conclusions are $\{a; p; \bar{s}; t; r; q; \bar{p}\}$, and these are not consistent.

This problem, of course, is exactly the one that was pointed by [21], and which led to the definition of the rationality postulates. As discussed above, the solution suggested by [21] and implemented in ASPIC⁺ is to close theories under transposition.

Let's consider what happens when we do this to the example above:

Example 5.3 Add two more strict rules to the ASPIC⁺ theory $AT_{5.2}$ of Example 5.2, to get a theory we will call $AT_{5.3}$. The new rules are:

$$p, t \rightarrow \bar{q}$$

$$p, q \rightarrow \bar{t}$$

By applying the two new strict rules, we have two new arguments

$$A_1^+ = [B'_1, A \rightarrow \bar{q}]; \quad A_2^+ = [A, B'_2 \rightarrow \bar{t}]$$

This leads, in turn, to new attack relations: A_1^+ attacks B'_2, B, A_2^+ ; A_2^+ attacks B'_1, B, A_1^+ ; B attacks A_1^+, A_2^+ . If the preference ordering is reasonable, as defined above, then the defeat relations are: A_1^+ defeats B ; A_2^+ defeats B, B'_1, A_1^+ ; B_2 defeats C ; B defeats A_1^+ . The justified conclusions are then $\{a; p; \bar{s}; \bar{t}; r; q\}$ which are now consistent.

These examples show how closing a theory under transposition “fixes” it, making sure that the set of justified conclusions are consistent. However, there is a downside to doing this, and before we continue, we will explore this. In Example 5.2, we have an argument B whose `TopRule` is a strict rule $t, q \rightarrow \bar{p}$, and another argument A is rebutted by B . By introducing transposed strict rules, as in Example 5.3, we have two more arguments A_1^+ and A_2^+ , whose conclusions are \bar{q} and \bar{t} . Moreover, every pair of arguments of A_1^+ , A_2^+ and B attack each other, i.e., the attack relation of A_1^+ , A_2^+ and B forms a complete graph. Therefore, at most one is justified. We can generalize this to the following:

Proposition 5.1 *In any well-defined ASPIC⁺ framework, for any argument A with a strict `TopRule` $a_1, \dots, a_n \rightarrow a$, if there exists an argument B with conclusion \bar{a} , then at most one of $\bar{a}_1, \dots, \bar{a}_n, a$ is justified.*

Proof We denote

$$A = [A_1, \dots, A_n \rightarrow a]$$

Since the strict rule is closed under transposition, and there is a strict rule

$a_1, \dots, a_n \rightarrow a$, therefore, we have the following strict rules:

$$\begin{aligned} \bar{a}, a_2, \dots, a_n &\rightarrow \bar{a}_1 \\ a_1, \bar{a}, a_3, \dots, a_n &\rightarrow \bar{a}_2 \\ &\dots \\ a_1, \dots, a_{n-1}, \bar{a} &\rightarrow \bar{a}_n \end{aligned}$$

Moreover, we can construct the following arguments:

$$\begin{aligned} C_1 &= [B, A_2, \dots, A_n \rightarrow \bar{a}_1] \\ C_2 &= [A_1, B, A_3, \dots, A_n \rightarrow \bar{a}_2] \\ &\dots \\ C_n &= [A_1, \dots, A_{n-1}, B \rightarrow \bar{a}_n] \end{aligned}$$

We note that C_i rebuts A_i (and thus A and $C_j, i \neq j$), A rebuts B (and thus C_i). Thus every pair of arguments in A, C_1, \dots, C_n attack each other. Thus, at most one is in the extension. Therefore, at most one of $\bar{a}_1, \dots, \bar{a}_n, a$ is justified. \square

The reason we consider this to be a downside is that if we consider $a_1, \dots, a_n \rightarrow a$ in propositional logic, it means that $\bar{a}_1 \vee \dots \vee \bar{a}_n \vee a$. Thus, at least one of $\bar{a}_1, \dots, \bar{a}_n, a$ must be true. In contrast, in ASPIC⁺, at most one of $\bar{a}_1, \dots, \bar{a}_n, a$ is justified, and so strict rules in ASPIC⁺ clearly behave rather

differently from the way that propositional logic handles material implication, which is the widely considered to be the natural way to model strict rules in propositional logic.

5.1.2 Under the grounded semantics

Returning to the issue of handling ASPIC⁺ theories that are ill-defined, an alternative method of ensuring that a theory is rational, rather than relying on closure under transposition, is to translate the theory into ASPIC_D⁺:

Example 5.4 *Now consider the theory $AT_{5.4}$, the defeasible version of $AT_{5.2}$:*

$$\mathcal{R}_s = \{\}$$

$$\mathcal{R}_d = \{\bar{s} \Rightarrow t; r \Rightarrow q; a \Rightarrow p; t, q \Rightarrow \bar{p}\}$$

$$\mathcal{K}_n = \emptyset$$

$$\mathcal{K}_p = \{a; r; \bar{r}; \bar{s}\}$$

From this theory, we can construct the arguments:

$$A' = [a]; \quad A = [A' \Rightarrow p];$$

$$B_1 = [\bar{s}]; \quad B'_1 = [B_1 \Rightarrow t];$$

$$B_2 = [r]; \quad B'_2 = [B_2 \Rightarrow q];$$

$$B = [B'_1, B'_2 \Rightarrow \bar{p}];$$

$$C = [\bar{r}];$$

where all the arguments except B are exactly the same as the arguments of $AT_{5.2}$, and B is the defeasible version of argument B from $AT_{5.2}$. As a result, this theory has all the attacks of $AT_{5.2}$ — that is C attacks B_2 , C attacks B'_2 , C attacks B ; B_2 attacks C ; B attacks A — and one additional attack: A attacks B . By applying the weakest link principle, the defeat relations are B_2 defeats C and A defeats B (in $AT_{5.2}$ we only have B_2 defeats C). The justified conclusions are then the consistent set: $\{a; p; \bar{s}; t; r; q\}$.

The two sets of justified conclusions of $AT_{5.3}$ and $AT_{5.4}$ (the two rational versions of $AT_{5.2}$) are not exactly same because $AT_{5.3}$ has more rules than $AT_{5.4}$. ($AT_{5.3}$ has the transposed strict rules and $AT_{5.4}$ does not.) However, if we translate $AT_{5.3}$ into $ASPIC_D^+$ — that is we create the defeasible version of $AT_{5.3}$ — the justified conclusions of $AT_{5.3}$ and its defeasible version will be exactly same under the conditions established in the previous section. The attack and defeat relations are shown in Figure 5.3.

Examples 5.2–5.4 show the relationship between an $ASPIC^+$ theory that is ill-defined and two ways to make it rational (adding the transposition of the strict rules and translating it into $ASPIC_D^+$) for a specific example. The general situation is summarized in Figure 5.2 — given an $ASPIC^+$ theory that is ill-defined, the two mechanisms for ensuring¹ it is rational have sets of

¹We say “ensuring” advisedly — as Example 5.1 shows, a theory that is ill-defined

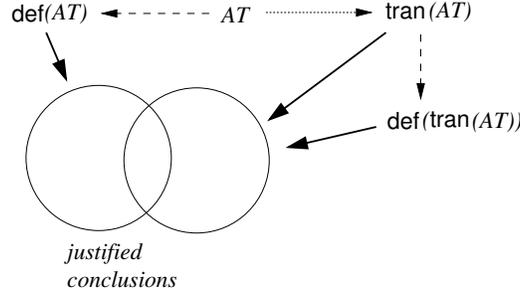


Figure 5.2: The relationship between the conclusions of ASPIC^+ and ASPIC_D^+ theories. AT denotes an ASPIC^+ theory. Heavy arrows denote inference from a given theory. Dashed arrows denote the creation of a defeasible version of a theory. The dotted arrow denotes the creation of the transposed version of a theory.

justified conclusions that overlap, but which are not equivalent. In the rest of this section, we characterize the arguments that will have the same status whichever mechanism is used to rationalize an ASPIC^+ theory. Once we know which arguments have the same status, we can easily establish when the sets of justified conclusions will agree.

To answer this question, we first need the following definition:

Definition 5.1 (Transposed version) ASPIC^+ theory AT_{tr} is the transposed version of ASPIC^+ theory $AT = \langle AS, \mathcal{K}_n \cup \mathcal{K}_p \rangle$ where $AS = \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, n \rangle$ iff:

- $AS_{tr} = \langle \mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d \cup \mathcal{R}_{tr}, n \rangle$, where $\mathcal{R}_{tr} = \bigcup_{\phi_1, \dots, \phi_n \rightarrow \phi \in \mathcal{R}_s} \text{transpose}(\phi_1, \dots, \phi_n \rightarrow$

might be rational.

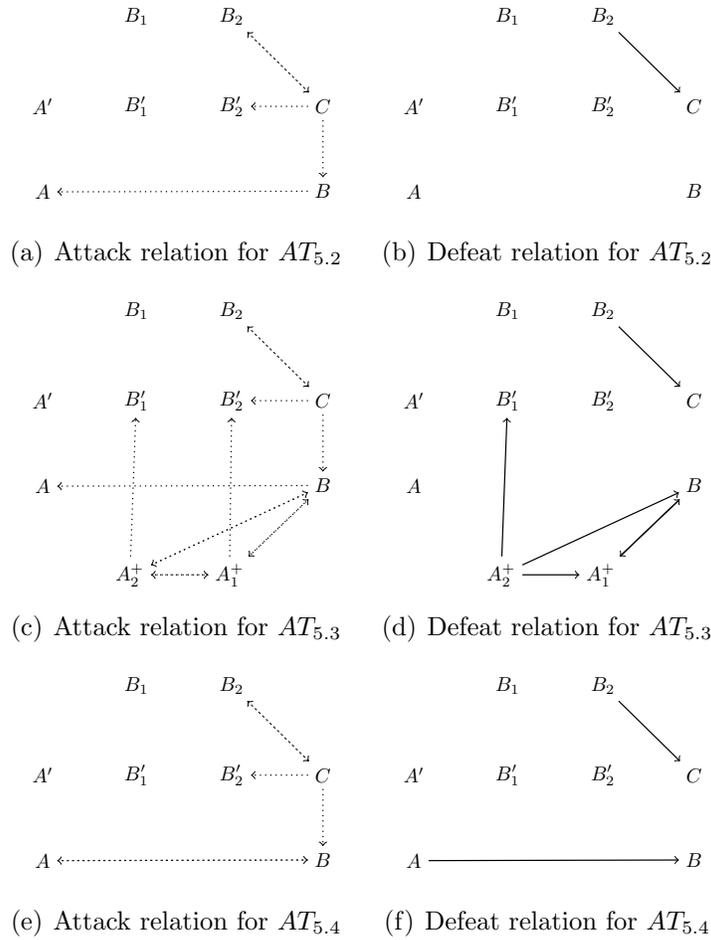


Figure 5.3: Attack and defeat relations for the argumentation theories $AT_{5.2}$ from Example 5.2 (top row), $AT_{5.3}$ from Example 5.3 ($AT_{5.2}$ closed under transposition, middle row) and $AT_{5.4}$ from Example 5.4 (the defeasible version of $AT_{5.2}$, bottom row).

ϕ) where:

$$\begin{aligned} \text{transpose}(\phi_1, \dots, \phi_n \rightarrow \phi) &= \overline{\phi}, \phi_2, \dots, \phi_n \rightarrow \overline{\phi_1} \\ &\cup \phi_1, \overline{\phi}, \phi_3, \dots, \phi_n \rightarrow \overline{\phi_2} \\ &\quad \vdots \\ &\cup \phi_1, \dots, \phi_{n-1}, \overline{\phi} \rightarrow \overline{\phi_n} \end{aligned}$$

- $AT_{tr} = \langle AS_{tr}, \mathcal{K}_n \cup \mathcal{K}_p \rangle$.

If AT_{tr} is the transposed version of AT , we write $AT_{tr} = \text{tr}(AT)$. Note that there are some ASPIC⁺ theories AT that are their own transposed version, so that $AT_{tr} = AT$. Such theories will be those with no strict rules, and those which include all the transposed versions of every strict rule that they contain.

Now, consider an ASPIC⁺ theory AT_+ which is not closed under transposition. The defeasible version of AT_+ is AT_D , and the transposed version of AT_+ is AT_{tr} . The defeasible version of AT_{tr} is AT_{trD} . For any argument $A_+ \in \mathcal{A}(AT_+)$, the corresponding arguments in AT_D , AT_{tr} and AT_{trD} are denoted A_D , A_{tr} and A_{trD} respectively. We then consider, in turn, the different arguments that might be in AT_{tr} and their relationship with the arguments in AT_D .

Considering basic arguments which consist of only premises, we have the following result:

Lemma 5.1 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the elitist weakest link principle, for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ with defeasible version A_D , if $\text{TopRule}(A_{tr}) = \text{undefined}$ and $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then A_D and A_{tr} have the same status.*

Proof *Since there is no $\text{TopRule}(A_{tr})$, the only possible attack on A_{tr} is a rebut. However, there is no rule with conclusion $\overline{\text{Conc}(A_{tr})}$ by definition, and the only difference between AT_D and AT_{tr} is the existence of rules that were added to AT_{tr} to close it under transposition, therefore, A_D and A_{tr} have the same status. \square*

This means that if the conclusion of A_{tr} is a justified conclusion of AT_{tr} , then it is a justified conclusion of A_D . Next, we consider arguments with strict TopRules.

Lemma 5.2 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over*

AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the elitist weakest link principle, for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ with defeasible version A_D , if $\text{TopRule}(A_{tr})$ is a strict rule, and $\text{Sub}(A_D)$ and $\text{Sub}(A_{tr})$ have the same status, then A_D and A_{tr} have the same status.

Proof Before proving the proposition, we need to define some additional notation. We will write:

$$A_+ : A_1^+, \dots, A_n^+ \rightarrow a$$

$$A_D : A_1^D, \dots, A_n^D \Rightarrow a$$

$$A_{tr} : A_1^{tr}, \dots, A_n^{tr} \rightarrow a$$

$$A_{trD} : A_1^{trD}, \dots, A_n^{trD} \Rightarrow a$$

$$\text{TopRule}(A_+) = a_1, \dots, a_n \rightarrow a$$

$$\text{TopRule}(A_D) = a_1, \dots, a_n \Rightarrow a$$

$$\text{TopRule}(A_{tr}) = a_1, \dots, a_n \rightarrow a$$

$$\text{TopRule}(A_{trD}) = a_1, \dots, a_n \Rightarrow a$$

Since AT_{tr} is a well-defined ASPIC⁺ theory, from Corollary 4.1, we know that A_{trD} and A_{tr} have the same status. Therefore, in the rest of proof, we will only consider the status of AT_{trD} instead of AT_{tr} . The only difference

between AT_D and AT_{trD} is that AT_{trD} has more transposed rules like:

$$a_1, \dots, a_{i-1}, \bar{a}, a_{i+1}, \dots, a_n \Rightarrow \bar{a}_i$$

If we have no arguments with the conclusion \bar{a} , then these transposed rules cannot be used to generate attacks on A_{trD} , and so A_D and A_{trD} , and hence AT_{tr} , have the same status. Now, we consider the case where we have an argument B with conclusion \bar{a} . As before, we need to consider four versions of B :

$$B_+ : B_1^+, \dots, B_n^+ \rightarrow \bar{a}$$

$$B_D : B_1^D, \dots, B_n^D \Rightarrow \bar{a}$$

$$B_{tr} : B_1^{tr}, \dots, B_n^{tr} \rightarrow \bar{a}$$

$$B_{trD} : B_1^{trD}, \dots, B_n^{trD} \Rightarrow \bar{a}$$

$$\text{TopRule}(B_+) = b_1, \dots, b_n \rightarrow \bar{a}$$

$$\text{TopRule}(B_D) = b_1, \dots, b_n \Rightarrow \bar{a}$$

$$\text{TopRule}(B_{tr}) = b_1, \dots, b_n \rightarrow \bar{a}$$

$$\text{TopRule}(B_{trD}) = b_1, \dots, b_n \Rightarrow \bar{a}$$

Combining these with the transposed rules indicated above, we have a number of additional arguments C_i :

$$C_i^{trD} : A_1^{trD}, \dots, A_{i-1}^{trD}, B_{trD}, A_{i+1}^{trD}, \dots, A_n^{trD} \Rightarrow \bar{a}_i$$

Furthermore, the attack relations of AT_{trD} and AT_{tr} are shown in Figure 5.4(a) and 5.4(b). Consider the 3 preference orderings in AT_{trD} and AT_{tr} :

1. $B_{trD} \prec A_{trD}$, from Proposition 4.1, we know that $C_i^{trD} \prec A_{trD}$ as well.

The defeat relations for AT_{trD} and AT_D are shown in Figure 5.4(c) and 5.4(d). The additional defeat relation (A_{trD}, C_i^{tr}) has no effect.

2. $A_{trD} \prec B_{trD}$, from Proposition 4.1, we know that $A_{trD} \prec C_i^{trD}$ as well in AT_{trD} . The defeat relations for AT_{trD} and AT_D are shown in Figure 5.4(e) and 5.4(f). If B_{trD} is labeled IN, the additional defeat (C_i^{trD}, A_{trD}) has no effect. If B_{trD} is labeled OUT or UNDEC, then C_i^{trD} must be labeled OUT or UNDEC (B_{trD} is a sub-argument of C_i^{trD} , so all the arguments which defeat B_{trD} defeat C_i^{trD}). Thus the additional defeat (C_i^{trD}, A_{trD}) has no effect.

3. $A_{trD} \sim B_{trD}$ ². The defeat relations for AT_{trD} and AT_D are the same as the attack relations, shown in Figure 5.4(a) and 5.4(b). If B_{trD} is labeled IN in AT_{trD} , then A_{trD} must be labeled OUT in AT_{trD} . If B_{trD} is labeled UNDEC in AT_{trD} , then A_{trD} cannot be labeled IN in AT_{trD} . Therefore, neither A_{trD} nor B_{trD} is in the extension. If B_{trD} is labeled OUT in AT_{trD} , then C_i^{trD} must be labeled OUT (for the same reason as

² $A_{trD} \preceq B_{trD}$ and $B_{trD} \preceq A_{trD}$

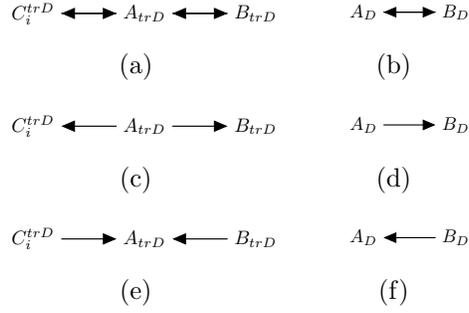


Figure 5.4: Attack and defeat relations for AT_{trD} and AT_D from the proof of Proposition 5.2. (a) Attack relations for AT_{trD} and Defeat relations for AT_{trD} in case 3, (b) Attack relations for AT_D and Defeat relations for AT_D in case 3, (c) Defeat relations for AT_{trD} in case 1, (d) Defeat relations for AT_D in case 1, (e) Defeat relations for AT_{trD} case 2, (f) Defeat relations for AT_D case 2

the previous case). Therefore, the additional defeats (C_i^{trD}, A_{trD}) and (A_{trD}, C_i^{trD}) have no effect.

Therefore, in all cases, A_{trD} and A_D have the same status provided their sub-arguments have the same status, and the same holds for A_{tr} and A_D . \square

Finally — since arguments are either premises, or end with a strict rule or end with a defeasible rule — we consider arguments with defeasible **TopRule**.

This case is similar to Lemma 5.1.

Lemma 5.3 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed*

version AT_{tr} . Under the elitist weakest link principle, for any argument A_{tr} , if $\text{TopRule}(A_{tr})$ is a defeasible rule, $\text{Sub}(A_D)$ and $\text{Sub}(A_{tr})$ have the same status, and $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then A_D and A_{tr} have the same status.

Proof Since all the arguments in $\text{Sub}(A_D)$ have the same status as the corresponding arguments in $\text{Sub}(A_{tr})$, there are no additional defeats on the sub-argument of A_{tr} . Therefore, the only possible additional defeat is due to a rebut on the conclusion. However, there are no rules with conclusion $\overline{\text{Conc}(A_{tr})}$, so A_D and A_{tr} have the same status. \square

These results allow us to identify arguments from AT_{tr} and AT_D that have the same status. Next, we will work through some examples where arguments from these theories do not have the same status. First, let's consider an example in which an argument that only contains a single premise may not have the same status in both the defeasible and transposed versions of the underlying theory.

Example 5.5 Consider three theories, $AT_+^{5.5}$ with premises K_p and rules R , its defeasible version $AT_D^{5.5}$ with premises K_{p_D} and rules R_D , and its transposed version $AT_{tr}^{5.5}$ with premises $K_{p_{tr}}$ and rules R_{tr} :

$$\begin{array}{lll} K_p = \{a, c\} & K_{p_D} = \{a, c\} & K_{p_{tr}} = \{a, c\} \\ R = \{\bar{b} \rightarrow \bar{c}; b \Rightarrow \bar{a}\} & R_D = \{\bar{b} \Rightarrow \bar{c}; b \Rightarrow \bar{a}\} & R_{tr} = \{\bar{b} \rightarrow \bar{c}; b \Rightarrow \bar{a}; c \rightarrow b\} \end{array}$$

We have the following arguments:

$$\begin{array}{lll} A = [a]; & A_D = [a]; & A_{tr} = [a]; \\ B = [c]; & B_D = [c]; & B_{tr} = [c]; \\ & & C_{tr} = [B_{tr} \rightarrow b] \\ & & D_{tr} = [C_{tr} \Rightarrow \bar{a}] \end{array}$$

In this example, B_D and B_{tr} have the same status, however, A_D and A_{tr} do not have the same status if $A_{tr} \prec D_{tr}$. This illustrates a case where two corresponding arguments with no **TopRule**, one from $AT_D^{5.5}$ and one from $AT_{tr}^{5.5}$, do not have the same status.

Next, we consider an example in which the **TopRule** is a strict rule, and the two corresponding arguments do not have the same status.

Example 5.6 Consider three theories, $AT_+^{5.6}$ with premises K_p and rules R , its defeasible version $AT_D^{5.6}$ with premises K_{p_D} and rules R_D , and its transposed version $AT_{tr}^{5.6}$ with premises $K_{p_{tr}}$ and rules R_{tr} :

$$\begin{array}{lll} K_p = \{a, \bar{d}\} & K_{p_D} = \{a, \bar{d}\} & K_{p_{tr}} = \{a, \bar{d}\} \\ R = \{a \rightarrow c; a \rightarrow d\} & R_D = \{a \Rightarrow c; a \Rightarrow d\} & R_{tr} = \{a \rightarrow c; a \rightarrow d; \bar{c} \rightarrow \bar{a}; \bar{d} \rightarrow \bar{a}\} \end{array}$$

We have the following arguments:

$$\begin{array}{lll} A = [a]; & A_D = [a]; & A_{tr} = [a]; \\ B = [A \rightarrow c]; & B_D = [A_D \Rightarrow c]; & B_{tr} = [A_{tr} \rightarrow c]; \\ C = [A \rightarrow d]; & C_D = [A_D \Rightarrow d]; & C_{tr} = [A_{tr} \rightarrow d]; \\ D = [\bar{d}]; & D_D = [\bar{d}]; & D_{tr} = [\bar{d}]; \\ & & E_{tr} = [D_{tr} \rightarrow \bar{a}] \end{array}$$

Assume $a \prec' d$, then $C_D \prec D_D$; $C_{tr} \prec D_{tr}$; $A_{tr} \prec E_{tr}$; $C_{tr} \prec E_{tr}$. Therefore, B_D is labeled **IN**, however, B_{tr} is labeled **OUT**. This illustrates a case where

two corresponding arguments, one from $AT_D^{5.6}$ and one from $AT_{tr}^{5.6}$, do not have the same status and the **TopRule** of the argument from $AT_{tr}^{5.6}$ is strict.

Finally, we have two examples illustrating the case where the **TopRule** is a defeasible rule.

Example 5.7 Consider three theories, $AT_+^{5.7}$ with premises K_p and rules R , its defeasible version $AT_D^{5.7}$ with premises K_{pD} and rules R_D , and its transposed version $AT_{tr}^{5.7}$ with premises K_{ptr} and rules R_{tr} :

$$\begin{array}{lll} K_p = \{a, \bar{c}\} & K_{pD} = \{a, \bar{c}\} & K_{ptr} = \{a, \bar{c}\} \\ R = \{a \Rightarrow b; b \rightarrow c\} & R_D = \{a \Rightarrow b; b \Rightarrow c\} & R_{tr} = \{a \Rightarrow b; b \rightarrow c; \bar{c} \rightarrow \bar{b}\} \end{array}$$

We have the following arguments:

$$\begin{array}{lll} A = [a]; & A_D = [a]; & A_{tr} = [a]; \\ B = [A \Rightarrow b]; & B_D = [A_D \Rightarrow b]; & B_{tr} = [A_{tr} \Rightarrow b]; \\ C = [B \rightarrow c]; & C_D = [B_D \Rightarrow c]; & C_{tr} = [B_{tr} \rightarrow c]; \\ D = [\bar{c}]; & D_D = [\bar{c}]; & D_{tr} = [\bar{c}]; \\ & & E_{tr} = [D_{tr} \rightarrow \bar{b}] \end{array}$$

If $a <' \bar{c}$, then B_D is **IN**, however, B_{tr} is **OUT**. If $\bar{c} <' a$, then both B_D and B_{tr} are **IN**. This illustrates a case where two corresponding arguments with a defeasible **TopRule**, one from $AT_D^{5.7}$ and one from $AT_{tr}^{5.7}$, have the same status and do not have the same status, respectively.

Example 5.8 Consider three theories, $AT_+^{5.8}$ with premises K_p and rules R , its defeasible version $AT_D^{5.8}$ with premises K_{pD} and rules R_D , and its

transposed version $AT_{tr}^{5.8}$ with premises $K_{p_{tr}}$ and rules R_{tr} :

$$\begin{array}{lll} K_p = \{a, b, d, e, \bar{f}\} & K_{p_D} = \{a, b, d, e, \bar{f}\} & K_{p_{tr}} = \{a, b, d, e, \bar{f}\} \\ R = \{a, b \Rightarrow c; d \Rightarrow \bar{b}; & R_D = \{a, b \Rightarrow c; d \Rightarrow \bar{b}; & R_{tr} = \{a, b \Rightarrow c; d \Rightarrow \bar{b}; \\ & d \rightarrow \bar{e}; a \rightarrow f\} & d \Rightarrow \bar{e}; a \Rightarrow f\} & d \rightarrow \bar{e}; e \rightarrow \bar{d}; a \rightarrow f; \bar{f} \rightarrow \bar{a}\} \end{array}$$

We have the following arguments:

$$\begin{array}{lll} A = [a]; & A_D = [a]; & A_{tr} = [a]; \\ B = [b]; & B_D = [b]; & B_{tr} = [b]; \\ C = [A, B \Rightarrow c]; & C_D = [A_D, B_D \Rightarrow c]; & C_{tr} = [A_{tr}, B_{tr} \Rightarrow c]; \\ D = [d]; & D_D = [d]; & D_{tr} = [d]; \\ E = [D \Rightarrow \bar{b}]; & E_D = [D_D \Rightarrow \bar{b}]; & E_{tr} = [D_{tr} \Rightarrow \bar{b}]; \\ F = [D \rightarrow \bar{e}]; & F_D = [D_D \Rightarrow \bar{e}]; & F_{tr} = [D_{tr} \rightarrow \bar{e}]; \\ G = [e]; & G_D = [e]; & G_{tr} = [e]; \\ H = [\bar{f}]; & H_D = [\bar{f}]; & H_{tr} = [\bar{f}]; \\ I = [A \rightarrow f]; & I_D = [A_D \rightarrow f]; & I_{tr} = [A_{tr} \rightarrow f]; \\ & & J_{tr} = [G_{tr} \rightarrow \bar{d}]; \\ & & K_{tr} = [H_{tr} \rightarrow \bar{a}]; \end{array}$$

Assume that the labeling of A_D, B_D, C_D is IN, OUT, OUT. If A_{tr}, B_{tr}, C_{tr} are IN, IN, IN, then C_D and C_{tr} do not have the same labeling. However, if A_{tr}, B_{tr}, C_{tr} are OUT, OUT, OUT, then C_D and C_{tr} do have the same labeling. This illustrates a case where two corresponding arguments with a defeasible TopRule, one from $AT_D^{5.8}$ and one from $AT_{tr}^{5.8}$, may or may not have the same status.

Taking Lemmas 5.1, 5.2 and 5.3 together, we have:

Proposition 5.2 Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its

transposed version AT_{tr} . Under the weakest link principle, for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ with defeasible version A_D , we have the following:

1. If $\text{TopRule}(A_{tr}) = \text{undefined}$,
 - (a) If $\text{Conc}(A_{tr})$ is an axiom, then A_D and A_{tr} have the same status.
 - (b) If $\text{Conc}(A_{tr})$ is an ordinary premise,
 - i. If $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then A_D and A_{tr} have the same status.
 - ii. If $\overline{\text{Conc}(A_{tr})}$ is the conclusion of a rule in AT_{tr} , then A_D and A_{tr} may not have the same status.
2. If $\text{TopRule}(A_{tr})$ is a strict rule,
 - (a) If all the arguments in $\text{Sub}(A_D)$ have the same labeling as the corresponding arguments in $\text{Sub}(A_{tr})$, then A_D and A_{tr} have the same status.
 - (b) If not all the arguments in $\text{Sub}(A_D)$ have the labeling as the corresponding arguments in $\text{Sub}(A_{tr})$, then A_D and A_{tr} may not have the same status.
3. If $\text{TopRule}(A_{tr})$ is a defeasible rule,

- (a) If all the arguments in $\text{Sub}(A_D)$ have the same status as the corresponding arguments in $\text{Sub}(A_{tr})$,
- i. If $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then A_D and A_{tr} have the same status.
 - ii. If $\overline{\text{Conc}(A_{tr})}$ is the conclusion of a rule in AT_{tr} , then A_D and A_{tr} may not have the same status.
- (b) If not all the arguments in $\text{Sub}(A_D)$ have the same status as the corresponding arguments in $\text{Sub}(A_{tr})$, then A_D and A_{tr} may not have the same status.

Proof The proof follows from the previous propositions and examples. 1(a) follows because in a well-defined theory, every axiom is in the set of justified conclusions; 1(b)(i) follows from Lemma 5.1; and 1(b)(ii) follows from Example 5.5 which shows a case in which the status is different. 2(a) follows from Lemma 5.2. 2(b) follows from Example 5.6 which shows a case in which the status is different. Finally, 3(a)(i) follows from Lemma 5.3; 3(a)(ii) follows from Example 5.7 which shows a case in which the status is different. and 3(b) follows from Example 5.8 which shows a case in which the status is different. \square

A more compact, but less explicit, way of stating the result in Proposition 5.2

is:

Corollary 5.1 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the weakest link principle, for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ with defeasible version A_D , then:*

- *If $\overline{\text{Conc}(A_{tr})}$ or any $\overline{\text{Conc}(\text{Sub}(A_{tr}))}$ is the conclusion of a rule in AT_{tr} , then A_{tr} may not have the same status as A_D ;*
- *Otherwise, A_{tr} will have the same status as A_D .*

Between them, Proposition 5.2 and Corollary 5.1 allow us to say for sure which arguments will definitely have the same status in $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$, and which may not. That is all arguments that neither have a conclusion that is the negation of the conclusion of a transposed strict rule, nor include a sub-argument with such a conclusion, will have the same status in $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$, while any argument which does not meet these conditions may not have the same status in $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$.

However, we can only make use of these results once we have constructed the set of arguments $\mathcal{A}(AT_{tr})$. To try to improve on this, we can go a step further and identify which subset of AT_{tr} we can be sure will generate arguments

that will have the same status as those from AT_D . As the details of the proofs of Lemmas 5.1–5.3 show, the reason that A_{tr} and A_D may have different status is that A_{tr} may be the subject of an attack by a transposed strict rule. Bearing this in mind, we adapt notation from [22], defining $\text{Atoms}_R(\mathcal{R})$ to be the atoms in a set of rules \mathcal{R} but unlike [22] we distinguish between atoms and their negations when we do this. Thus, if $\mathcal{R} = \{a, b \rightarrow c; \bar{a}, c \rightarrow d\}$ then $\text{Atoms}_R(\mathcal{R}) = \{a, \bar{a}, b, c, d\}$. Similarly, we can define $\text{Atoms}_K(\mathcal{K})$ to be the set of atoms in the knowledge-base \mathcal{K} , and we can further define $\text{Atoms}_T(AT)$ for an argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$ to be $\text{Atoms}_R(R) \cup \text{Atoms}_K(\mathcal{K})$.

Then:

Definition 5.2 (Safety) *Consider an ASPIC⁺ theory $AT_+ = \langle \langle \mathcal{L}, \mathcal{R}_+, n \rangle, \mathcal{K}_+ \rangle$ with a transposed version AT_{tr} . We say that $AT_S = \langle \langle \mathcal{L}, \mathcal{R}_S, n \rangle, \mathcal{K}_S \rangle$, such that $\mathcal{R}_S \subseteq \mathcal{R}_+$ and $\mathcal{K}_S \subseteq \mathcal{K}_+$, is safe with respect to AT_+ if $\text{Atoms}_T(AT_S) \cup \{\text{Conc}(r) \mid r \in (\mathcal{R}_{tr} - \mathcal{R}_+)\}$ is consistent.*

Thus a subset AT_S of an argumentation theory AT_+ is safe if the knowledge base and rules of AT_S are such that closing AT_+ under transposition does not lead to the knowledge base of AT_+ having any rules added to it with conclusions that are the negation of any atoms in AT_S and which thus might lead to attacks on arguments constructed from AT_S . In other words, AT_S

is safe if closing AT_+ under transposition does not open up any possible new lines of attack on arguments from AT_S because there are no atoms in AT_S that can be the subject of an attack from a transposed rule. Since no argument in $\mathcal{A}(AT_S)$ can be subject to any more attacks after closing it under transposition, every $A_S \in \mathcal{A}(AT_S)$ will have the same status in the transposed and defeasible versions of AT_+ . (That is the essence of a safe subset.) This gives us:

Proposition 5.3 *Consider an ASPIC⁺ theory $AT_+ = \langle \langle \mathcal{L}, \mathcal{R}_+, n \rangle, \mathcal{K}_+ \rangle$ with a transposed version AT_{tr} and a safe subset AT_S . Any argument $A_S \in \mathcal{A}(AT_{tr})$ will have the same status as its defeasible version A_D .*

Proof *As established in Proposition 5.2, A_S and A_D have the same status as long as A_S is not attacked by one of the transposed strict rules. Now, since AT_S is safe, it does not include the contrary atoms that feature in any of the transposed strict rules that are added to AT_+ to close it. Thus A_S cannot suffer any additional attacks, and so must have the same status as A_D . \square*

This leaves us in the following position. We use Proposition 5.3 to define which parts of theory AT , which is ill-defined, will give use the same results (the same arguments with the same status, and hence the same set of justified conclusions) whether it is made well-defined by transposition or by creating

its defeasible version. This will give us a rather conservative answer which amounts to “you can’t have strict rules which reference any other propositions in the theory”. Alternatively, we could use Proposition 5.2 or Corollary 5.1 which only flag up potential differences in status when there is an argument that clashes with a transposed strict rule. As a result, there will be situations in which Proposition 5.3 suggests a difference in status might exist when in fact none exists and one would not be predicted by Proposition 5.2 and Corollary 5.1. Such a situation is shown in Example 5.9. This will also be the case for any theory which contains strict rules and is its own transposed version because it already contains the transposed versions of every strict rule. (We considered such a case above.) However, to apply Proposition 5.2 and Corollary 5.1, we need to construct all the arguments from the theory which is unappealing from a computational perspective. Despite these issues, we have done what we set out to: we have characterized exactly when our two approaches to making a ASPIC^+ theory well-defined — translating into ASPIC_D^+ and completing the theory under the transposition of strict rules — will differ in terms of the conclusions that one can draw from the resulting theories under the grounded semantics.

Example 5.9 First consider the following theory $AT_+^{5.9}$:

$$K_p = \{a, \bar{c}\} \quad R = \{a \Rightarrow b; b \rightarrow c; a \Rightarrow d\}$$

In this case a safe subset of $AT_+^{5.9}$ is $AT_s^{5.9}$ where:

$$K_{p_s} = \{a, \bar{c}\} \quad R_s = \{b \rightarrow c, a \Rightarrow d\}$$

The rule $a \Rightarrow b$ is not included because it conflicts with the conclusion of the rule $\bar{c} \rightarrow \bar{b}$ which is the rule that would be introduced were $AT_+^{5.9}$ closed under transposition. The result, as we want from a safe theory, is that $\mathcal{A}(AT_s)$ includes no arguments that will have a different status than the corresponding arguments in $\mathcal{A}(AT_{tr})$. $\mathcal{A}(AT_s)$ makes sense in this case because the \bar{c} in the knowledge base might combine with the rule $\bar{c} \rightarrow \bar{b}$ to change the status of the argument for b that uses $a \Rightarrow b$. In the very similar theory $AT^{5.9}$:

$$K_p = \{\bar{c}\} \quad R = \{a \Rightarrow b; b \rightarrow c; a \Rightarrow d\}$$

a safe subset of $AT^{5.9}$ is $AT_{s'}^{5.9}$ where:

$$K_{p,s'} = \{a\} \quad R_{s'} = \{b \rightarrow c; a \Rightarrow d\}$$

In this theory, $a \Rightarrow b$ is still excluded because it conflicts with the conclusion of $\bar{c} \rightarrow \bar{b}$, despite the fact that there is no way — given the knowledge base — that an argument with conclusion \bar{b} could be constructed. But there is no

way of knowing this without constructing all the arguments. Of course, there is another safe subset of AT' , one which includes $a \Rightarrow b$ and excludes $b \rightarrow c$.

5.1.3 Under the preferred semantics

We have already established the relationship between the conclusions of ASPIC⁺ theories that are ill-defined and ASPIC_D⁺ theories under the grounded semantics. In this section, we will discuss the relationship under the preferred semantics and discuss how the results we give can be extended to other complete semantics.

We start by considering a special case of the preferred semantics, when the argumentation framework is relatively grounded. As described in Section 2.1.1, if a framework is relatively grounded, the grounded extension coincides with the intersection of the preferred extensions, and hence with the sceptical preferred extension. For a structured argumentation system, we can describe theories as being relatively grounded. As Section 2.1.2 shows, from a structured argumentation theory, we can derive an argumentation framework by constructing arguments and looking for attacks between them. Then we can define a theory AT as being relatively grounded if the corresponding framework $AF = \langle \mathcal{A}(AT), att, \preceq \rangle$ is relatively grounded. Then:

Corollary 5.2 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the weakest link principle, if AT_{tr} and AT_D are relatively grounded, then for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ that is in the sceptical preferred extension with defeasible version A_D also in the sceptical preferred extension, it is the case that:*

- *If $\overline{\text{Conc}(A_{tr})}$ or any $\overline{\text{Conc}(\text{Sub}(A_{tr}))}$ is the conclusion of a rule in AT_{tr} , then A_{tr} may not have the same status as A_D ;*
- *Otherwise, A_{tr} will have the same status as A_D .*

Proof *By definition, the sceptical preferred extension coincides with the grounded extension, and by Corollary 5.1, the above conditions hold for the grounded extension. Thus the conditions hold for the sceptical preferred extension. \square*

In other words, we can easily show that Corollary 5.1 holds for relatively grounded theories under the preferred semantics.

Extending the results to more general theories is hard. Using the same notation of AT_+ , AT_{tr} and AT_D , Example 4.6 has already shown that AT_D

may have more preferred extensions than AT_{tr} when AT_+ is well-defined. This makes it hard to compare the justified conclusions of the theories because such a comparison only makes sense between extensions that bear some relation to one another (the same issue limited our ability to extend our results beyond the grounded extension in the previous section). We make a slight change to Example 4.6 to show the same result for an AT_+ that is ill-defined:

Example 5.10 Consider three theories, $AT_+^{5.10}$ with premises K_p and rules R , its defeasible version $AT_D^{5.10}$ with premises K_{p_D} and rules R_D , and its transposed version $AT_{tr}^{5.10}$ with premises $K_{p_{tr}}$ and rules R_{tr} :

$$\begin{array}{lll} K_p = \{a, b\} & K_{p_D} = \{a, b\} & K_{p_{tr}} = \{a, b\} \\ R = \{a \rightarrow \bar{b}; b \rightarrow \bar{a}; \bar{c} \rightarrow \bar{a}\} & R_D = \{a \Rightarrow \bar{b}; b \Rightarrow \bar{a}; \bar{c} \Rightarrow \bar{a}\} & R_{tr} = \{a \rightarrow \bar{b}; b \rightarrow \bar{a}; \bar{c} \rightarrow \bar{a}; a \rightarrow c\} \end{array}$$

We have the following arguments:

$$\begin{array}{lll} A = [a]; & A_D = [a]; & A_{tr} = [a]; \\ B = [b]; & B_D = [b]; & B_{tr} = [b]; \\ C = [A \rightarrow \bar{b}] & C_D = [A \Rightarrow \bar{b}] & C_{tr} = [A \rightarrow \bar{b}] \\ D = [B \rightarrow \bar{a}] & D_D = [B \Rightarrow \bar{a}] & D_{tr} = [B \rightarrow \bar{a}] \\ & & E_{tr} = [A \rightarrow c] \end{array}$$

In this example, there are two preferred extensions for $AT_{tr}^{5.10}$ — $\{A_{tr}, C_{tr}, E_{tr}\}$ and $\{B_{tr}, D_{tr}\}$. However, there are three preferred extensions for $AT_D^{5.10}$ — $\{A_D, B_D\}$, $\{A_D, C_D\}$ and $\{B_D, D_D\}$.

In this case, then, the $ASPIC_D^+$ version of AT_+ has more extensions. Note that there are two preferred extensions which, in some sense, correspond.

That is $\{A_{tr}, C_{tr}, E_{tr}\}$ corresponds to $\{A_D, C_D\}$ because the justified conclusions overlap, and $\{B_{tr}, D_{tr}\}$ corresponds to $\{B_D, D_D\}$ because they have the same justified conclusions. The following example demonstrates that AT_{tr} may have more preferred extensions than AT_D .

Example 5.11 Consider three theories, $AT_+^{5.11}$ with premises K_p and rules R , its defeasible version $AT_D^{5.11}$ with premises K_{p_D} and rules R_D , and its transposed version $AT_{tr}^{5.11}$ with premises $K_{p_{tr}}$ and rules R_{tr} :

$$\begin{array}{lll} K_p = \{a, c\} & K_{p_D} = \{a, c\} & K_{p_{tr}} = \{a, c\} \\ R = \{\bar{b} \rightarrow \bar{c}; b \Rightarrow \bar{a}\} & R_D = \{\bar{b} \Rightarrow \bar{c}; b \Rightarrow \bar{a}\} & R_{tr} = \{\bar{b} \rightarrow \bar{c}; b \Rightarrow \bar{a}; c \rightarrow b\} \end{array}$$

We have the following arguments:

$$\begin{array}{lll} A = [a]; & A_D = [a]; & A_{tr} = [a]; \\ B = [c]; & B_D = [c]; & B_{tr} = [c]; \\ & & C_{tr} = [B_{tr} \rightarrow b] \\ & & D_{tr} = [C_{tr} \Rightarrow \bar{a}] \end{array}$$

In this example, if A_{tr} and D_{tr} are equally preferred, there are two preferred extensions for $AT_{tr}^{5.11}$ — $\{A_{tr}, B_{tr}, C_{tr}\}$ and $\{B_{tr}, C_{tr}, D_{tr}\}$. However, there is only one preferred extension for $AT_D^{5.11}$ — $\{A_D, B_D\}$.

In this case, there is one pair of preferred extensions that correspond since $\{A_{tr}, B_{tr}, C_{tr}\}$ corresponds to $\{A_D, B_D\}$.

The above examples demonstrate that the number of preferred extensions in AT_D may be greater than or less than the number of preferred extensions

in AT_{tr} . Thus, in general, it is not the case that every extension of AT_D has a corresponding extension in the set of extensions of AT_{tr} , and vice versa. This, in turn, makes it hard to say anything about whether a particular argument will be sceptically or credulously justified in both the defeasible and transposed versions of some AT_+ . The best we can do is to give the following results to illustrate the relations between particular members of the set of the preferred extensions of AT_D and AT_{tr} .

Lemma 5.4 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the weakest link principle, for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ with defeasible version A_D , if $\text{TopRule}(A_{tr}) = \text{undefined}$ and $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then there exists a preferred extension E_D of AT_D and a preferred extension E_{tr} of AT_{tr} such that A_D and A_{tr} have the same status.*

Proof *Since there is no $\text{TopRule}(A_{tr})$, the only possible attack on A_{tr} is a rebut. However, there is no rule with conclusion $\overline{\text{Conc}(A_{tr})}$ by definition, and the only difference between AT_D and AT_{tr} is the existence of rules that were added to AT_{tr} to close it under transposition, therefore there can be no*

additional attacks in AT_{tr} beyond those in AT_D and so A_D and A_{tr} have exactly same defeaters. Thus, whatever the extensions are, there exists at least one preferred extension E_D of AT_D and at least one preferred extension E_{tr} of AT_{tr} such that A_D and A_{tr} have the same status. \square

Lemma 5.5 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the weakest link principle, for any argument $A_{tr} \in \mathcal{A}(AT_{tr})$ with defeasible version A_D , if $\text{TopRule}(A_{tr})$ is a strict rule, and $\text{Sub}(A_D)$ and $\text{Sub}(A_{tr})$ have the same status, then there exists a preferred extension E_D of AT_D and a preferred extension E_{tr} of AT_{tr} such that A_D and A_{tr} have the same status.*

Proof Recall that the difference between AT_D and AT_{tr} is that AT_D has more attacks generated by unrestricted rebut, while AT_{tr} has more transposed rules, which may generate more arguments, and these arguments will generate more attacks. Thus either set of arguments may include attacks that the other does not (which, in turn, is why we find it so hard to relate the status of arguments from those two sets of arguments.) Recall also the notation in the proof of Proposition 5.2. Here we are comparing AT_D with AT_{tr} . The attack relations

of AT_{tr} and AT_D are shown in Figure 5.5(a) and 5.5(b). Consider the 3 preference orderings in AT_{tr} and AT_D .

1. $B_{tr} \prec A_{tr}$. As in the proof of Proposition 5.2, since B_{tr} is a sub-argument of C_i^{tr} we know that $C_i^{tr} \prec A_{tr}$. The defeat relations for AT_{tr} and AT_D are shown in Figure 5.5(c) and 5.5(d). The additional defeats (A_{tr}, C_i^{tr}) have no effect.
2. $A_{tr} \prec B_{tr}$. As in the proof of Proposition 5.2, since C_i^{tr} is constructed from A and B we know that $A_{tr} \prec C_i^{tr}$. The defeat relations for AT_{tr} and AT_D are shown in Figure 5.5(e) and 5.5(f). If B_{tr} is labeled **IN** in E_{tr} , then the attackers of C_i^{tr} are:

- the attackers of A_{tr} (attacks from A_i^{tr}); or
- C_j^{tr} ($i \neq j$) ($\text{Conc}(C_j^{tr})$ is \bar{a}_j); or
- defended in E_{tr} (B_{tr} is labeled **IN**).

In all cases, either $\{A_{tr}, C_1^{tr}, \dots, C_n^{tr}\}$ are all labeled **OUT** in E_{tr} or there are several preferred extensions where A_{tr} is labeled **OUT** in at least one of them. If B_{tr} is labeled **OUT** or **UNDEC** in E_{tr} , then all C_i^{tr} are labeled **OUT** in E_{tr} since B_{tr} is a sub-argument of C_i^{tr} . The additional defeats (C_i^{tr}, A_{tr}) have no effect.

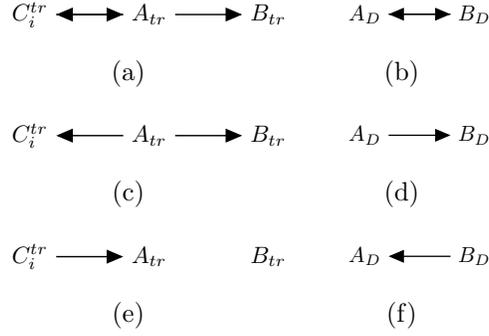


Figure 5.5: Attack and defeat relations for AT_{tr} and AT_D from the proof of Proposition 5.4. (a) Attack relations for AT_{tr} and Defeat relations for AT_{tr} in case 3, (b) Attack relations for AT_D and Defeat relations for AT_D in case 3, (c) Defeat relations for AT_{tr} in case 1, (d) Defeat relations for AT_D in case 1, (e) Defeat relations for AT_{tr} case 2, (f) Defeat relations for AT_D case 2

3. $A_{tr} \sim B_{tr}$. The defeat relations for AT_{tr} and AT_D are the same as the attack relations, shown in Figure 5.5(a) and 5.5(b). If B_{tr} is labeled **IN** in E_{tr} , then A_{tr} must be labeled **OUT** in E_{tr} . If B_{tr} is labeled **UNDEC** in E_{tr} , then A_{tr} cannot be labeled **IN** in E_{tr} . Therefore, neither A_{tr} nor B_{tr} is in the extension E_{tr} . If B_{tr} is labeled **OUT** in E_{tr} , then C_i^{tr} must be labeled **OUT** (for the same reason as the previous case). Therefore, the additional defeats (C_i^{tr}, A_{tr}) and (A_{tr}, C_i^{tr}) have no effect.

Therefore, in all cases, there exists at least one preferred extension E_D of AT_D and at least one preferred extension E_{tr} of AT_{tr} where A_{tr} and A_D have the same status provided their sub-arguments have the same status. \square

Lemma 5.6 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its transposed version AT_{tr} . Under the weakest link principle, for any argument A_{tr} , if $\text{TopRule}(A_{tr})$ is a defeasible rule, $\text{Sub}(A_D)$ and $\text{Sub}(A_{tr})$ have the same status, and $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then there exists a preferred extension E_D of AT_D and a preferred extension E_{tr} of AT_{tr} such that A_D and A_{tr} have the same status.*

Proof *Since all the arguments in $\text{Sub}(A_D)$ have the same status as the corresponding arguments in $\text{Sub}(A_{tr})$, there are no additional defeats on the sub-argument of A_{tr} . Therefore, the only possible additional defeat is due to a rebut on the conclusion. However, there are no rules with conclusion $\overline{\text{Conc}(A_{tr})}$, so A_D and A_{tr} have exactly same defeaters. Therefore, there exists at least one preferred extension E_D of AT_D and at least one preferred extension E_{tr} of AT_{tr} such that A_D and A_{tr} have the same status. \square*

Combining these results we have:

Proposition 5.4 *Consider an ASPIC⁺ theory AT_+ which is not closed under transposition, its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and its trans-*

posed version AT_{tr} . Under the weakest link principle, there exists a preferred extension E_D of AT_D and a preferred extension E_{tr} of AT_{tr} such that for any argument $A_{tr} \in E_{tr}$ with defeasible version A_D , we have the following:

1. If $\text{TopRule}(A_{tr}) = \text{undefined}$,
 - (a) If $\text{Conc}(A_{tr})$ is an axiom, then A_D and A_{tr} have the same status.
 - (b) If $\text{Conc}(A_{tr})$ is an ordinary premise,
 - i. If $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then A_D and A_{tr} have the same status.
 - ii. If $\overline{\text{Conc}(A_{tr})}$ is the conclusion of a rule in AT_{tr} , then A_D and A_{tr} may not have the same status.
2. If $\text{TopRule}(A_{tr})$ is a strict rule,
 - (a) If all the arguments in $\text{Sub}(A_D)$ have the same labeling as the corresponding arguments in $\text{Sub}(A_{tr})$, then A_D and A_{tr} have the same status.
 - (b) If not all the arguments in $\text{Sub}(A_D)$ have the labeling as the corresponding arguments in $\text{Sub}(A_{tr})$, then A_D and A_{tr} may not have the same status.
3. If $\text{TopRule}(A_{tr})$ is a defeasible rule,

- (a) If all the arguments in $\text{Sub}(A_D)$ have the same status as the corresponding arguments in $\text{Sub}(A_{tr})$,
- i. If $\overline{\text{Conc}(A_{tr})}$ is not the conclusion of any rule in AT_{tr} , then A_D and A_{tr} have the same status.
 - ii. If $\overline{\text{Conc}(A_{tr})}$ is the conclusion of a rule in AT_{tr} , then A_D and A_{tr} may not have the same status.
- (b) If not all the arguments in $\text{Sub}(A_D)$ have the same status as the corresponding arguments in $\text{Sub}(A_{tr})$, then A_D and A_{tr} may not have the same status.

Proof The proof follows from the previous propositions and examples. 1(a) follows because in a well-defined theory, every axiom is in the set of justified conclusions; 1(b)(i) follows from Lemma 5.4; and 1(b)(ii) follows from Example 5.5 which shows a case in which the status is different. 2(a) follows from Lemma 5.5. 2(b) follows from Example 5.6 which shows a case in which the status is different. Finally, 3(a)(i) follows from Lemma 5.6; 3(a)(ii) follows from Example 5.7 which shows a case in which the status is different. and 3(b) follows from Example 5.8 which shows a case in which the status is different.

□

What this means is that, for any extension E_{tr} in AT_{tr} , if E_{tr} does not contain any arguments constructed using transposed rules, then there is an extension E_D in AT_D , such that E_D and E_{tr} satisfy Proposition 5.4. The reason is that any additional preferred extensions of AT_{tr} are created by additional transposed rules, and if all the arguments constructed by transposed rules are labeled **OUT**, then there is a corresponding extension E_D in AT_D which satisfies the conditions in Proposition 5.4. Similarly, for any extension E_D in AT_D , if any subset of AT_D is not one of the cases shown in Example 4.6, then there is an extension E_{tr} in AT_{tr} , such that E_D and E_{tr} satisfy the conditions in Proposition 5.4. The reason is that the only situation in which AT_D may have more preferred extension than AT_{tr} is shown in Example 4.6. If none of the subsets of AT_{tr} contain this case, we know that every preferred extension in AT_D has a corresponding extension E_{tr} in AT_{tr} which satisfies Proposition 5.4.

As any stable extension is a semi-stable extension, and any semi-stable extension is a preferred extension, an analogous result to Proposition 5.4 holds for semi-stable extensions and stable extensions. Note that, the stable extension does not always exist. Furthermore, it is possible that every preferred extension is a stable extension, so we cannot say anything more than Proposition 5.4.

5.2 Theories that are axiom inconsistent

Under the definition of a well-defined ASPIC⁺ theory, Definition 2.31, a theory will be well-defined iff it is closed under transposition, and axiom consistent. The previous section investigated the relationship between the ways in which ASPIC⁺ and ASPIC_D⁺ handle theories that are not closed under transposition. In this section, we will consider the other case that makes ASPIC⁺ theories ill-defined — when they are axiom inconsistent.

An ASPIC⁺ theory is axiom inconsistent iff its strict part is not consistent. In other words, there will be at least two strict and firm arguments such that the conclusion of one is the negation of the conclusion of the other. There are four ways that this can occur:

Definition 5.3 (Axiom Inconsistent Theory) *Consider an ASPIC⁺ theory AT . AT is axiom inconsistent iff there exist two strict and firm arguments A and B in $\mathcal{A}(AT)$, such that $\text{Conc}(A) = \overline{\text{Conc}(B)}$. This will be the case iff one of the following cases holds:*

1. *Both A and B only consist of an axiom, i.e., $\text{TopRule}(A) = \text{undefined}$ and $\text{TopRule}(B) = \text{undefined}$*
2. *A consists of an axiom, B is constructed using axioms and strict rules, i.e., $\text{TopRule}(A) = \text{undefined}$ and $\text{TopRule}(B) \neq \text{undefined}$*

3. Both A and B are constructed using axioms and strict rules, i.e.,
 $\text{TopRule}(A) \neq \text{undefined}$ and $\text{TopRule}(B) \neq \text{undefined}$

Since strict and firm arguments cannot be attacked, the inconsistent axiom and its negation will be justified conclusions of the theory. Thus the set of justified conclusions will not be consistent, and so the theory is not rational. However, if we convert the theory to its defeasible version, it will become rational, since all ASPIC_D^+ theories are rational. The main question for us, which is analagous to the question we answered in Section 5.1, is the following. Given an ASPIC^+ theory AT_+ , which is axiom inconsistent, when and how the justified conclusions differ between AT_D , the ASPIC_D^+ version of AT_+ , and AT_C , the version of AT that has been made axiom consistent? Naturally this depends on *how* the inconsistency is resolved. We consider three simple approaches which remove elements of AT_+ that lead to the inconsistency. Clearly any operator from the belief revision literature [42] could be used here. The ones we choose, as we shall see, have the property of aligning, in some sense, with the use of ASPIC_D^+ .

Considering Definition 5.3, the first of these approaches is formalised as follows:

Definition 5.4 (Belief Revision 1) *The function $BR_1(\cdot)$ converts an ax-*

iom inconsistent theory AT_+ into an axiom consistent theory $AT_C = BR_1(AT_+)$ if for every pair of strict and firm arguments A and B in $\mathcal{A}(AT_+)$, such that $\text{Conc}(A) = \overline{\text{Conc}(B)}$:

1. If $\text{TopRule}(A) = \text{undefined}$ and $\text{TopRule}(B) = \text{undefined}$, then $AT_C = AT_+ \setminus \{\text{Prem}(A), \text{Prem}(B)\}$
2. If $\text{TopRule}(A) = \text{undefined}$ and $\text{TopRule}(B) \neq \text{undefined}$, then $AT_C = AT_+ \setminus \{\text{Prem}(A), \text{TopRule}(B)\}$
3. If $\text{TopRule}(A) \neq \text{undefined}$ and $\text{TopRule}(B) \neq \text{undefined}$, then $AT_C = AT_+ \setminus \{\text{TopRule}(A), \text{TopRule}(B)\}$

Thus $BR_1(\cdot)$ makes an axiom inconsistent theory AT consistent by removing the elements that directly clash. If there is an axiom that is the negation of the conclusion of a strict and firm argument, that axiom is removed. If there is a strict rule with a conclusion that is the negation of a strict and firm argument, then that rule is removed. Now:

Proposition 5.5 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ and its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ . Under the elitist weakest link principle, for every pair of conflicting strict and firm arguments $A, B \in$*

$\mathcal{A}(AT_+)$, the corresponding defeasible arguments $A_D, B_D \in \mathcal{A}(AT_D)$ will not be justified under the grounded semantics.

Proof Call any two of the conflicting strict and firm arguments of $\mathcal{A}(AT_+)$ A and B . Without loss of generality, we assume there is no argument attacking the strict sub-argument of A and B . (Such an argument would have to be strict and firm since a defeasible or plausible argument cannot attack a strict and firm argument.) If there is an argument C , where C attacks A' , and A' is a strict sub-argument of A , then rename C as B and A' as A , and continue until neither A nor B has a attacking sub-argument. In AT_D , we have the defeasible versions of these arguments A_D and B_D , which attack each other, and there are no other arguments attacking A_D or B_D . In addition, A_D and B_D are equally preferred. Therefore, the only defeater of A_D is B_D and the only defeater of B_D is A_D . Under the grounded semantics, therefore, both A_D and B_D are labeled UNDEC. Furthermore, any argument that contains A_D as a sub-argument is defeated by B_D , and any argument that contains B_D as sub-argument is defeated by A_D . Since both A_D and B_D are labeled UNDEC, the arguments with either A_D or B_D as sub-arguments are labeled UNDEC or OUT³. The same holds for any other conflicting pairs of strict and

³If there are no IN attackers, the argument is labeled UNDEC, otherwise it is labeled OUT.

firm arguments A' and B' . Thus the defeasible versions of all the conflicting strict and firm arguments of $\mathcal{A}(AT_+)$ will not be justified under the grounded semantics. \square

So any arguments in $\mathcal{A}(AT_D)$ that are the defeasible version of conflicting strict and firm arguments in $\mathcal{A}(AT_+)$ won't be justified. This establishes a relationship between $\mathcal{A}(AT_+)$ and $\mathcal{A}(AT_D)$. We can go further:

Proposition 5.6 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ , its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and $AT_C = BR_1(AT_+)$. Under the elitist weakest link principle, for every argument A in AT_+ for which there is a $p \in \text{Prem}(A)$ but $p \notin AT_C$, or for which there is a $r \in \mathcal{R}(A)$ but $r \notin AT_C$, the corresponding defeasible argument A_D will not be justified under the grounded semantics.*

Proof *Consider an argument A_D which has a premise $p \in \text{Prem}(A_D)$ such that $p \notin AT_C$. This must be an element that was removed from AT_+ by applying BR_1 . Thus, by Definition 5.4, AT_D either contains \bar{p} or an argument with conclusion \bar{p} . Both premise and argument, if they exist, will be at the highest preference level (because in AT_+ they were strict and firm), and so A_D will be UNDEC under the grounded semantics, and thus will not be justified.*

Now consider that A_D includes a rule $r \notin AT_C$. Again, since this was removed by BR_1 , there is an attacking argument at the highest preference level which rebuts A_D on r . This time A_D may be **OUT** (if it contains elements that are not at the highest preference level), but whether it is **UNDEC** or **OUT** it will not be justified. \square

This says that although AT_D includes the defeasible version of the knowledge that made AT_+ axiom inconsistent, this knowledge cannot be used to create any arguments that are justified. So the effects of the inconsistency are isolated — the result is the creation of some new **UNDEC** arguments. Indeed, we can show:

Proposition 5.7 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ , its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and $AT_C = BR_1(AT_+)$. Under the elitist weakest link principle and the grounded semantics, the justified conclusions of AT_D are the same as the justified conclusions of AT_C .*

Proof We know, from Proposition 4.7 that $\mathcal{A}(AT_D)$ includes exactly one argument A_D for each argument A_+ of $\mathcal{A}(AT_+)$, such that A_D is the defeasible version of A_+ . Furthermore, together Definition 5.3 and 5.4 tell us that the difference between AT_+ and AT_C is that AT_C does not contain the elements

necessary to create the conflicting strict and firm arguments of AT_+ that lead to AT_+ being axiom inconsistent. So the difference between $\mathcal{A}(AT_+)$ and $\mathcal{A}(AT_C)$ will be that $\mathcal{A}(AT_C)$ does not include any $A \in \mathcal{A}(AT_+)$ that includes one of the elements that is in AT_+ but not AT_C . These elements are exactly the premises and rules removed by BR_1 . Thus $\mathcal{A}(AT_D)$ will include defeasible versions of all the arguments in $\mathcal{A}(AT_C)$ and defeasible versions of all the arguments in $\mathcal{A}(AT_+)$ which include elements removed by BR_1 . Let's call the set of arguments which are the defeasible versions of the arguments in $\mathcal{A}(AT_+) - \mathcal{A}(AT_C)$ by the name \mathbf{A}_{BR} . Now, consider the justified conclusions of $\mathcal{A}(AT_D)$ and $\mathcal{A}(AT_C)$. Proposition 5.6 tells us that none of the arguments in \mathbf{A}_{BR} will be justified under the grounded semantics, so they won't be responsible for any additional justified conclusions of AT_D beyond those of AT_C . In other words, AT_D has no justified conclusions that are not justified conclusions of AT_C . Furthermore, since all of the arguments in \mathbf{A}_{BR} will be labelled **OUT** or **UNDEC**, any attacks that they make on other arguments in $\mathcal{A}(AT_D)$, arguments that are the defeasible version of arguments in $\mathcal{A}(AT_C)$, cannot prevent those arguments being **IN**. Thus AT_C has no justified conclusions that are not justified conclusions of AT_D . Thus the result holds. \square

Thus, under the grounded semantics, the justified conclusions of AT_D are

exactly same as the justified conclusions of AT_C . This means that if we use $ASPIC_D^+$ there is no need to worry about axiom inconsistent theories — we get exactly the same justified conclusions as if we had removed the source of the inconsistency (albeit if the removal was done by BR_1).

Now let's consider the the sceptically preferred conclusions:

Proposition 5.8 *Consider an axiom inconsistent $ASPIC^+$ theory AT_+ and its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ . Under the elitist weakest link principle, for every pair of conflicting strict and firm arguments $A, B \in \mathcal{A}(AT_+)$, the corresponding defeasible arguments $A_D, B_D \in \mathcal{A}(AT_D)$ will not be in the sceptically preferred extension.*

Proof *This proof proceeds like the proof of Proposition 5.5. The only difference is that A_D and B_D will not be labelled UNDEC, rather there will be pairs of extensions in one of which A_D is IN and B_D is OUT, and in the other of which A_D is OUT and B_D is IN. This, of course, means that neither A_D nor B_D will be in the sceptically preferred extension. \square*

Proposition 5.9 *Consider an axiom inconsistent $ASPIC^+$ theory AT_+ , its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and $AT_C = BR_1(AT_+)$. Under the*

elitist weakest link principle, for every argument A in AT_+ for which there is a $p \in \text{Prem}(A)$ but $p \notin AT_C$, or for which there is a $r \in \mathcal{R}(A)$ but $r \notin AT_C$, the corresponding defeasible argument A_D will not be in the sceptically preferred extension.

Proof *This proof proceeds just like that of Proposition 5.6, except that cases in which A_D was UNDEC under the grounded semantics will now become cases in which there are multiple extensions under the preferred semantics such that A_D is IN in some extensions and OUT in others. However, this means that A_D will not be in the sceptically preferred extension. \square*

Thus we have analogues of Propositions 5.5 and 5.6, and we can show that none of the elements that cause the axioms to be inconsistent can be involved in arguments that lead to sceptically preferred conclusions. We cannot, however, show that the sceptically preferred conclusions are the same for AT_C and AT_D . Rather we can show that:

Proposition 5.10 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ , its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and $AT_C = BR_1(AT_+)$. Under the elitist weakest link principle the sceptically justified conclusions of AT_D are a subset of the sceptically preferred conclusions of AT_C .*

Proof As in the proof of Proposition 5.7, $\mathcal{A}(AT_D)$ will include defeasible versions of all the arguments in $\mathcal{A}(AT_C)$ and defeasible versions of all the arguments in $\mathcal{A}(AT_+)$ which include elements removed by BR_1 . Again, we will call the set of arguments which are the defeasible versions of the arguments in $\mathcal{A}(AT_+) - \mathcal{A}(AT_C)$ by the name \mathbf{A}_{BR} . Now, consider the justified conclusions of $\mathcal{A}(AT_D)$ and $\mathcal{A}(AT_C)$. Proposition 5.9 tells us that none of the arguments in \mathbf{A}_{BR} will be in the sceptically justified extension. In other words, AT_D has no arguments in its sceptically justified extension that are not in the sceptically justified extension of AT_C . Now (and here we diverge from the line of proof of Proposition 5.7), the arguments in \mathbf{A}_{BR} will be labelled **IN** in some preferred extensions. (As we know from before, such arguments come in mutually attacking pairs, and each one will be **IN** in at least one preferred extension.) When such arguments are **IN**, any attacks that they make on other arguments in $\mathcal{A}(AT_D)$, arguments that are the defeasible version of arguments in $\mathcal{A}(AT_C)$, may make those arguments **OUT** when their counterparts in all the preferred extensions of AT_C are **IN**. Thus there may be arguments in $\mathcal{A}(AT_D)$ that are not in the sceptically justified extension of AT_D but whose counterparts are in the sceptically justified extension of AT_C . Thus the sceptically justified conclusions of AT_D are a subset of the sceptically justified conclusions of AT_C . \square

Thus, unlike the case for the grounded semantics, if we translate a theory with inconsistent axioms, AT_+ , into the $ASPIC_D^+$ theory AT_D , we will get a different set of (sceptically) justified conclusions than if we revise the theory with $BR_1(\cdot)$ to remove the inconsistency. Given what we already know about justified conclusions and $ASPIC_D^+$, this is not a surprise, but it is helpful to know that the difference arises because while AT_D will not have any more justified conclusions than the revised theory (which we might take as a form of soundness result), it might not be able to draw all the same conclusions (a kind of incompleteness).

However, if we use a different function for belief revision, it turns out that we will be able to draw the same sceptically justified conclusions from AT_+ and AT_D . We define:

Definition 5.5 (Belief Revision 2) *The function $BR_2(\cdot)$ converts an axiom inconsistent theory AT_+ into an axiom consistent theory $AT_C = BR_2(AT_+)$ if for every pair of strict and firm arguments A and B in $\mathcal{A}(AT_+)$, such that $\text{Conc}(A) = \overline{\text{Conc}(B)}$, $AT_C = AT_+ \setminus (\mathcal{R}_A \cup \mathcal{R}_B \cup \mathcal{K}_A \cup \mathcal{K}_B)$:*

1. $\mathcal{R}_A = \{r \in \mathcal{R} \mid \text{conc}(r) = \text{Conc}(A)\}$
2. $\mathcal{R}_B = \{r \in \mathcal{R} \mid \text{conc}(r) = \text{Conc}(B)\}$
3. $\mathcal{K}_A = \{p \in \mathcal{K} \mid p = \text{Conc}(A)\}$

$$4. \mathcal{K}_B = \{p \in \mathcal{K} \mid p = \text{Conc}(B)\}$$

where $\text{conc}(r)$ denotes the conclusion of the rule r .

The difference between $BR_2(\cdot)$ and $BR_1(\cdot)$ is that $BR_1(\cdot)$ removes just the strict elements that conflict, whereas $BR_2(\cdot)$ is less conservative and also removes any defeasible elements (premises or rules) with the same conclusion as either of the strict elements. We then have:

Proposition 5.11 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ , its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ , and $AT_C = BR_2(AT_+)$. Under the elitist weakest link principle, the sceptically justified conclusions of AT_D are same as the sceptically justified conclusions of AT_C .*

Proof *The proof again starts like that of Proposition 5.7. We know, from Proposition 4.7 that $\mathcal{A}(AT_D)$ includes exactly one argument A_D for each argument A_+ of $\mathcal{A}(AT_+)$, such that A_D is the defeasible version of A_+ . Furthermore, together Definition 5.3 and 5.5 tell us that the difference between AT_+ and AT_C is that AT_C does not contain the strict elements necessary to create the conflicting strict and firm arguments of AT_+ that lead to AT_+ being axiom inconsistent, or any defeasible elements that have the same conclusions as these strict elements. The difference between $\mathcal{A}(AT_+)$ and $\mathcal{A}(AT_C)$ will be*

that $\mathcal{A}(AT_C)$ does not include any $A \in \mathcal{A}(AT_+)$ that includes one of the elements that is in AT_+ but not AT_C . These elements are exactly the premises and rules removed by BR_2 . Thus $\mathcal{A}(AT_D)$ will include defeasible versions of all the arguments in $\mathcal{A}(AT_C)$ and defeasible versions of all the arguments in $\mathcal{A}(AT_+)$ which include elements removed by BR_2 . Clearly any differences between the sceptically justified conclusions of AT_D and AT_C are going to be down to the arguments that are constructed from these elements. Consider, then an argument A_D which has a premise $p \in \text{Prem}(A_D)$ such that $p \notin AT_C$. This premise p must be an element that was removed from AT_+ by applying BR_2 . Thus, by Definition 5.5, AT_D either contains at least one premise \bar{p} or at least one argument with conclusion \bar{p} (there may be several, because $BR_2(\cdot)$ removes all such premises and rules). Now, of these premises and arguments that are not in AT_C but whose defeasible version(s) are in AT_D , at least one will be at the highest preference level (because in AT_+ it was included in a strict and firm argument), and so A_D will be defeated and thus will not be in the sceptically justified conclusions. The same holds for an A_D which includes a rule r that is not in AT_C . In other words, AT_D has no arguments in its sceptically justified extension that are not in the sceptically justified extension of AT_C .

That gives us one half of the result. To get the other half we have to con-

sider whether there are any arguments in the sceptically justified conclusions of AT_C that are not in the sceptically justified conclusions of AT_D . If we, as before, call the set of arguments which are the defeasible versions of the arguments in $\mathcal{A}(AT_+) - \mathcal{A}(AT_C)$ by the name \mathbf{A}_{BR} , then we are considering if any of the arguments in \mathbf{A}_{BR} defeat any of the arguments in the defeasible version of $\mathcal{A}(AT_C)$. In theory, this might happen (as in the proof of Proposition 5.10) because there are arguments A_{BR} in \mathbf{A}_{BR} which are labeled **IN** in some preferred extensions and so force some other arguments A'_D , that are not in A_{BR} , to be **OUT** in those extensions, while there are also extensions in which A_{BR} are **OUT**, and so A'_D are **IN**. (In such a case, neither the A_{BR} nor the A'_D will be in the sceptically preferred extension of AT_D , but the non-defeasible versions of the A'_D will be in the sceptically preferred extension of AT_C . However, in this case, we can show that any argument in \mathbf{A}_{BR} will not attack any argument in $\mathcal{A}(AT_D)$, since:

- for any argument $A_{BR} \in \mathbf{A}_{BR}$, there is no argument in $\mathcal{A}(AT_C)$ which has the conclusion $\overline{\text{Conc}(A_{BR})}$, since $BR_2(\cdot)$ removed any such A_{BR} from AT_C . So A_{BR} does not rebut or undermine any argument in the defeasible version of $\mathcal{A}(AT_C)$.
- for any argument $A_{BR} \in \mathbf{A}_{BR}$, $\text{Conc}(A_{BR})$ is not the name of a defeasi-

ble rule since names are not subject to the revision function. Thus A_{BR} does not undercut any argument in the defeasible version of $\mathcal{A}(AT_C)$.

Thus there are no sceptically justified conclusions of AT_C that are not sceptically justified conclusions of AT_D , and the result holds. \square

So the advantage of using the less conservative revision operator $BR_2(\cdot)$ is that the revised theory AT_C has exactly the same sceptically justified conclusions as the defeasible version of the original theory AT_D .

Given the above, we have revision operators for an axiom inconsistent theory AT_+ that will create axiom consistent theories AT_C which have the same justified conclusions as the defeasible version of AT_+ whether “justified conclusion” is determined using the grounded extension or the sceptically preferred extension. A natural question to ask is whether there is a revision function that has a similar alignment with the preferred extensions. It turns out that there is. We call the function $BR_3(\cdot)$. Whereas $BR_1(\cdot)$ and $BR_2(\cdot)$ revise AT_+ by removing all the conflicting elements to create a single axiom-consistent theory, the idea is that for each conflicting pair of elements $BR_3(\cdot)$ creates a pair of theories, each of which contains one of the conflicting elements:

Definition 5.6 (Belief Revision, Pairwise) *The function $BR_*(\cdot)$ converts an axiom inconsistent theory AT_+ into a pair of theories AT' and AT'' :*

$$BR_*(AT_+) = \{AT', AT''\}$$

such that for a pair of strict and firm arguments A and B in $\mathcal{A}(AT_+)$, where $\text{Conc}(A) = \overline{\text{Conc}(B)}$:

1. *If $\text{TopRule}(A) = \text{undefined}$ and $\text{TopRule}(B) = \text{undefined}$, then $AT' = AT \setminus \{\text{Prem}(A)\}$ and $AT'' = AT \setminus \{\text{Prem}(B)\}$*
2. *If $\text{TopRule}(A) = \text{undefined}$ and $\text{TopRule}(B) \neq \text{undefined}$, then $AT' = AT \setminus \{\text{Prem}(A)\}$ and $AT'' = AT \setminus \{\text{TopRule}(B)\}$*
3. *If $\text{TopRule}(A) \neq \text{undefined}$ and $\text{TopRule}(B) \neq \text{undefined}$, then $AT' = AT \setminus \{\text{TopRule}(A)\}$ and $AT'' = AT \setminus \{\text{TopRule}(B)\}$*

Of course, AT_+ may be axiom inconsistent in such a way that just resolving one conflict between two strict arguments is not enough to render it axiom consistent, so we need to apply $BR_*(\cdot)$ to the source of every conflict:

Definition 5.7 (Belief Revision 3) *The function $BR_+(\cdot) : AT \mapsto \mathbf{AT}$, where \mathbf{AT} denotes a set of argumentation theories, is defined as:*

$$BR_+(AT) = \begin{cases} BR_*(AT) & \text{if } AT \text{ is axiom inconsistent,} \\ \{AT\}, & \text{otherwise} \end{cases}$$

then $BR_{\#}(\cdot) : \mathbf{AT} \mapsto \mathbf{AT}$ is the function:

$$BR_{\#}(\mathbf{AT}) = \bigcup_i BR_+(AT_i), \forall AT_i \in \mathbf{AT}$$

$BR_3(\cdot)$ is then the fixpoint combinator for $BR_{\#}$:

$$BR_3(AT) = BR_{\#}(BR_{\#}(BR_{\#}(\dots BR_{\#}(\{AT\}) \dots)))$$

Thus $BR_+(AT)$ returns AT if AT is axiom consistent, otherwise it returns the two argumentation theories generated by $BR_*(\cdot)$; $BR_{\#}(\cdot)$ applies $BR_+(\cdot)$ to every argumentation theory in a set of argumentation theories; and $BR_3(\cdot)$ applies $BR_+(\cdot)$ until one element from each of the conflicts caused by axiom inconsistency has been removed. (At this point applying $BR_+(\cdot)$ will not cause the output to change and the fixpoint will have been reached.) After $BR_3(\cdot)$ has been applied to an axiom inconsistent theory AT_+ with n conflicts, the result will be a set of 2^n theories $\{AT_i\}$, such that $\mathcal{A}(AT_i)$ contains one strict argument from each of the n conflicts.

Given this definition of $BR_3(\cdot)$, we can relate the extensions of AT_D and AT_C . However, since AT_D is the defeasible version of AT_+ , their extensions (and by extension, the extensions of AT_C) do not involve the same arguments — the extensions of AT_D will involve the defeasible versions of the arguments in the extensions of AT_+ . As a result, to make comparisons, we need the notion of one extension E_D being the *defeasible version* of an extension E .

This idea follows from Definition 4.5 and associated definitions of the term “defeasible version” applying to theories and arguments:

Definition 5.8 (Defeasible Version of an Extension) *Given two extensions, E and E_D , E_D is the defeasible version of E , iff for every argument $A \in E$, there is an argument $A_D \in E_D$ such that A_D is the defeasible version of A , and there are no arguments $A_D \in E_D$ such that A_D is not the defeasible version of an argument $A \in E$.*

In other words E_D is the defeasible version of E if E_D is exactly the set of arguments that are the defeasible versions of the arguments in E . Then we have:

Lemma 5.7 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ , such that $\mathcal{A}(AT_+)$ includes one pair of conflicting strict and firm arguments, and its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ . Let AT_C and AT'_C be the result of applying BR_* to AT_+ , $BR_*(AT_+) = \{AT_C, AT'_C\}$. Then the preferred extensions of AT_D can be split into two disjoint sets E'_D and E''_D , where E'_D and E''_D are the defeasible versions of E' and E'' , and E' are the preferred extensions of AT_C and E'' are the preferred extensions of AT'_C .*

Proof *Call the two conflicting strict and firm arguments are A and B . With-*

out loss of generality, we assume there is no argument attacking any of the strict sub-arguments of A and B . From the proof of Proposition 5.5, we know that, in AT_D , the only defeater of A_D is B_D and the only defeater of B_D is A_D . Under the preferred semantics, either A_D is labeled **IN** and B is labelled **OUT**, or A_D is labelled **OUT** and B_D is labelled **IN**. Thus, in the absence of any other pairs of mutually defeating arguments, there will be exactly two preferred extensions, one including A_D and one including B_D . If there is another pair of mutually attacking arguments in AT_D which are not the result of conflicting strict and first arguments in AT_+ , call them C_D and D_D , then there will be four preferred extensions. Each extension will correspond to one of the labellings:

$$A_D : \text{IN}, B_D : \text{OUT}, C_D : \text{IN}, D_D : \text{OUT}$$

$$A_D : \text{IN}, B_D : \text{OUT}, C_D : \text{OUT}, D_D : \text{IN}$$

$$A_D : \text{OUT}, B_D : \text{IN}, C_D : \text{IN}, D_D : \text{OUT}$$

$$A_D : \text{OUT}, B_D : \text{IN}, C_D : \text{OUT}, D_D : \text{IN}$$

Given the way that $BR_*(\cdot)$ works, AT_C and AT'_C will each contain one of A and B . Without loss of generality, let us assume that AT_C contains A , and AT'_C contains B . The extensions of AT_C will then correspond to the

labellings:

$$A : \text{IN}, C : \text{IN}, D : \text{OUT}$$

$$A : \text{IN}, C_D : \text{OUT}, D_D : \text{IN}$$

while the extensions of AT'_C will correspond to the labellings:

$$B : \text{IN}, C : \text{IN}, D : \text{OUT}$$

$$B : \text{IN}, C_D : \text{OUT}, D_D : \text{IN}$$

Thus for this specific case, the preferred extensions of AT_D can be split into two disjoint sets E' and E'' , where E' is the preferred extensions of AT_C and E'' is the preferred extensions of AT'_C — and the result holds. Since we made no assumptions about the other arguments in AT_+ and AT_D , nor about the other extensions, we can see that whatever the extensions of AT_D , half will contain A_D and half B_D . Further, the half that include A_D will exactly correspond to the extensions of whichever of AT_C and AT'_C contains A , and the other half of the extensions will exactly correspond to the extensions of whichever of AT_C and AT'_C contains B . Since A and B were arbitrary names for the pair of strict and firm arguments in AT_+ , the result holds in the general case. \square

This deals with the case where AT_+ contains 2 conflicting strict and firm arguments only. This result easily generalises:

Proposition 5.12 *Consider an axiom inconsistent ASPIC⁺ theory AT_+ , and its defeasible version AT_D , where the preference ordering over AT_D is the strict-first version of the ordering over AT_+ . Let AT_c^1, \dots, AT_c^m be the result of applying BR_+ to AT_+ , $BR_3(AT_+) = \{AT_c^1, \dots, AT_c^m\}$. Then the preferred extensions of AT_D can be split into m disjoint sets E^1, \dots, E^m , where E^i are the defeasible versions of the preferred extensions of AT_c^i .*

Proof We know from Lemma 5.7 that if AT_+ includes a pair of conflicting arguments, then the extensions of AT_D can be split into two disjoint sets of extensions E^1 and E^2 , such that each E^i corresponds to the set of extensions of the two argumentation theories AT_c^i created by $BR_*(AT_+)$. Since $BR_3(\cdot)$ is the fixpoint combinator of $BR_*(\cdot)$, and since each application of $BR_*(\cdot)$ will further split the set of extensions of A_D into two disjoint sets, the result follows. \square

As noted above, if we start with n pairs of conflicting strict and firm arguments in $\mathcal{A}(AT_+)$, then the m in the statement of Proposition 5.12 will be 2^n .

Now, what $BR_3(\cdot)$ does, in effect, is to provide us with a number (2^n) of different choices about how to resolve the conflicts that arise from the inconsistent axioms. Making these choices picks a particular (sub)theory

AT_c^i , with an associated set of extensions. What Proposition 5.12 tells us is that that set of extensions corresponds exactly to a set of the preferred extensions of AT_D , and that these extensions are disjoint from those produced by any other set of choices about how to resolve the conflicts in AT_+ . In addition, should we want to recover the set of extensions of AT_D which correspond to a specific set of choices, that is easy to do. As the proof of Proposition 5.12 shows, if we want the extensions of AT_D that correspond to the extensions of the $AT^i c$ that we get when we prefer A to B , we just pick all the extensions of AT_D that include A not B ⁴.

5.3 Summary

Sections 5.1 and 5.2 considered $ASPIC_D^+$ versions of $ASPIC^+$ theories that are ill-defined, and answer the three research question raised in Chapter 3 in term of the ill-defined $ASPIC^+$ theory. Here we summarize the results of that investigation.

Section 5.1 looked at theories that are not closed under transposition. Given an $ASPIC^+$ theory AT_+ that is not closed under transposition, with transposed version AT_{tr} and defeasible version AT_D :

- Under the grounded semantics, we can identify which arguments in

⁴Of course, which arguments stem from the inconsistent axioms may not immediately be apparent.

$\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$ will have the same status, and so represent conclusions that hold whichever means we use to make the original theory well-defined. These arguments are identified by Proposition 5.2 and Corollary 5.1.

- The results in Proposition 5.2 and Corollary 5.1 work at the level of arguments — given an argument we can tell whether its status may be different between $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$. Proposition 5.3 extends this to identify which sub-theories AT'_+ of AT_+ will produce arguments with the same status in both the transposed version $\mathcal{A}(AT'_{tr})$ and the defeasible version $\mathcal{A}(AT'_D)$.
- Under all other semantics, there is no way to establish any general relationship between the members of $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$. The best we can do is to state, as Proposition 5.4 does, relationships that hold between arguments in at least one pair of extensions of AT_{tr} and AT_D .

As we pointed out above, these are not particularly strong results. Proposition 5.2 and Corollary 5.1 give us a way to vet individual arguments under the grounded semantics, but we have to construct them before we can tell whether they may have different status between $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$. Proposition 5.3 extends this to be able to predict which theories AT_+ will

generate arguments that will have the same status whether or not AT_+ is made well-defined by completing it with by transposing strict rules or by converting it into its defeasible counterpart.

Finally, once we step away from the grounded semantics, we can say very little in general about the relationship between the members of $\mathcal{A}(AT_{tr})$ and $\mathcal{A}(AT_D)$. Indeed, we are only able to give conditions, as in Proposition 5.4, under which arguments in at least one pair of extensions, one from $\mathcal{A}(AT_{tr})$ and one from $\mathcal{A}(AT_D)$, have the same status. However, given the results for defeasible theories under semantics other than the grounded semantics, this is not surprising. This answers the research question RQ2.

Of course, these results do not say that $ASPIC_D^+$ is bound to give different results from $ASPIC^+$ when handling theories that are ill-defined. What we have shown is that making an $ASPIC^+$ theory rational by translating it into $ASPIC_D^+$ may (where “may” is explained in detail in Section 5.1) produce different results to making an $ASPIC^+$ theory well-defined by adding all the transposed strict rules. Of course, if we take that well-defined $ASPIC^+$ theory and translate it into $ASPIC_D^+$, we are back in the scenario studied in Section 4.4, where under the grounded semantics both the $ASPIC^+$ and $ASPIC_D^+$ versions of the theory will have the same justified conclusions.

The results of Section 5.2 are stronger than those of Section 5.1. They

show that ASPIC_D^+ provides a natural way to deal with axiom inconsistency. Propositions 5.5 and 5.6 show that, under the grounded semantics, inconsistent axioms cannot lead to any justified conclusions. That is no inconsistent axiom can be part of an argument with a justified conclusion under the grounded semantics. Propositions 5.8 and 5.9 show the same for the sceptically preferred semantics. (Naturally, under the preferred semantics, inconsistent axioms lead to more preferred extensions.)

The results also show that the results obtained from ASPIC_D^+ theories align with some simple ways to revise the inconsistency, thus illuminating the way that ASPIC_D^+ handles inconsistency. If we have an axiom inconsistent ASPIC^+ theory AT_+ , with defeasible counterpart A_D , then we identified three simple ways to revise AT_C to create an axiom consistent theory AT_C . For one such revision, Proposition 5.7 shows us that the grounded extension of AT_C is the same as the grounded extension of AT_D . For a second revision, Proposition 5.10 shows us that the sceptically preferred extension of AT_C is the same as the sceptically preferred extension of AT_C . And for the third revision, Proposition 5.12 proves that the union of the preferred extensions of the AT_C are preferred extensions of AT_D . This answers the research questions RQ2 and RQ3.

Having investigated the theories that are both well-defined and ill-defined,

I will compare the ASPIC^+ and ASPIC_D^+ theories in the context of non-monotonic reasoning.

Chapter 6

Non-monotonic Properties

The previous chapters introduced the defeasible subset of ASPIC⁺ framework — the ASPIC_D⁺ framework, and investigated the similarity and difference between the two frameworks with respect to the justified conclusions. In this chapter, we continue considering the link between the two framework, but from another perspective — non-monotonic axioms. As we already know, both ASPIC⁺ and ASPIC_D⁺ are non-monotonic reasoning frameworks. By investigating the non-monotonic axioms, we can understand the exact nature of this non-monotonicity and get the underlying connection between the two frameworks.

In this chapter, I will investigate the links between instantiated argumentation systems and the axioms for non-monotonic reasoning described in [46] with the aim of characterizing the nature of argument based reasoning. It will answer the first and second research questions. In doing so, we consider

two possible interpretations of the consequence relation, and describe which axioms are satisfied by three different kinds of ASPIC⁺ theory ¹ under each of these interpretations. The three kinds of ASPIC⁺ theory that we will consider are: (1) ASPIC⁺ theories which contain only strict information, axioms and strict rules; (2) ASPIC⁺ theories that contain strict and defeasible information, axioms ordinary premises, strict rules and defeasible rules; and (3) ASPIC⁺ theories that contain only defeasible information, ordinary premises and defeasible rules. We will call these kinds of theory strict theories, regular theories and defeasible theories, respectively. At the end, we will compare the difference between an ASPIC⁺ theory and an ASPIC_D⁺ theory.

6.1 Axiomatic Reasoning and ASPIC⁺

Kraus *et al.* [46], building on earlier work by Gabbay [40], identified a set of axioms which characterise non-monotonic inference in logical systems, and studied the relationships between sets of these axioms. Their goal was to characterise different kinds of reasoning; to pin down what it means for a logical system to be monotonic or non-monotonic; and — in particular — to be able to distinguish between the two. Table 6.1 presents the axioms of

¹All theories in this chapter are well-defined.

Abbr.	Axiom	Name
Ref	$\alpha \sim \alpha$	Reflexivity
LLE	$\frac{\models \alpha \equiv \beta \quad \alpha \sim \gamma}{\beta \sim \gamma}$	Left Logical Equivalence
RW	$\frac{\models \alpha \leftrightarrow \beta \quad \gamma \sim \alpha}{\gamma \sim \beta}$	Right Weakening
Cut	$\frac{\alpha \wedge \beta \sim \gamma \quad \alpha \sim \beta}{\alpha \sim \gamma}$	Cut
CM	$\frac{\alpha \sim \beta \quad \alpha \sim \gamma}{\alpha \wedge \beta \sim \gamma}$	Cautious Monotonicity
M	$\frac{\models \alpha \leftrightarrow \beta \quad \beta \sim \gamma}{\alpha \sim \gamma}$	Monotonicity
T	$\frac{\alpha \sim \beta \quad \beta \sim \gamma}{\alpha \sim \gamma}$	Transitivity
CP	$\frac{\alpha \sim \beta}{\bar{\beta} \sim \bar{\alpha}}$	Contraposition

Table 6.1: The axioms from [46] that we will consider.

[46], which we will use to characterise reasoning in ASPIC⁺. The symbol \sim encodes a consequence relation, while \models identifies the statements obtainable from the underlying theory. We have altered some of the symbols used in [46] to avoid confusion with the notation of ASPIC⁺. Equivalence is denoted \equiv (rather than \leftrightarrow), and \leftrightarrow (rather than \rightarrow) denotes the existence of a strict or defeasible rule.

Consequence relations that satisfy Ref, LLE, RW, Cut and CM are said to be *cumulative*, and [46] describes them as being the weakest interesting

logical system. Cumulative consequence relations which also satisfy CP are *monotonic*, while consequence relations that are cumulative and satisfy M are called *cumulative monotonic*. Such relations are stronger than cumulative but not monotonic in the usual sense.

To determine which axioms ASPIC⁺ does or does not comply with, we must decide how different aspects of the axioms should be interpreted. We interpret the consequence relation \vdash in two ways that are natural in the context of ASPIC⁺—describing these in detail later— and which fit with the high level meaning of “if α is in the knowledge base, then β follows”, or “ β is a consequence of α ”.

Assuming such an interpretation of $\alpha \vdash \beta$ we can consider the meaning of the axioms. Some axioms are clear. For example, axiom T says that if β is a consequence of α , and γ is a consequence of β , then γ is a consequence of α . Other axioms are more ambiguous. Does $\alpha \wedge \beta \vdash \gamma$ in Cut mean that γ is a consequence of the conjunction $\alpha \wedge \beta$, or a consequence of α and β together? In other words is \wedge a feature of the language underlying the reasoning system, or a feature of the meta-language in which the properties are written? Similarly, given the distinction between strict and defeasible rules, is $\alpha \leftrightarrow \beta$ a strict rule in ASPIC⁺, a defeasible rule, or some statement in the property meta-language?

We interpret the symbols found in the axioms as follows:

- $\models \alpha$ means that α is an element of the relevant knowledge base.
- $\alpha \wedge \beta$ means both α and β , in particular in Cut and CM, \wedge means that both α and β are in the knowledge base.
- $\alpha \equiv \beta$ is taken — as usual — to abbreviate the formula $(\alpha \leftrightarrow \beta) \wedge (\beta \leftrightarrow \alpha)$. We assume $\alpha \leftrightarrow \beta$ and $\beta \leftrightarrow \alpha$ have the same interpretation, i.e., both or neither are strict.
- $\alpha \leftrightarrow \beta$ has two interpretations. We have the *strict* interpretation in which $\alpha \leftrightarrow \beta$ denotes a strict rule $\alpha \rightarrow \beta$ in ASPIC⁺, and the *defeasible* interpretation in which $\alpha \leftrightarrow \beta$ denotes either a strict or defeasible rule. We denote the latter interpretation by writing $\alpha \rightsquigarrow \beta$.

To evaluate ASPIC⁺, we have to be a bit more precise about exactly what we are evaluating. We start by saying that we assume an arbitrary ASPIC⁺ argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$, in the sense that we say nothing about the contents of the knowledge base, or what domain-specific rules it contains. However, we distinguish between two classes of theory, with respect to the *base logic* that the theory contains.

The idea we capture by this is that in addition to domain specific rules — rules, for example, about birds and penguins flying — an ASPIC⁺ theory

might also contain rules for reasoning in some logic. For example, we might equip an ASPIC⁺ theory with the axioms and inference rules of classical logic. Such a theory would be able to construct arguments using all the rules of classical logic, as well as all the domain-specific rules in the theory. The two base logics that we consider are classical logic, and what we call the “empty” base logic, where the ASPIC⁺ theory only contains domain-specific rules. (We make some observations about other base logics — intuitionistic logic and defeasible logic [16], but show no formal results for them.)

For each of the base logics, we consider the two different interpretations of the non-monotonic consequence relation $|\sim$ described above, identifying which axioms each interpretation satisfies. For our theory AT , we write AT_x to denote an extension of this augmentation theory also containing proposition x : $AT_x = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \cup \{x\} \rangle$. An argument present in the latter, but not former, theory is denoted A^x . In the following, I will compare the strength of several systems. By saying system A is stronger than system B , I mean that A satisfies more non-monotonic axioms than B .

6.2 Argument Construction

We begin by considering the consequence relation as representing argument construction. In other words, we interpret $\alpha |\sim \beta$ as meaning that if α is in

the axioms or ordinary premises of a theory, we can construct an argument for β . More precisely:

Definition 6.1 *We write $\alpha \vdash_{T,B,a} \beta$, if for every T ASPIC⁺ argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$ with base logic B such that $\beta \notin \text{Concs}(\mathcal{A}(AT))$, it is the case that $\beta \in \text{Concs}(\mathcal{A}(AT_\alpha))$, where $T = \{S, R, D\}$, representing strict, regular and defeasible ASPIC⁺ argumentation theories respectively; and $B = \{\emptyset, c\}$, representing empty and classical base logics respectively.*

Proposition 6.1 *Ref, LLE, RW, Cut and CM hold for $\vdash_{T,\emptyset,a}$ where $T = \{S, R, D\}$.*

Proof *Consider an arbitrary theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$.*

[Ref] *Given a theory AT_α , we have an argument $A^\alpha = [\alpha]$, so Ref holds for $\vdash_{T,\emptyset,a}$.*

[LLE] *Since $\alpha \vdash_{T,\emptyset,a} \gamma$, AT_α contains a chain of arguments $A_1^\alpha, A_2^\alpha, \dots, A_n^\alpha$ with $A_1^\alpha = [\alpha]$ and $\text{Conc}(A_n^\alpha) = \gamma$. Given $\models \alpha \equiv \beta$, we have that both $\alpha \rightsquigarrow \beta$ and $\beta \rightsquigarrow \alpha$ are in the theory AT , so are in the theory AT_β . Within AT_β , we obtain a chain of arguments $B_0^\beta = [\beta], B_1^\beta = [B_0^\beta \rightsquigarrow \alpha], A_2^\beta, \dots, A_n^\beta$. That is $\beta \vdash_{T,\emptyset,a} \gamma$. Therefore, both strict and defeasible versions of LLE hold for $\vdash_{T,\emptyset,a}$.*

[RW] *Since $\gamma \vdash_{T,\emptyset,a} \alpha$ in theory AT_γ , there is a chain of arguments*

$A_1^\gamma, A_2^\gamma, \dots, A_n^\gamma$ with $A_1^\gamma = [\gamma]$ and $\text{Conc}(A_n^\gamma) = \alpha$. Given $\models \alpha \leftrightarrow \beta$, theory AT must contain $\alpha \rightsquigarrow \beta$, so must in AT_γ . In AT_γ , we have a chain of arguments $A_1^\gamma, \dots, A_n^\gamma, A_{n+1}^\gamma = [A_n^\gamma \rightsquigarrow \beta]$. Thus, $\gamma \sim_{T, \emptyset, a} \beta$, and both strict and defeasible versions of RW hold for $\sim_{T, \emptyset, a}$.

[Cut] Since $\alpha \wedge \beta \sim_{T, \emptyset, a} \gamma$, there is a chain of arguments $A_1^{\alpha, \beta}, A_2^{\alpha, \beta}, \dots, A_n^{\alpha, \beta}$ with $A_1^{\alpha, \beta} = [\alpha]$, $A_2^{\alpha, \beta} = [\beta]$ in theory $AT_{\alpha, \beta}$, and $\text{Conc}(A_n^{\alpha, \beta}) = \gamma$. In theory AT_α , since $\alpha \sim_{T, \emptyset, a} \beta$, there is a chain of arguments $B_1^\alpha, B_2^\alpha, \dots, B_m^\alpha$ with $B_1^\alpha = [\alpha]$ and $\text{Conc}(B_m^\alpha) = \beta$. There is also a chain of arguments $B_1^\alpha, B_2^\alpha, \dots, B_m^\alpha, A_3^\alpha, \dots, A_n^\alpha$ in AT_α . That is $\alpha \sim_{T, \emptyset, a} \gamma$. Therefore, cut holds for $\sim_{T, \emptyset, a}$.

[CM] Since $\alpha \sim_{T, \emptyset, a} \gamma$, AT_α has a chain of arguments $A_1^\alpha, \dots, A_n^\alpha$ with $A_1^\alpha = [\alpha]$ and $\text{Conc}(A_n^\alpha) = \gamma$. $AT_{\alpha, \beta}$ has a similar chain of arguments $A_1^{\alpha, \beta}, \dots, A_n^{\alpha, \beta}$, so $\alpha \wedge \beta \sim_{T, \emptyset, a} \gamma$. CM thus holds for $\sim_{T, \emptyset, a}$. \square

Since Ref, LLE, RW, Cut and CM hold, $\sim_{T, \emptyset, a}$ is cumulative where $T = \{S, R, D\}$.

Proposition 6.2 *M and T hold for $\sim_{T, \emptyset, a}$ where $T = \{S, R, D\}$.*

Proof Consider an arbitrary theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$.

[M] Since $\beta \sim_{T, \emptyset, a} \gamma$, in the theory AT_β , there is a chain of arguments $A_1^\beta, A_2^\beta, \dots, A_n^\beta$ with $A_1^\beta = [\beta]$ and $\text{Conc}(A_n^\beta) = \gamma$. Given $\models \alpha \leftrightarrow \beta$, we

have $\alpha \rightsquigarrow \beta$ in the theory AT , and also in the theory AT_α . In the latter, there is a chain of arguments $B_0^\alpha = [\alpha], B_1^\alpha = [B_0^\alpha \rightsquigarrow \beta], A_2^\alpha, \dots, A_n^\alpha$. That is $\alpha \vdash_{T, \emptyset, a} \gamma$. Therefore, both strict and defeasible versions of M hold for $\vdash_{T, \emptyset, a}$.

[T] Since $\beta \vdash_{T, \emptyset, a} \gamma$, in AT_β , there is a chain of arguments $B_1^\beta, B_2^\beta, \dots, B_m^\beta$ with $B_1^\beta = [\beta]$ and $\text{Conc}(B_m^\beta) = \gamma$. Similarly, since $\alpha \vdash_{T, \emptyset, a} \beta$, in AT_α , there is a chain of arguments $A_1^\alpha, A_2^\alpha, \dots, A_n^\alpha$ with $A_1^\alpha = [\alpha]$ and $\text{Conc}(A_n^\alpha) = \beta$. Combining this with $B_1^\alpha, B_2^\alpha, \dots, B_m^\alpha$, we obtain the combined chain of arguments $A_1^\alpha, A_2^\alpha, \dots, A_n^\alpha, B_2^\alpha, \dots, B_m^\alpha$. That is $\alpha \vdash_{T, \emptyset, a} \gamma$. Therefore, T holds for $\vdash_{T, \emptyset, a}$. \square

Thus $\vdash_{T, \emptyset, a}$ is cumulative monotonic where $T = \{S, R, D\}$. It is not, however, monotonic.

Proposition 6.3 *CP does not hold for any $\vdash_{T, \emptyset, a}$ where $T = \{S, R\}$.*

Proof Consider an ASPIC⁺ theory which contains: $\mathcal{K} = \{c\}$, $\mathcal{R}_s = \{\alpha, c \rightarrow d; \alpha, \bar{d} \rightarrow \bar{c}; c, \bar{d} \rightarrow \bar{\alpha}; \alpha \rightarrow e; \bar{e} \rightarrow \bar{\alpha}; d, e \rightarrow \beta; d, \bar{\beta} \rightarrow \bar{e}; \bar{\beta}, e \rightarrow \bar{d}\}$. We have $\alpha \vdash_{T, \emptyset, a} \beta$ but not $\bar{\beta} \vdash_{T, \emptyset, a} \bar{\alpha}$. Therefore, CP does not hold for $\vdash_{T, \emptyset, a}$. \square

Proposition 6.4 *CP does not hold for any $\vdash_{D, \emptyset, a}$.*

Proof Consider the counter-example from Proposition 6.3 where all rules are

defeasible. It shows that CP does not hold for any $\sim_{D,\emptyset,a}$. \square

Having characterised $\sim_{T,\emptyset,a}$, we consider $\sim_{T,c,a}$. Clearly this will satisfy all the properties that are satisfied by $\sim_{T,\emptyset,a}$, since it includes all the inference rules of $\sim_{T,\emptyset,a}$. In addition, we have the following.

Proposition 6.5 *CP holds for $\sim_{S,c,a}$.*

Proof Any strict ASPIC⁺ theory with a classical base logic will generate the same set of consequences as classical logic. Furthermore, we know that CP is satisfied under classical logic. Therefore, the consequence relation $\sim_{S,c,a}$ satisfies CP. \square

Thus $\sim_{S,c,a}$ is monotonic. However:

Proposition 6.6 *CP does not hold for $\sim_{T,c,a}$ where $T = \{R, D\}$.*

Proof Consider the counter-example from Proposition 6.3 where all rules are defeasible. Since the defeasible portion of the theory does not contain a rule of the form $\bar{\beta} \rightarrow \bar{d} \vee \bar{e}$, CP will not be satisfied. \square

In this section, we have investigated the non-monotonic axioms under the argument construction interpretation. The results are shown in Table 6.2. Recall from Section 6.1 that a consequence relation which satisfies axioms

Table 6.2: Summary of axioms satisfied under the argumentation construction interpretation.

	$\sim_{S,\emptyset,a}$	$\sim_{R,\emptyset,a}$	$\sim_{D,\emptyset,a}$	$\sim_{S,c,a}$	$\sim_{R,c,a}$	$\sim_{D,c,a}$
Ref	Y	Y	Y	Y	Y	Y
LLE	Y	Y	Y	Y	Y	Y
RW	Y	Y	Y	Y	Y	Y
Cut	Y	Y	Y	Y	Y	Y
CM	Y	Y	Y	Y	Y	Y
M	Y	Y	Y	Y	Y	Y
T	Y	Y	Y	Y	Y	Y
CP	N	N	N	Y	N	N

Ref, LLE, RW, Cut and CM is said to be “cumulative”, a cumulative consequence relation that also satisfies M is said to be “cumulative monotonic”, and a consequence relation that satisfies CP is monotonic. From the table, we can see that all three kinds of ASPIC⁺ theory with empty base logic satisfy the same axioms, therefore, in term of inference under the argument construction interpretation, all three theories are of equal strength. As we can see from the table, all the three consequence relations $\sim_{S,\emptyset,a}$, $\sim_{R,\emptyset,a}$, $\sim_{D,\emptyset,a}$ are cumulative monotonic, but none of them are monotonic, this is because none of them satisfies CP. However, as we know, in any argumentation framework, the argument construction process is monotonic — the new constructed argument does not affect the existing arguments. In order to strengthen the three consequence relations, we added classical logic as base logic. This makes $\sim_{S,c,a}$ monotonic, and therefore as strong as classical

logic. However, neither $\sim_{R,c,a}$ nor $\sim_{D,c,a}$ is monotonic, since the defeasible inference rules are not closed under transposition.

6.3 Justified Conclusions

Next we interpret $\alpha \sim \beta$ as meaning that if α is in a theory, we can construct an argument for β such that β is in the set of justified conclusions. We will consider only the grounded and preferred semantics, but, as we will see, we have to bring in the ideas from Definition 2.30 since different kinds of justified conclusion lead to $\alpha \sim \beta$ satisfying different properties. We start with:

Definition 6.2 *Let $AF = \langle \mathcal{A}, Defeats \rangle$ be an abstract argumentation framework, we define*

$$\mathbf{Just}_g(\mathcal{A}(AT)) = \{\phi \mid \phi \text{ is a grounded justified conclusion}\}$$

$$\mathbf{Just}_p^c(\mathcal{A}(AT)) = \{\phi \mid \phi \text{ is a preferred credulously justified conclusion}\}$$

$$\mathbf{Just}_p^s(\mathcal{A}(AT)) = \{\phi \mid \phi \text{ is a preferred sceptically justified conclusion}\}$$

$$\mathbf{Just}_p^u(\mathcal{A}(AT)) = \{\phi \mid \phi \text{ is a preferred universally justified conclusion}\}$$

Note that we don't have to distinguish between different classes of grounded justified conclusion because, since there is exactly one grounded extension, the three different classes of grounded justified conclusion coincide. Then:

Definition 6.3 We write $\alpha \sim_{T,B,j}^g \beta$, if for every T ASPIC⁺ argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$ with the B base logic such that $\beta \notin \text{Just}_g(\mathcal{A}(AT))$, it is the case that $\beta \in \text{Just}_g(\mathcal{A}(AT_\alpha))$, where $T = \{S, R, D\}$ and $B = \{\emptyset, c\}$.

Definition 6.4 We write $\alpha \sim_{T,B,j}^{p, Sem} \beta$, if for every T ASPIC⁺ argumentation theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$ with the B base logic such that $\beta \notin \text{Just}_p^{Sem}(\mathcal{A}(AT))$, it is the case that $\beta \in \text{Just}_p^{Sem}(\mathcal{A}(AT_\alpha))$, where $T = \{S, R, D\}$, $B = \{\emptyset, c\}$ and $Sem = \{c, s, u\}$.

It is worth noting the following result.

Proposition 6.7 If $\alpha \sim_{T,B,j}^g \beta$ or $\alpha \sim_{T,B,j}^{p, Sem} \beta$ then $\alpha \sim_{T,B,a} \beta$ where $Sem = \{c, s, u\}$.

Proof Follows immediately from the definitions — for β to be a justified conclusion, there must first be an argument with β as a conclusion. \square

Since there are, in general, less justified conclusions of a theory than there are arguments, $\sim_{T,\emptyset,j}^g$ and $\sim_{T,\emptyset,j}^{p, Sem}$ are more restrictive notions of consequence than $\sim_{T,\emptyset,a}$. Similarly, $\sim_{T,c,j}^g$ and $\sim_{T,c,j}^{p, Sem}$ are more restrictive notions of consequence than $\sim_{T,c,a}$. It is therefore no surprise to find that fewer of the axioms from [46] hold.

6.3.1 Justified Conclusions in Strict ASPIC⁺ Theories

We start with strict ASPIC⁺ theories.

Proposition 6.8 *Ref, LLE, RW, Cut, CM, M and T hold for $\sim_{S,\emptyset,j}^g$ and $\sim_{S,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof *If the theory is strict, then for any argumentation theory, all conclusions are justified. Therefore, for any strict theory, if $\alpha \sim_{T,\emptyset,a} \beta$, then $\alpha \sim_{S,\emptyset,j}^g \beta$ and $\alpha \sim_{S,\emptyset,j}^{p,Sem} \beta$. We know that $\sim_{T,\emptyset,a}$ holds for Ref, LLE, RW, Cut, CM, M and T, therefore, $\sim_{S,\emptyset,j}^g$ and $\sim_{S,\emptyset,j}^{p,Sem}$ hold for Ref, LLE, RW, Cut, CM, M and T. \square*

Proposition 6.9 *CP does not hold for $\sim_{S,\emptyset,j}^g$ or $\sim_{S,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof *Since CP does not hold for $\sim_{S,\emptyset,a}$, CP can not hold for $\sim_{S,\emptyset,j}^g$ or $\sim_{S,\emptyset,j}^{p,Sem}$. \square*

This completes the characterisation of $\sim_{S,\emptyset,j}^g$ and $\sim_{S,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$. As we argued above, adding classical logic as a base logic will create consequence relations that satisfy the same properties since they will include all the same inference rules. In addition, we have the following:

Proposition 6.10 *CP holds for $\sim_{S,c,j}^g$ and $\sim_{S,c,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Table 6.3: Summary of axioms satisfied under the justified conclusion interpretation for strict theories. Here $Sem = \{c, s, u\}$.

	$\sim_{S,\emptyset,j}^g$	$\sim_{S,\emptyset,j}^{p,Sem}$	$\sim_{S,c,j}^g$	$\sim_{S,c,j}^{p,Sem}$
Ref	Y	Y	Y	Y
LLE	Y	Y	Y	Y
RW	Y	Y	Y	Y
Cut	Y	Y	Y	Y
CM	Y	Y	Y	Y
M	Y	Y	Y	Y
T	Y	Y	Y	Y
CP	N	N	Y	Y

Proof As above, $\sim_{S,c,a}$ satisfies CP. Since the strict part of the theory is always consistent, any conclusions from the argument construction are justified. Therefore, the consequence relation $\sim_{S,c,j}^g$ and $\sim_{S,c,j}^{p,Sem}$ satisfies CP.

□

In this section, we have investigated the non-monotonic axioms under the justified conclusions interpretation in strict ASPIC⁺ theories. The result is shown in Table 6.3. Since all the elements are strict, all the conclusions are justified. Therefore, it is not surprising that $\sim_{S,\emptyset,j}^g$ and $\sim_{S,\emptyset,j}^{p,Sem}$ have the exactly behaviors as $\sim_{S,\emptyset,a}$; $\sim_{S,c,j}^g$ and $\sim_{S,c,j}^{p,Sem}$ has the exactly behaviors as $\sim_{S,c,a}$ where $Sem = \{c, s, u\}$.

6.3.2 Justified Conclusions in Regular ASPIC⁺ Theories

Next we will consider regular ASPIC⁺ theories, namely, any ASPIC⁺ theory which contains both strict and defeasible elements. Furthermore, all the defeasible elements are equally preferred.

Proposition 6.11 *The premise version of Ref, and the defeasible versions of LLE and RW, do not hold for $\sim_{R,\emptyset,j}^g$, $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof [Ref (premise)] *Consider an ASPIC⁺ theory that contains: $\mathcal{K}_n = \{\bar{\alpha}\}$ and $\mathcal{R} = \emptyset$. Here, we have an argument $A = [\bar{\alpha}]$. If a is in the knowledge base \mathcal{K}_p , we have another argument $B = [a]$. However, B is defeated by A , but not vice versa. So B is not in any extension. Thus, Ref does not hold for either $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

[LLE (defeasible version)] *Consider an ASPIC⁺ theory that contains $\mathcal{K}_n = \{c\}$ and $\mathcal{R} = \{\alpha \Rightarrow \beta; \beta \Rightarrow \alpha; \alpha \Rightarrow \gamma; c \rightarrow \bar{n}_1\}$ where $n(\beta \Rightarrow \alpha) = n_1$. Here, $\alpha \sim_{R,\emptyset,j}^g \gamma$ and $\alpha \sim_{R,\emptyset,j}^{p,Sem} \gamma$, but, $\beta \not\sim_{R,\emptyset,j}^g \gamma$ and $\beta \not\sim_{R,\emptyset,j}^{p,Sem} \gamma$. Therefore, the defeasible version of LLE does not hold for either $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

[RW (defeasible version)] *Consider an ASPIC⁺ theory that contains $\bar{\beta}$ in its axioms. For such a theory, β will not appear in any justified conclu-*

sions. Therefore, the defeasible version of RW does not hold for either $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$. \square

Proposition 6.12 *The axiom version of Ref, and the strict version of LLE and RW hold for $\sim_{R,\emptyset,j}^g$ and $\sim_{R,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$.*

Proof Consider an arbitrary theory $AT = \langle \langle \mathcal{L}, \mathcal{R}, n \rangle, \mathcal{K} \rangle$.

[Ref (axiom)] Consider the ASPIC⁺ theory AT_α . Since α is an axiom in AT_α , $A = [\alpha]$ is always in any extension. Therefore, the axiom version of Ref holds for $\sim_{R,\emptyset,j}^g$ and $\sim_{R,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$.

[RW (strict version)] Consider the extension E_γ in AT_γ containing an argument A^γ with $\text{Conc}(A^\gamma) = \alpha$. Since $\models \alpha \rightsquigarrow \beta$, under the strict interpretation, we know that $\alpha \rightarrow \beta$ is in AT_γ . Therefore, we can construct an argument $B^\gamma = A^\gamma \rightarrow \beta$. Furthermore, the attackers of B are the attackers of A because $\text{TopRule}(B)$ is a strict rule. Since A^γ is in the extension E_γ , B^γ is in the same extension E_γ . Therefore the strict version of RW holds for $\sim_{R,\emptyset,j}^g$ and $\sim_{R,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$.

[LLE (strict version)] Since $\models \alpha \equiv \beta$, under the strict interpretation, the rules $\beta \rightarrow \alpha$ and $\alpha \rightarrow \beta$ are in AT, AT_α, AT_β and $AT_{\alpha,\beta}$. Thus $AT_\alpha, AT_\beta, AT_{\alpha,\beta}$ have the same extensions, just as for RW(strict version). If $\alpha \sim_{R,\emptyset,j}^g \gamma$, then $\beta \sim_{R,\emptyset,j}^g \gamma$. If $\alpha \sim_{R,\emptyset,j}^{p, Sem} \gamma$, then $\beta \sim_{R,\emptyset,j}^{p, Sem} \gamma$. Therefore, the

strict version of LLE holds for $\sim_{R,\emptyset,j}^g$ and $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$. \square

Proposition 6.13 *Cut holds for $\sim_{R,\emptyset,j}^g$ and $\sim_{R,\emptyset,j}^{p,s}$.*

Proof Since $\alpha \sim_{R,\emptyset,j}^g \beta$, the grounded justified conclusions of AT_α contain α and β . By adding β into the knowledge base, the grounded justified conclusions will not change – if the newly added β is not justified, then it has no effect; if the newly added β is justified, it will remain in the justified conclusions. The same argument applies for $\sim_{R,\emptyset,j}^{p,s}$. \square

Proposition 6.14 *Cut does not hold for $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$.*

Proof We will give a counter-example. Consider the ASPIC⁺ theory that include $\mathcal{K} = \emptyset$ and $\mathcal{R} = \{a \Rightarrow c; c \Rightarrow b; b \Rightarrow \bar{c}; \bar{c} \Rightarrow r\}$. The credulous or universal justified conclusions of AT_α are $\{a, b, c\}$. The credulous or universal justified conclusions of $AT_{\alpha,\beta}$ are $\{a, b, \bar{c}, r, c\}$. That is $a \wedge b \sim_{R,\emptyset,j}^{p,Sem} r$, $a \sim_{R,\emptyset,j}^{p,Sem} b$, but $a \not\sim_{R,\emptyset,j}^{p,Sem} r$ where $Sem = \{c, u\}$. Therefore Cut does not hold for $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$. \square

Proposition 6.15 *CM holds for $\sim_{R,\emptyset,j}^g$.*

Proof Since $\alpha \sim_{R,\emptyset,j}^g \gamma$, the grounded justified conclusions of AT_α contain α and γ . By adding β into the knowledge base, the grounded justified conclu-

sions will not change. The justification is same as in the proof of Proposition 6.13 \square

Proposition 6.16 *CM does not hold for $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof *We will give counter-examples. Consider an ASPIC⁺ theory that include $\mathcal{K} = \emptyset$ and $\mathcal{R} = \{a \Rightarrow b; a \Rightarrow r; b \Rightarrow \overline{n1}; r \Rightarrow \overline{n2}\}$, where $n(a \Rightarrow b) = n1$ and $n(a \Rightarrow r) = n2$. The credulous or universal justified conclusions of AT_α are $\{a, r, \overline{n1}, b, \overline{n2}\}$. And the credulous or universal justified conclusions of $AT_{\alpha,\beta}$ are $\{a, b, \overline{n2}\}$. That is $a \sim_{R,\emptyset,j}^{p,Sem} b$, $a \sim_{R,\emptyset,j}^{p,Sem} r$, but $a \wedge b \not\sim_{R,\emptyset,j}^{p,Sem} r$ where $Sem = \{c, u\}$. Therefore CM does not hold for $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$.*

Now, consider an ASPIC⁺ theory that include $\mathcal{K} = \emptyset$, $\mathcal{R} = \{a \Rightarrow r; r \Rightarrow b; b \Rightarrow \overline{r}\}$. The sceptical justified conclusions of AT_α are $\{a, b, r\}$. And the sceptical justified conclusions of $AT_{\alpha,\beta}$ are $\{a, b\}$. That is $a \sim_{R,\emptyset,j}^{p,s} b$, $a \sim_{R,\emptyset,j}^{p,s} r$, but $a \wedge b \not\sim_{R,\emptyset,j}^{p,s} r$. Therefore CM does not hold for $\sim_{R,\emptyset,j}^{p,s}$. \square

Proposition 6.17 *M, T and CP do not hold for $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof *We will give counter-examples.*

[M] *Consider an ASPIC⁺ theory which includes $\mathcal{K}_n = \{\overline{\alpha}\}$ and $\mathcal{R} = \{\alpha \rightarrow \beta; \overline{\beta} \rightarrow \overline{\alpha}; \beta \Rightarrow \gamma\}$. Thus, $\beta \sim_{R,\emptyset,j}^g \gamma$, but $\alpha \not\sim_{R,\emptyset,j}^g \gamma$; $\beta \sim_{R,\emptyset,j}^{p,Sem} \gamma$, but*

Table 6.4: Summary of axioms satisfied under the justified conclusion interpretation for regular theories. Here $Sem = \{c, u\}$.

		$\sim_{R,\emptyset,j}^g$	$\sim_{R,\emptyset,j}^{p,s}$	$\sim_{R,\emptyset,j}^{p,Sem}$	$\sim_{R,c,j}^g$	$\sim_{R,c,j}^{p,s}$	$\sim_{R,c,j}^{p,Sem}$
Ref	premise	N	N	N	N	N	N
	axiom	Y	Y	Y	Y	Y	Y
LLE	defeasible rule	N	N	N	N	N	N
	strict rule	Y	Y	Y	Y	Y	Y
RW	defeasible rule	N	N	N	N	N	N
	strict rule	Y	Y	Y	Y	Y	Y
Cut		Y	Y	N	Y	Y	N
CM		Y	N	N	Y	N	N
M		N	N	N	N	N	N
T		N	N	N	N	N	N
CP		N	N	N	N	N	N

$\alpha \not\sim_{R,\emptyset,j}^{p,Sem} \gamma$ where $Sem = \{c, s, u\}$. Therefore, M does not hold for $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.

[T] Consider an ASPIC⁺ theory which includes $\mathcal{K} = \emptyset$ and $\mathcal{R} = \{\alpha \Rightarrow \beta; \beta \Rightarrow c; c \Rightarrow \gamma; \alpha \Rightarrow \overline{n_1}\}$ where $n(c \Rightarrow \gamma) = n_1$. Thus, $a \sim_{R,\emptyset,j}^g b$, $b \sim_{R,\emptyset,j}^g r$, but $a \not\sim_{R,\emptyset,j}^g r$; $a \sim_{R,\emptyset,j}^{p,Sem} b$, $b \sim_{R,\emptyset,j}^{p,Sem} r$, but $a \not\sim_{R,\emptyset,j}^{p,Sem} r$ where $Sem = \{c, s, u\}$. Therefore, T does not hold for $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.

[CP] Since contraposition does not hold for $\sim_{\emptyset,a}$, by Proposition 6.3 it cannot hold for $\sim_{R,\emptyset,j}^g$ or $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$. \square

For regular theories, $\sim_{R,c,j}^g$ behaves exactly same as $\sim_{R,\emptyset,j}^g$; and $\sim_{R,c,j}^{p,Sem}$ behaves exactly same as $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.

In this section, we have investigated the non-monotonic axioms under

the justified conclusion interpretation in regular ASPIC⁺ theories. The results are shown in Table 6.4. $\sim_{R,\emptyset,j}$ is perhaps a more reasonable notions of consequence for ASPIC⁺ than $\sim_{T,\emptyset,a}$ or $\sim_{S,\emptyset,j}$, since it contains the inference process using the full argumentation framework. $\sim_{R,\emptyset,j}$ is quite a restrictive notion of consequence in a representation that allows for conflicting information, because in the strict only theories, there is no conflicting information. As Table 6.4 makes clear, even $\sim_{R,\emptyset,j}^g$, which is the strongest of the consequence relations based on justified conclusions, is a relatively weak notion of consequence in the sense that it obeys less of the axioms than the non-monotonic logics analyzed in [46], for example.

$\sim_{R,\emptyset,j}^g$ is not cumulative, and only satisfies LLE and RW if the rules applied in those axioms are strict. As we pointed out above, at the time that [46] was published, cumulativity was considered the minimum requirement of a useful logic². Whether or not one accepts this, it is clear that $\sim_{R,\emptyset,j}^g$ is weaker in the sense that it obeys less of the axioms. But is it too weak? To answer this, we should consider reason why $\sim_{R,\emptyset,j}^g$ is not cumulative, which as Table 6.4 shows is due to LLE, RW and Ref.

LLE and RW only hold in the case of strict rules. For both LLE and RW,

²This position was doubtless a side-effect of the fact that at that time there were no logics that did not obey cumulativity. The subsequent discovery of logics of causality that are not cumulative suggests that this view should be revised.

the effect of the axiom is to extend an existing argument, either switching one premise for another (LLE), or adding a rule to the conclusion of an argument (RW). While having these axioms hold for defeasible rules would allow $\sim_{R,\emptyset,j}^g$ to be cumulative, this is not reasonable. Using LLE or RW to extend arguments with defeasible rules —by definition— means that the new arguments created by this extension can be defeated. Thus their conclusions may not be justified, and $\sim_{R,\emptyset,j}^g$ must not be cumulative for defeasible rules. In other words $\sim_{R,\emptyset,j}^g$ is not cumulative for defeasible rules exactly because it makes no sense for a system of defeasible rules to be cumulative.

A similar argument applies to Ref. If Ref were to hold for $\sim_{R,\emptyset,j}^g$, α could be a premise. But premises can be defeated, again by definition, so it is not appropriate to directly conclude that any premise is a justified conclusion (it is necessary to go through the whole process of constructing arguments and establishing extensions to determine this).

Turning to the preferred semantics, since there may be more than one preferred extension, we consider three different preferred justified conclusions — sceptical, credulous and universal. Among them, credulous and universal justified conclusion satisfy the exactly same axioms, therefore we will consider them as a whole. From the Table 6.4, we can see that $\sim_{R,\emptyset,j}^{p,s}$ is strictly weaker than $\sim_{R,\emptyset,j}^g$, due to CM. And $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$ is strictly weaker

than $\sim_{R,\emptyset,j}^g$ due to CM and Cut.

CM expresses the fact that learning a new proposition, which has been previously justified, should not invalidate previous conclusions. This is true for grounded semantics, however, it is not true for sceptical preferred semantics. The reason is that ASPIC⁺ allows self-attacking arguments. Under the grounded semantics, self-attacking arguments are all labeled UNDEC, therefore any arguments attacked by self-attacking argument are not justified. However, this is not true under the preferred semantics, since the preferred semantics maximizes the number of justified conclusions, which may make the arguments attacked by self-attacking argument justified³.

Next we will consider why the credulous/universal justified conclusions interpretation is weaker than the grounded interpretation. This is because adding a new proposition will introduce more arguments, and these arguments may cause more attacks, and therefore lead to a smaller number of preferred extensions. This is why CM fails under the credulous/universal justified conclusions interpretation — the newly added proposition may decrease the number of preferred extensions⁴. Therefore, less justified conclusions will

³If the self-attacking argument A has an IN attacker, then A will be labeled OUT. Then the argument B attacked by A can be labeled IN if B has no other IN attackers.

⁴As shown in the counter-example in the proof of Proposition 6.16, the number of preferred extension change from two to one by adding a new proposition which is previously justified.

be obtained.

What is the difference between the sceptical justified conclusion interpretation and the credulous/universal justified conclusion interpretation? Again, from the Table 6.4, we can see that $\sim_{R,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$ is strictly weaker than $\sim_{R,\emptyset,j}^{p,s}$ due to Cut. Cut expresses the reasoning that is the opposite of CM, where removing a proposition, which has been previously justified should not invalidate previous conclusions. Again, the reason why Cut does not hold under the credulous/universal justified conclusion interpretation is that removing a proposition may cause the number of preferred extension to increase. Therefore, less justified conclusion will be obtained.

Overall, despite the fact that regular ASPIC⁺ is weaker in term of the number of non-monotonic axioms that it conforms to than the minimum requirements for a reasonable non-monotonic system proposed by [46], it is reasonable to have these behaviors in terms of an argumentation framework as discussed above. Again, this weakness raises the question of whether reasoning in ASPIC⁺ can be strengthened. Adding a classical logic as base logic does not help in strengthening conclusions — we gain nothing from adding the classical base logic, as the results above show.

6.3.3 Justified Conclusions in ASPIC_D^+ Theories

Finally, we consider defeasible only ASPIC^+ theories — that are the theories of ASPIC_D^+ which is introduced in previous chapters. We interpret all the rules in the non-monotonic axioms as defeasible rules. But they can not be undercut⁵. Again, as in regular theories, all the elements are equally preferred.

Proposition 6.18 *Ref, LLE, RW holds for $\sim_{D,\emptyset,j}^{p, Sem}$ where $Sem = \{c, u\}$.*

Proof [Ref] *Consider any ASPIC_D^+ theory AT , there is an argument $A_\alpha = [\alpha]$ in AT_α . And the only possible attack of A_α is rebut. Since AT_α is purely defeasible, the rebut relation is symmetric. Therefore, A_α is in at least one preferred extension. Thus, $\alpha \sim_{D,\emptyset,j}^{p, Sem} \alpha$ in AT_α where $Sem = \{c, u\}$. Hence, Ref holds for $\sim_{D,\emptyset,j}^{p, Sem}$ where $Sem = \{c, u\}$.*

[LLE] *Consider any ASPIC_D^+ theory AT . Since $\beta \Rightarrow \alpha$ is a defeasible rule in AT , we can construct arguments $A_\beta = [\beta]$ and $B_\beta = [A \Rightarrow \alpha]$ in AT_β . Since $\alpha \sim_{D,\emptyset,j}^{p, Sem} \gamma$, there is a chain of arguments $C_\alpha^1 = [\alpha], \dots, C_\alpha^n$, where $\text{Conc}(C_\alpha^n) = \gamma$ in a preferred extension E_α . Furthermore, there is an argument $D_\alpha = [C_\alpha^1 \Rightarrow \beta]$ in AT_α . Now, consider the chain of arguments $A_\beta, B_\beta, C_\beta^2, \dots, C_\beta^m$. It is possible that there is a direct attack to A_β or there is*

⁵Otherwise, we can have a rule that undercut all the inference rules in the non-monotonic axioms that makes them not hold

a direct attack to $B, C_\beta^2, \dots, C_\beta^n$. The former is symmetric attacking, and the latter is defended in E_α which contains argument with conclusion β . Therefore, $A_\beta, B_\beta, C_\beta^2, \dots, C_\beta^n$ is in at least one preferred extension in AT_β . Thus, LLE holds for $\sim_{D, \emptyset, j}^{p, Sem}$ where $Sem = \{c, u\}$.

[RW] Consider any purely defeasible ASPIC⁺ theory AT . Since $\gamma \sim_{D, \emptyset, j}^{p, Sem} \alpha$ there is an argument A_γ with $\text{Conc}(A_\gamma) = \alpha$ in at least one preferred extension E_γ in AT_γ . We know that there is a defeasible rule $\alpha \Rightarrow \beta$ which is not undercut by any arguments. So can construct argument $B_\gamma = [A_\gamma \Rightarrow \beta]$. Furthermore, any argument attacking B either attacks A or directly rebuts B . The former one is defended in E_γ , the latter one is a symmetric attack. Therefore, B is in at least one preferred extension in AT_γ . Thus, RW holds for $\sim_{D, \emptyset, j}^{p, Sem}$ where $Sem = \{c, u\}$. \square

Proposition 6.19 *Ref, LLE, RW do not hold for $\sim_{D, \emptyset, j}^g, \sim_{D, \emptyset, j}^{p, s}$.*

Proof We will give counter-examples.

[Ref] Consider an ASPIC⁺ theory that includes $\mathcal{K}_p = \{\bar{\alpha}\}$. The grounded extension of AT_α is \emptyset . Thus $\alpha \not\sim_{D, \emptyset, j}^g \alpha$. The preferred extension of AT_α is $\{\{\alpha\}, \{\bar{\alpha}\}\}$. Thus $\alpha \not\sim_{D, \emptyset, j}^{p, s} \alpha$. Therefore, Ref does not hold for either $\sim_{D, \emptyset, j}^g$ or $\sim_{D, \emptyset, j}^{p, s}$. \square

[LLE] Consider an ASPIC⁺ theory that includes $\mathcal{K}_p = \{\bar{\beta}\}$ and $\mathcal{R} =$

$\{\alpha \Rightarrow \beta; \beta \Rightarrow \alpha; \alpha \Rightarrow \gamma\}$. Here, $\alpha \sim_{D,\emptyset,j}^g \gamma$ and $\alpha \sim_{D,\emptyset,j}^{p,s} \gamma$, but, $\beta \not\sim_{D,\emptyset,j}^g \gamma$ and $\beta \not\sim_{D,\emptyset,j}^{p,s} \gamma$. Therefore, LLE does not hold for either $\sim_{D,\emptyset,j}^g$ or $\sim_{D,\emptyset,j}^{p,s}$.

[RW] Consider an ASPIC⁺ theory that includes $\mathcal{K} = \{\bar{\beta}\}$ and $\mathcal{R} = \{\alpha \Rightarrow \beta; \gamma \Rightarrow \alpha\}$. Here $\gamma \sim_{D,\emptyset,j}^g \alpha$ and $\gamma \sim_{D,\emptyset,j}^{p,s} \alpha$ but $\gamma \not\sim_{D,\emptyset,j}^g \beta$ and $\gamma \not\sim_{D,\emptyset,j}^{p,s} \beta$. Therefore, RW does not hold for either $\sim_{D,\emptyset,j}^g$ or $\sim_{D,\emptyset,j}^{p,s}$ in purely defeasible theory.

Proposition 6.20 *Cut holds for $\sim_{D,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^{p,s}$.*

Proof The proof in Proposition 6.13 hold for ASPIC_D⁺ theory. □

Proposition 6.21 *Cut does not hold for $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$.*

Proof The counter-example is same as the counter-example given in the proof of Proposition 6.14. □

Proposition 6.22 *CM holds for $\sim_{D,\emptyset,j}^g$.*

Proof The proof in Proposition 6.15 hold for ASPIC_D⁺ theory. □

Proposition 6.23 *CM does not hold for $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof The counter-example is same as the counter-example given in the proof of Proposition 6.16. □

Proposition 6.24 *M, T and CP do not hold for $\sim_{D,\emptyset,j}^g$ or $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

Proof *We will give counter-examples.*

[M] *Consider an ASPIC⁺ theory that contains $\mathcal{K}_p = \emptyset$ and $\mathcal{R} = \{\alpha \Rightarrow \beta; \beta \Rightarrow c; c \Rightarrow \gamma; \alpha \Rightarrow \overline{n1}\}$ where $n(\beta \Rightarrow c) = n1$. Thus, $\beta \sim_{D,\emptyset,j}^g \gamma$ and $\beta \sim_{D,\emptyset,j}^{p,Sem} \gamma$, however, $\alpha \not\sim_{D,\emptyset,j}^g \gamma$ and $\alpha \not\sim_{D,\emptyset,j}^{p,Sem} \gamma$. Therefore, M does not hold for $\sim_{D,\emptyset,j}^g$ or $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

[T] *Consider an ASPIC⁺ theory the same as above. We have $a \sim_{D,\emptyset,j}^g b$, $b \sim_{D,\emptyset,j}^g r$ but $a \not\sim_{D,\emptyset,j}^g r$; $a \sim_{D,\emptyset,j}^{p,Sem} b$, $b \sim_{D,\emptyset,j}^{p,Sem} r$, but $a \not\sim_{D,\emptyset,j}^{p,Sem} r$ where $Sem = \{c, s, u\}$. Therefore, T does not hold for $\sim_{D,\emptyset,j}^g$ or $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$.*

[CP] *Since contraposition does not hold for $\sim_{T,\emptyset,a}$, by Proposition 6.3 it cannot hold for $\sim_{D,\emptyset,j}^g$ or $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$. \square*

In the above interpretation, the non-monotonic axioms are interpreted as referring to defeasible rules, which corresponds to the defeasible interpretation in regular ASPIC⁺ theories. However, we also investigate the strict interpretation in regular ASPIC⁺ theories, in order to compare the connection with regular ASPIC⁺ theories. To do this, we make the inference rules in the non-monotonic axioms be “strict”, i.e., there is no argument directly attack-

ing the relevant argument, the one for which `TopRule` is the rules referred to in the non-monotonic axioms. For example, in LLE, we interpret $\models \alpha \equiv \beta$ as $\alpha \Rightarrow \beta$ and $\beta \Rightarrow \alpha$, and any argument A with $\text{TopRule}(A) = \alpha \Rightarrow \beta$ or $\text{TopRule}(A) = \beta \Rightarrow \alpha$ can not be directly attacked.⁶ This interpretation corresponds to the strict interpretation in regular ASPIC⁺ theories. We call this the **strict** version interpretation.

Proposition 6.25 *The strict version of Ref, LLE and RW hold for $\sim_{D,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$.*

Proof [Ref] *Consider any ASPIC_D⁺ theory AT, there is an argument $A_\alpha = [\alpha]$ in AT_α . The only possible attack of A_α is rebut. With strict interpretation, A_α can not be defeated. So $\alpha \sim_{D,\emptyset,j}^g \alpha$ and $\alpha \sim_{D,\emptyset,j}^{p, Sem} \alpha$ where $Sem = \{c, s, u\}$. Therefore, Ref holds for $\sim_{D,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$.*

[LLE] *Consider any ASPIC_D⁺ theory AT. With strict interpretation, $AT_\alpha, AT_{\alpha,\beta}, AT_\beta$ have the same extensions. If $\alpha \sim_{D,\emptyset,j}^{p, Sem} \gamma$, then $\beta \sim_{D,\emptyset,j}^{p, Sem} \gamma$ where $Sem = \{c, s, u\}$. If $\alpha \sim_{D,\emptyset,j}^g \gamma$, then $\beta \sim_{D,\emptyset,j}^g \gamma$. Therefore, LLE holds for $\sim_{D,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^{p, Sem}$ where $Sem = \{c, s, u\}$.*

[RW] *Consider any ASPIC_D⁺ theory AT. Since $\gamma \sim_{D,\emptyset,j}^{p, Sem} \alpha$ there is an argument A_γ with $\text{Conc}(A_\gamma) = \alpha$ in every preferred extension E_γ in AT_γ . We*

⁶But you can still attack the strict sub-argument of A .

Table 6.5: Summary of axioms satisfied under the justified conclusion interpretation for ASPIC_D^+ theories. Here $Sem = \{c, u\}$.

		$\sim_{D,\emptyset,j}^g$	$\sim_{D,\emptyset,j}^{p,s}$	$\sim_{D,\emptyset,j}^{p,Sem}$	$\sim_{D,c,j}^g$	$\sim_{D,c,j}^{p,s}$	$\sim_{D,c,j}^{p,Sem}$
Ref	premise	N	N	Y	N	N	Y
	axiom	Y	Y	Y	Y	Y	Y
LLE	defeasible	N	N	Y	N	N	Y
	strict	Y	Y	Y	Y	Y	Y
RW	defeasible	N	N	Y	N	N	Y
	strict	Y	Y	Y	Y	Y	Y
Cut		Y	Y	N	Y	Y	N
CM		Y	N	N	Y	N	N
M		N	N	N	N	N	N
T		N	N	N	N	N	N
CP		N	N	N	N	N	N

know that there is a defeasible rule $\alpha \Rightarrow \beta$ which is a strict interpretation. So we can construct argument $B_\gamma = [A_\gamma \Rightarrow \beta]$. Furthermore, any argument attacking B either attacks A or directly rebuts B . The former one is defended in E_γ , the latter is not successful due to the strict interpretation. Therefore, B is in every preferred extension in AT_γ . Thus, RW holds for $\sim_{D,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, s, u\}$. \square

In this section, we have investigated the non-monotonic axioms under the justified conclusion interpretation in ASPIC_D^+ theories. The results are shown in Table 6.5. In ASPIC_D^+ theories, all the inference rules are defeasible, and therefore can be undercut. It is reasonable to set some constraints that the rules in the non-monotonic axioms can not be undercut. In order to

match the defeasible/strict version interpretation in regular ASPIC⁺ theories, we consider two interpretations. For the strict version interpretation, we add the constraint that the argument whose `TopRule` is the rules referred to in the non-monotonic axioms can not be directly attacked. Since ASPIC_D⁺ is a subset of ASPIC⁺, and ASPIC_D⁺ has less constraints than ASPIC⁺, such as not requiring restricted rebut and that theories are not closed under transposition, we see that more axioms hold in ASPIC_D⁺ theories than regular ASPIC⁺ theories. The detail of differences will be discussed in the following section.

6.4 Discussion

Table 6.6: Summary of axioms satisfied under justified conclusion for strict ASPIC⁺, regular ASPIC⁺ and ASPIC_D⁺ theories with empty base logic. Here $Sem = \{c, u\}$.

		$\sim_{S,\emptyset,j}^g$	$\sim_{R,\emptyset,j}^g$	$\sim_{D,\emptyset,j}^g$	$\sim_{S,\emptyset,j}^{p,s}$	$\sim_{R,\emptyset,j}^{p,s}$	$\sim_{D,\emptyset,j}^{p,s}$	$\sim_{S,\emptyset,j}^{p,Sem}$	$\sim_{R,\emptyset,j}^{p,Sem}$	$\sim_{D,\emptyset,j}^{p,Sem}$
Ref	premise		N	N		N	N		N	Y
	axiom	Y	Y	Y	Y	Y	Y	Y	Y	Y
LLE	defeasible		N	N		N	N		N	Y
	strict	Y	Y	Y	Y	Y	Y	Y	Y	Y
RW	defeasible		N	N		N	N		N	Y
	strict	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cut		Y	Y	Y	Y	Y	Y	Y	N	N
CM		Y	Y	Y	Y	N	N	Y	N	N
M		Y	N	N	Y	N	N	Y	N	N
T		Y	N	N	Y	N	N	Y	N	N
CP		N	N	N	N	N	N	N	N	N

Table 6.7: Summary of axioms satisfied under justified conclusion for strict ASPIC⁺, regular ASPIC⁺ and ASPIC_D⁺ theories with classical base logic. Here $Sem = \{c, u\}$.

		$\sim^g_{S,c,j}$	$\sim^g_{R,c,j}$	$\sim^g_{D,c,j}$	$\sim^{p,s}_{S,c,j}$	$\sim^{p,s}_{R,c,j}$	$\sim^{p,s}_{D,c,j}$	$\sim^{p,Sem}_{S,c,j}$	$\sim^{p,Sem}_{R,c,j}$	$\sim^{p,Sem}_{D,c,j}$
Ref	premise		N	N		N	N		N	Y
	axiom	Y	Y	Y	Y	Y	Y	Y	Y	Y
LLE	defeasible		N	N		N	N		N	Y
	strict	Y	Y	Y	Y	Y	Y	Y	Y	Y
RW	defeasible		N	N		N	N		N	Y
	strict	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cut		Y	Y	Y	Y	Y	Y	Y	N	N
CM		Y	Y	Y	Y	N	N	Y	N	N
M		Y	N	N	Y	N	N	Y	N	N
T		Y	N	N	Y	N	N	Y	N	N
CP		Y	N	N	Y	N	N	Y	N	N

What are the difference between ASPIC⁺ and ASPIC_D⁺ argumentation framework with respect to non-monotonic axioms in general? We will answer this question by considering each of the consequence relations in turn.

Starting with the consequence relations where inference corresponds to argument construction, it is no surprise that the three consequence relations, $\vdash_{S,\emptyset,a}$, $\vdash_{R,\emptyset,a}$ and $\vdash_{D,\emptyset,a}$, have the same strength. They all cumulative monotonic and satisfy the axiom M which captures a form of monotonicity. It is clear from the detail of ASPIC⁺, and indeed any argumentation system, that the number of arguments grows over time, and that once introduced, arguments do not disappear. However, the fact that $\vdash_{S,\emptyset,a}$, $\vdash_{R,\emptyset,a}$ and $\vdash_{D,\emptyset,a}$ are not monotonic in the same strict sense as classical logic, and so are strictly weaker, as a result of not satisfying CP, is a bit more interesting. This is, of course, because arguments are not subject to the law of the excluded middle — it is perfectly possible for neither α nor $\bar{\alpha}$ to be the consequence of a given theory.

When we add classical logic as base logic, $\vdash_{S,c,a}$ satisfies CP and therefore becomes monotonic. This is because we added the law of the excluded middle into the theory, and therefore the theory is the same strength as classical logic. However, $\vdash_{R,c,a}$ and $\vdash_{D,c,a}$ are still not monotonic even if we add the law of the excluded middle. The reason is that the inference rules in regular

ASPIC⁺ theories and ASPIC_D⁺ theories are not closed under transposition. Overall, with respect to argument construction, regular ASPIC⁺ theories and ASPIC_D⁺ theories behave exactly the same.

Turning to the various versions of consequence relations built around justified conclusions, we start by noting that they are perhaps more reasonable notions of consequence relations than argument construction. If β is a justified conclusion of α , then there is an argument for β which holds despite any attacks (in the scenario we have considered, where all attacks may be defeats for some preference ordering — and therefore succeed — there can still be attacks on the argument for β , but the attacking arguments must themselves be defeated). Table 6.6 and Table 6.7 compares different consequence relations under three different theories, strict ASPIC⁺ regular ASPIC⁺ and ASPIC_D⁺.

Again, we are starting with the grounded semantics. Recall Chapter 4, we have shown that ASPIC⁺ and ASPIC_D⁺ theories have the same expressiveness under grounded semantics. So it is quite reasonable that $\sim_{R,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^g$ behave exactly the same. Neither of them is cumulative in general, however, for the strict interpretation, both $\sim_{R,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^g$ are cumulative. In comparison with $\sim_{S,\emptyset,j}^g$, even with strict interpretation of non-monotonic axioms, $\sim_{D,\emptyset,j}^g$ is strict weaker than $\sim_{S,\emptyset,j}^g$. From the Table 6.6, we know

that the weakness is coming from the axioms M and T, which capture a form of monotonicity. For strict ASPIC⁺ theories, anything constructed will be justified, which is somehow monotonic. However, for ASPIC_D⁺ theories, everything is defeasible, we can not guarantee that newly added arguments are justified.

Now we turn to the preferred semantics. Recall Chapter 4, where we have shown that any preferred extension of an ASPIC⁺ theory is a preferred extension of the corresponding ASPIC_D⁺ theory, however, the ASPIC_D⁺ theory may contain more preferred extensions than the corresponding ASPIC⁺ theory. This is caused by unrestricted rebut. Because of unrestricted rebut, ASPIC_D⁺ satisfies more axioms than ASPIC⁺. That is the defeasible version of Ref, LLE and RW with respect to credulous and universal preferred semantics, which is shown in Table 6.6 and Table 6.7.

Let's see why the unrestricted rebut makes differences. By definition, Ref means when we add something into the theory, it is justified. In ASPIC⁺ theory, the new added premise can be attacked by an axiom, but not vice versa. However, in ASPIC_D⁺ theory, because of unrestricted rebut, the new added premise can attack a justified argument, causing the justified argument to be unjustified in some preferred extensions. Therefore, Ref holds for credulous and universal preferred semantics.

Similarly, for both LLE and RW, the effect of the axiom is to extend an existing argument. Again, for restricted rebut, the extended argument can only be attacked by a strict argument, but not vice versa. However, under unrestricted rebut, the extended arguments can attack a justified argument causing it to become unjustified. This makes the extended arguments justified in some preferred extensions. Therefore, both defeasible LLE and RW hold for $\sim_{D,\emptyset,j}^{p,Sem}$ where $Sem = \{c, u\}$.

Overall, both $ASPIC^+$ and $ASPIC_D^+$ are weaker than the systems introduced in [46]. This weakness raises the question of whether reasoning in $ASPIC^+$ or $ASPIC_D^+$ can be strengthened. When we add classical logic as a base logic, the results are shown in Table 6.7, we get a family of consequence relations that satisfy CP. Thus $\sim_{S,c,j}^g$ and $\sim_{S,c,j}^{p,Sem}$ are monotonic where $Sem = \{c, s, u\}$. However, for theories with defeasible elements, even the strongest consequence relations $\sim_{R,c,j}^g$ and $\sim_{D,c,j}^g$ can not guarantee that CP will hold for arbitrary α and β . Adding a base logic that is weaker than classical logic does not help in strengthening conclusions. If we add intuitionistic logic, for example, we don't get CP, because intuitionistic logic explicitly rejects this pattern of reasoning.

From this we conclude that under the grounded semantics, $ASPIC^+$ and $ASPIC_D^+$ have the same strength in terms of which axioms from [46] they sat-

isfy. However, they are both weaker than non-monotonic logics like circumscription [55] and default logic [73] which, according to [46], is cumulative. Under the preferred semantics, ASPIC_D^+ is slightly stronger than ASPIC^+ , but again, they are both weaker than the above systems.

6.5 Summary

In this chapter, we have examined the non-monotonic properties proposed by [46] with respect to the ASPIC^+ argumentation framework as well as the ASPIC_D^+ argumentation framework. We considered which of the axioms ASPIC^+ and ASPIC_D^+ satisfy based on two different interpretations of the consequence relation — argumentation construction and justified conclusions.

Table 6.2 shows that $\sim_{R,\emptyset,a}$ and $\sim_{D,\emptyset,a}$ satisfy the same non-monotonic axioms; $\sim_{R,c,a}$ and $\sim_{D,c,a}$ satisfy the same non-monotonic axioms. It means that regular ASPIC^+ and ASPIC_D^+ have the same strength of constructing arguments in the context of satisfying non-monotonic axioms. This provides another answer to research question RQ1.

Table 6.6 shows that under the grounded semantics, $\sim_{R,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^g$ satisfy the same non-monotonic axioms; Table 6.7 shows that under the grounded semantics, $\sim_{R,c,j}^g$ and $\sim_{D,c,j}^g$ satisfy the same non-monotonic ax-

ioms; It means that ASPIC^+ theories and its defeasible version ASPIC_D^+ theories have the same justified conclusion under grounded semantics. This provides another answer to research question RQ2 under the grounded semantics.

Turning into the preferred semantics, Table 6.6 shows that under the sceptical preferred justified conclusions, $|\sim_{R,\emptyset,j}^{p,s}$ and $|\sim_{D,\emptyset,j}^{p,s}$ satisfy the same non-monotonic axioms. Under the credulous and universal preferred justified conclusions, $|\sim_{D,\emptyset,j}^{p,Sem}$ is stronger than $|\sim_{R,\emptyset,j}^{p,Sem}$ in the context of satisfying non-monotonic axioms, which means that ASPIC^+ and ASPIC_D^+ may not have the same justified conclusions. By adding the classical logic as base logic, the results are exactly same as empty base logic. This provides another answers to research question RQ2 under the preferred semantics.

Chapter 7

Conclusions and Future Work

7.1 Summary of Contribution

The aim of this dissertation was to explore the reasoning that is possible using just the defeasible portion of ASPIC⁺, the argumentation system that we call ASPIC_D⁺. In doing so, we provided the first in-depth investigation of a purely defeasible argumentation system and drew a direct comparison to what can be achieved in an equivalent system that distinguishes strict and defeasible knowledge. Having provided a lengthy analysis, we have answered the three research questions raised in Chapter 3. In this section we will recap our justification for carrying out this work and summarize our investigation into the properties of ASPIC_D⁺.

First, we answer the three research questions from the details of the system:

RQ1 Since ASPIC_D^+ is an subset of ASPIC^+ , do we lose any expressiveness in removing the strict parts of ASPIC^+ argumentation framework?

Answer One might imagine that not having strict rules means that ASPIC_D^+ is unable to represent all of the knowledge that can be captured by ASPIC^+ . However, Propositions 4.6–4.8 show, in a straightforward way, that this is not the case. Any ASPIC^+ theory can be converted an ASPIC_D^+ theory so that there is an equivalent set of arguments, and this can be done in such a way that the same preference ordering exists over the two sets of arguments. This means we do not lose anything by removing strict elements of the ASPIC^+ argumentation framework.

RQ2 Do the ASPIC^+ and ASPIC_D^+ versions of the same theory have the same justified conclusions?

Answer The much bigger question is whether these two sets of arguments have the same conclusions. Proposition 4.11 shows that under the grounded semantics, and given some reasonable assumptions about the way the conversion from ASPIC^+ to ASPIC_D^+ is carried out, the two sets of arguments will, indeed, have the same conclusions. Under other semantics, things are more complicated. Corollary 4.2 shows that it is possible to ensure that the two sets of arguments will have the same

conclusions under any complete semantics, but only at the cost of (in effect) reinstating the notion of strict rules in ASPIC_D^+ . The results of Proposition 4.11 and Corollary 4.2 hold when we start with well-defined ASPIC^+ theories. Given that all ASPIC_D^+ theories are well-defined, the paper also investigated the question of whether, given an ASPIC^+ theory AT_+ that is ill-defined, it is possible to ensure that the conclusions of the ASPIC_D^+ version of this theory, AT_D , are the same as those of the version of AT_+ that has been made well-defined. When AT_+ is ill-defined because it is not closed under transposition of strict rules, it turns out that relatively little can be said about the relationship between the conclusions of AT_D and AT_{tr} , but Propositions 5.2, 5.4 and 5.3, lay down conditions under which particular arguments from the two theories will have the same status. More can be said when AT_+ is ill-defined because it is axiom inconsistent. Thanks to the ability of argumentation to handle inconsistent information, it turns out that AT_D will generate exactly the same conclusions as versions of AT_+ which have been revised with some simple, but rather natural, belief revision functions. Propositions 5.7, 5.11 and 5.12 detail when this is the case. Overall, in general, we can ensure ASPIC^+ theories and their defeasible versions expressed in ASPIC_D^+ have the same justified

conclusion given reasonable assumptions. The assumptions differ in different situations.

RQ3 What is the advantage, if any, of removing the strict elements from ASPIC_D^+ ?

Answer With a system that only contains defeasible elements, we shed much of the complexity of ASPIC^+ . This is natural since much of the complexity of ASPIC^+ , in particular the need to restrict rebuttals, and the need to transpose strict rules, is required in order to accommodate strict rules. (Or, more precisely, they are required to ensure that ASPIC^+ can both use strict rules and satisfy the rationality postulates of [21]). As we show in Propositions 4.2–4.4, ASPIC_D^+ conforms to the rationality postulates of [21], and hence meets the minimum requirements of any reasonable argumentation system. ASPIC_D^+ does this without need to transpose strict rules (because it does not use strict rules), and without needing restricted rebut. Furthermore, as highlighted in Proposition 4.5, ASPIC_D^+ conforms to the rationality postulates using unrestricted rebut for all Dung semantics, a modest extension of what was previously possible using unrestricted rebut [23]. Therefore, we do have some advantages by removing strict elements — we obtain a

simpler and more natural system with less restrictions.

Then we answer the first two research questions from another perspective — non-monotonic properties. we considered which of the axioms of [46] ASPIC^+ and ASPIC_D^+ meet based on two different interpretations of the consequence relation, argumentation construction and justified conclusions. Table 6.6 and Table 6.7 shows that under the grounded semantics, $\sim_{R,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^g$ behave exactly the same. Neither of them is cumulative in general, however, for the strict interpretation, both $\sim_{R,\emptyset,j}^g$ and $\sim_{D,\emptyset,j}^g$ are cumulative. Now we turn to the preferred semantics. Because of unrestricted rebut, ASPIC_D^+ satisfies more axioms than ASPIC^+ . That is the defeasible version of Ref, LLE and RW with respect to credulous and universal preferred semantics, which is shown in Table 6.6 and Table 6.7.

Overall, the purely defeasible system ASPIC_D^+ does not have any restrictions — you do not need to worry about whether the strict rules are closed under transposition or whether the axioms are consistent. Therefore, ASPIC_D^+ is easier for a non-expert user, and on that basis it has been used in ArgTrust [64] and Consult [45]. On the other hand, if you want a system to distinguish between strict elements and defeasible elements, ASPIC^+ is the better choice. In addition, since any ASPIC^+ theory has a defeasible version which is an ASPIC_D^+ theory, and since these two theories have the

same justified conclusions under the grounded semantics, we can use either of the two systems if we are considering the grounded semantics.

7.2 Limitations

As discussed in Section 2.1.2, [60] defines ASPIC⁺ in terms of a language with two forms of negation — contradictories, which are symmetric, and contraries, which are asymmetric. For simplicity, we only considered symmetric negation in our discussion of ASPIC⁺ and ASPIC_D⁺. Here we consider how our results would change were we to use asymmetric negation — contraries — as well. In brief, our results would not change. The reason is as follows.

As discussed earlier in the dissertation, we start by considering an ASPIC⁺ theory AT and its defeasible counterpart AT_D . Both of these use a language \mathcal{L} that only includes symmetric negation. Now consider re-writing these using a language \mathcal{L}' that includes asymmetric negation but is otherwise identical to the language used for AT and AT_D . Call these rewritten theories AT' and AT'_D . Under the assumption that AT and AT_D denote every conflict between the propositions in \mathcal{L} , then rewriting AT and AT_D using contraries as well as contradictories will lead to some of the contradictories being rewritten as contraries¹. Thus, the set of attacks $Attacks'$ between the arguments con-

¹Our assumption therefore suggests that AT and AT_D may have, due to the imprecision \mathcal{L} , have identified conflicts where there were none, by turning asymmetric conflicts into

constructed from AT' will be a subset of the set of attacks $Attacks$ between the arguments constructed from AT , and the set of attacks between the arguments $Attacks'_D$ constructed from AT'_D will be a subset of the set of attacks between the arguments $Attacks_D$ constructed from AT_D . However, the set of attacks $Attacks$ will be a subset of $Attacks_D$, and $Attacks'$ will be a subset of $Attacks'_D$, since, just as in so many of our results above, the use of unrestricted rebuts in $ASPIC^+_D$ may lead to more attacks. Since the only difference between AT and AT_D , and between AT' and AT'_D are these different numbers of attacks, and since this difference is exactly what is taken into account in our results, our results will not change.

Another way of looking at this is that our results hold between any $ASPIC^+$ theory and its $ASPIC^+_D$ counterpart. The results all hinge upon the fact that the set of arguments constructed from the $ASPIC^+_D$ theory includes a larger set of attacks (because of unrestricted rebut). Removing some of the attacks from both the arguments from the $ASPIC^+$ theory, and the arguments from the $ASPIC^+_D$ theory (and it will be the same attacks from both) will not change anything.

symmetric ones.

7.3 Future Work

There are a number of interesting avenues for future work around the topics investigated in this dissertation. One line of future work is to continue investigating the properties of the ASPIC_D^+ framework. As discussed above, it is possible to ensure that the two sets of arguments will have the same conclusions under any complete semantics, but only at the cost of (in effect) reinstating the notion of strict rules in ASPIC_D^+ . One question that one might investigate is whether there is any other way to ensure that the two sets of arguments will give the same justified conclusions under any complete semantics.

Another potential line of future work is to relate ASPIC_D^+ to other argumentation systems. For example, [23] introduces the ASPIC- system, which satisfies the rationality postulates using unrestricted rebut, but only under the grounded semantics. Both ASPIC- and ASPIC_D^+ systems satisfy the rationality postulates using unrestricted rebut, and it would be interesting to explore connections between the two systems.

Another area to investigate is applications. The ASPIC_D^+ system has been used in different applications, such as ArgTrust [64] and Consult [45]. In the future, it would be interesting to apply the ASPIC_D^+ system in more

areas, where its simplified representation, with respect to ASPIC^+ , is found to be useful.

Aside from developing the theory itself, it would be interesting to investigate combining ASPIC_D^+ with justification logic [10, 9, 11]. As we know, argumentation can generate consistent results by reasoning in the domain of the knowledge base, i.e., given the knowledge base, the argumentation framework can make the correct conclusions. However, one would ask where does the knowledge base itself come from? In general, the knowledge base is a combination of observations and common knowledge, in the form of agent A knows/believes x . Justification logic can be used in reasoning with the knowledge, so that given information that agent A knows/believes x , justification logic can identify what should be included in the knowledge base. Thus a combination of argumentation and justification logic might be able to provide a more complete account of an agent's reasoning, combining reasoning with knowledge from a knowledge base and reasoning about what knowledge should be in the knowledge base.

Bibliography

- [1] T. Alsinet, C. I. Chesnevar, L. Godo, and G. R. Simari. A logic programming framework for possibilistic argumentation: Formalization and logical properties. *Fuzzy Sets and Systems*, 159(10):1208–1228, 2008.
- [2] L. Amgoud. Five weaknesses of ASPIC+. *Advances in Computational Intelligence*, pages 122–131, 2012.
- [3] L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, H. Prakken, J. van Veenen, and G. Vreeswijk. Final review and report on formal argumentation system. deliverable d2. 6. Technical report, ASPIC IST-FP6-002307, 2006.
- [4] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence*, pages 1–7. Morgan Kaufmann Publishers Inc., 1998.
- [5] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(3):197–215, 2002.
- [6] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.
- [7] L. Amgoud, C. Cayrol, M. C. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23:1062–1093, 2008.

- [8] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. A flexible framework for defeasible logics. In *Proceedings of American National Conference on Artificial Intelligence*, pages 405–410, 2000.
- [9] S. Artemov. Justification logic. In *European Workshop on Logics in Artificial Intelligence*, pages 1–4. Springer, 2008.
- [10] S. Artemov. The logic of justification. *The Review of Symbolic Logic*, 1(4):477–513, 2008.
- [11] S. Artemov. Why do we need justification logic? In *Games, Norms and Reasons*, pages 23–38. Springer, 2011.
- [12] J. C. Augusto and G. R. Simari. Temporal defeasible reasoning. *Knowledge and Information Systems*, 3(3):287–318, 2001.
- [13] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.
- [14] T. Bench-Capon, H. Prakken, and G. Sartor. Argumentation in legal reasoning. In *Argumentation in Artificial Intelligence*, pages 363–382. Springer, 2009.
- [15] P. Besnard and A. Hunter. A logic-based theory of deductive arguments? *Artificial Intelligence*, 128:203–235, 2001.
- [16] D. Billington. Defeasible logic is stable. *Journal of Logic and Computation*, 3(4):379–400, 1993.
- [17] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1):63–101, 1997.
- [18] I. Bratko, J. Žabkar, and M. Možina. Argument-based machine learning. In *Argumentation in Artificial Intelligence*, pages 463–482. Springer, 2009.
- [19] M. Caminada. Contamination in formal argumentation systems. In *Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence*, 2005.

- [20] M. Caminada and L. Amgoud. An axiomatic account of formal argumentation. In *Proceedings of the 20th National Conference on Artificial Intelligence*, pages 608–613. AAAI Press, 2005.
- [21] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5):286–310, 2007.
- [22] M. Caminada, W. Carnielli, and P. E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22:1207–1254, 2012.
- [23] M. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. In *Proceedings of the 5th International Conference on Computational Models of Argument*, pages 209–220, 2014.
- [24] M. Caminada and Y. Wu. On the limitations of abstract argumentation. In *Proceedings of the 23rd Benelux Conference on Artificial Intelligence, Ghent, Belgium*, 2011.
- [25] C. Cayrol. On the relation between argumentation and non-monotonic coherence based entailment. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1443–1448. Morgan Kaufmann Publishers Inc., 1995.
- [26] C. Cayrol, C. Devred, and M.-C. Lagasquie-Schiex. Handling controversial arguments in bipolar argumentation frameworks. In *Proceedings of the First International Conference on Computational Models of Argument*, pages 261–272. IOS Press, 2006.
- [27] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolar abstract argumentation systems. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artificial Intelligence*, chapter 4, pages 65–84. Springer, Berlin, Germany, 2009.
- [28] C. I. Chesñevar, A. G. Maguitman, and R. P. Loui. Logical models of argument. *ACM Computing Surveys*, 32(4):337–383, 2000.
- [29] K. L. Clark. Negation as failure. In *Logic and Data Bases*, pages 293–322. Springer, 1978.

- [30] S. Coste-Marquis, C. Devred, and P. Marquis. Symmetric argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 317–328. Springer, 2005.
- [31] M. Croitoru and S. Vesic. What can argumentation do for inconsistent ontology query answering? In *International Conference on Scalable Uncertainty Management*, pages 15–29. Springer, 2013.
- [32] K. Čyras and F. Toni. Non-monotonic inference properties for assumption-based argumentation. In *International Workshop on Theorie and Applications of Formal Argumentation*, pages 92–111. Springer, 2015.
- [33] P. M. Dung. An axiomatic analysis of structured argumentation for prioritized default reasoning. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 267–272. IOS Press, 2014.
- [34] P. M. Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, 231:107–150, 2016.
- [35] P. M. Dung, R. A. Kowalski, and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170(2):114–159, 2006.
- [36] P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009.
- [37] P. M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(2):642–674, 2007.
- [38] J. Fox and S. K. Das. *Safe and sound: artificial intelligence in hazardous applications*. MIT Press, 2000.
- [39] J. Fox, P. Krause, and M. Elvang-Gøransson. Argumentation as a general framework for uncertain reasoning. In *Proceedings of the 9th International Conference on Uncertainty in Artificial Intelligence*, pages 428–434, 1993.

- [40] D. M. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In *Logics and Models of Concurrent Systems*, pages 439–457. Springer, 1985.
- [41] A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(2):95–138, 2004.
- [42] P. Gärdenfors, editor. *Belief Revision*. Cambridge University Press, 1992.
- [43] G. Governatori, M. J. Maher, G. Antoniou, and D. Billington. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14(5):675–702, 2004.
- [44] A. Hunter. Base logics in argumentation. *Frontiers in Artificial Intelligence and Applications*, 216:275–286, 2010.
- [45] N. Kokciyan, I. K. Sassoon, A. P. Young, M. D. Chapman, T. R. Porat, M. Ashworth, V. Curcin, S. Modgil, S. Parsons, and E. I. Sklar. Towards an argumentation system for supporting patients in self-managing their chronic conditions. In *AAAI Joint Workshop on Health Intelligence*, New Orleans, 2018.
- [46] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial intelligence*, 44(1):167–207, 1990.
- [47] P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.
- [48] P. Krause, S. Ambler, M. Elvang-Goransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11(1):113–131, 1995.
- [49] H. E. Kyburg Jr. *Probability and the logic of rational belief*. Wesleyan University Press, 1961.

- [50] Z. Li and S. Parsons. On argumentation with purely defeasible rules. In *Scalable Uncertainty Management*, pages 330–343. Springer, 2015.
- [51] F. Lin and Y. Shoham. Argument systems: A uniform basis for non-monotonic reasoning. *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning*, 89:245–255, 1989.
- [52] R. P. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3(3):100–106, 1987.
- [53] N. Mann and A. Hunter. Argumentation using temporal knowledge. In *Proceedings of the 2008 Conference on Computational Models of Argument*, pages 204–215. IOS Press, 2008.
- [54] P. McBurney and S. Parsons. Dialogue games for agent argumentation. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, pages 261–280. Springer, 2009.
- [55] J. McCarthy. Circumscription—a form of nonmonotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.
- [56] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28(1):89–116, 1986.
- [57] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. *Readings in Artificial Intelligence*, pages 431–450, 1969.
- [58] D. McDermott and J. Doyle. Non-monotonic logic i. *Artificial Intelligence*, 13(1-2):41–72, 1980.
- [59] S. Modgil and H. Prakken. Revisiting preferences and argumentation. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, pages 1021–1026. AAAI Press, 2011.
- [60] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 2012.

- [61] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: A tutorial. *Argument & Computation*, 5(1):31–62, 2014.
- [62] R. C. Moore. Possible-world semantics for autoepistemic logic. Technical report, DTIC Document, 1984.
- [63] D. Nute. Defeasible logic. In *International Conference on Applications of Prolog*, pages 151–169. Springer, 2001.
- [64] S. Parsons, E. Sklar, J. Salvit, H. Wall, and Z. Li. Argtrust: decision making with information from sources of varying trustworthiness. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 1395–1396. International Foundation for Autonomous Agents and Multiagent Systems, 2013.
- [65] D. Phan Minh. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-persons games. *Artificial Intelligence*, 77(2):321–358, 1995.
- [66] H. Prakken. AI & law, logic and argument schemes. *Argumentation*, 19(3):303–320, 2005.
- [67] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.
- [68] H. Prakken. Rethinking the rationality postulates for argumentation-based inference. *Proceedings of Computational Models of Argument 2016*, 287:419, 2016.
- [69] H. Prakken and S. Modgil. Clarifying some misconceptions on the ASPIC+ framework. *Computational Models of Argument: Proceedings of Comma 2012*, page 442, 2012.
- [70] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7(1-2):25–75, 1997.
- [71] H. Prakken and G. Vreeswijk. Logical systems for defeasible argumentation. *Handbook of Philosophical Logic*, 2002.

- [72] I. Rahwan and C. Reed. The argument interchange format. In *Argumentation in Artificial Intelligence*, pages 383–402. Springer, 2009.
- [73] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1):81–132, 1980.
- [74] G. R. Simari and R. P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2-3):125–157, 1992.
- [75] E. I. Sklar, S. Parsons, Z. Li, J. Salvit, S. Perumal, H. Wall, and J. Mangels. Evaluation of a trust-modulated argumentation-based interactive decision-making tool. *Autonomous Agents and Multi-Agent Systems*, 30(1):136–173, 2016.
- [76] T. Takahashi and H. Sawamura. A logic of multiple-valued argumentation. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 800–807. IEEE Computer Society, 2004.
- [77] Y. Tang, K. Cai, P. McBurney, E. Sklar, and S. Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, 22(5):979–1018, 2012.
- [78] Y. Tang, E. Sklar, and S. Parsons. An argumentation engine: Argtrust. In *Ninth International Workshop on Argumentation in Multiagent Systems*, 2012.
- [79] S. E. Toulmin. *The uses of argument*. Cambridge university press, 2003.
- [80] D. S. Touretzky. *The mathematics of inheritance systems*. Morgan Kaufmann, 1986.
- [81] B. Verheij. Two approaches to dialectical argumentation: Admissible sets and argumentation stages. In J.-J. Meyer and L. van der Gaag, editors, *Proceedings of the Eighth Dutch Conference on Artificial Intelligence*, pages 357–368, Utrecht, 1996.
- [82] G. A. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90(1):225–279, 1997.

- [83] D. Walton. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press, 1995.
- [84] Y. Wu and M. Podlaszewski. Implementing crash-resistance and non-interference in logic-based argumentation. *Journal of Logic and Computation*, pages 1–31, 2014.