UNIVERSITÉ DU LUXEMBOURG

PhD-FSTC-2015-13
The Faculty of Science, Technology and Communication

DISSERTATION

Defense held on 13/03/2015 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN INFORMATIQUE

by

Mikołaj PODLASZEWSKI
Born on 05/06/1979 in Gdańsk (Poland)

# Poles Apart
## Navigating the Space of Opinions in Argumentation

### Dissertation defense committee

Prof. Dr. Leon van der Torre, Supervisor
*Professor, Université du Luxembourg*

Prof. Dr. Pierre Kelsen, Chairman
*Professor, Université du Luxembourg*

Prof. Dr. Anthony Hunter, Vice Chairman
*Professor, University College London*

Dr. Sanjay Modgil, Member
*Lecturer, King's College London*

Dr. Martin Caminada, Member
*Research Fellow, University of Aberdeen*

Dr. Richard Booth, Expert in an advisory capacity
*Lecturer, Mahasarakham University, Thailand*

# ABSTRACT

Formal argumentation is a popular reasoning method in knowledge representation for intelligent systems. For the past 20 years it has been based on Dung's abstract argumentation theory. More recently several challenges have been made to this standard - for example in dynamics and aggregation of argumentation frameworks. To support these new developments in this thesis new foundations are developed based on distance measures. We introduce postulates for distance measures and we show their consistency by constructing concrete measures. In the process we develop the new notion of issue. Subsequently we use the distance measures in argumentation using distance based operators introduced by Miller and Osherson in judgment aggregation. Moreover in this thesis we also improve dialectical proof procedures for grounded semantics and study postulates of non-interference and crash resistance for Dung based non-monotonic inference.

# ACKNOWLEDGEMENTS

The initial goal of this project was to use argumentation in computer simulation explaining market inefficiencies and through many ups and downs ended up investigating some technical results in argumentation. I would have never done it without the help of the people around me.

I would like to thank my supervisors Leon van der Torre, Martin Caminada and Richard Booth for their guidance, patience and helping me to stay focused in the crucial moments. I am especially thankful to Richard who spent many hours listening to me and helping formulate my ideas.

I feel grateful to be surrounded by people from whom I could learn. Especially I would like to thank my coauthors Martin Caminada, Richard Booth, Yining Wu, Gabriella Pigozzi, Iyad Rahwan - working with them taught me a lot. Moreover, this period of my life wouldn't have been the same without teaching. I am thankful to Leon for insisting on that. I learned a lot assisting him as well as Emil Weydert, Srdjan Vesic and Richard Booth in classes.

My work was also influenced by the discussions, the positive attitude and the good atmosphere among my friends and fellow students Eugen Costezki, Pouyan Ziafati, Tjitze Rienstra, Muriel Foulonneau, Martin Kracheel, Simon Kramer, Wojtek Jamroga, Gabriela Gheorghe, Patrice Caire, Marija Slavkovik, Paolo Turrini, Diego Agustín Ambrossio, Silvano Colombo Tosatto, Marc van Zee, Llio Humphreys, Samy Soares Sa and Masoud Tabatabaei. Thank you.

Finally, I would like to thank my wife Svilena and my sons for their love, support and patience.

# CONTENTS


1 INTRODUCTION 1
- 1.1 Dung's Abstract Argumentation 2
- 1.2 Dung for Non-monotonic Inference 5
- 1.3 Research Questions 6
  - 1.3.1 Quantifying Disagreement 6
  - 1.3.2 Judgement Aggregation in Abstract Argumentation 7
  - 1.3.3 Grounded Persuasion Game 8
  - 1.3.4 Implementing Crash Resistance 9
- 1.4 Relevance 10
- 1.5 Thesis Overview 11

2 PRELIMINARIES 15
- 2.1 Argumentation Framework 15
- 2.2 Semantics 15
  - 2.2.1 Extension-based Semantics 15
  - 2.2.2 Labelling-based Semantics 16

3 QUANTIFYING DISAGREEMENT BETWEEN LABELLINGS: POSTULATES 19
- 3.1 Introduction 19
- 3.2 Distance between Labellings 20
- 3.3 Postulates for Distance Methods 20
  - 3.3.1 Group I - Metric Postulates 21
  - 3.3.2 Group II - Intuition-based Postulates 21
  - 3.3.3 Group III - Compositionality Postulates 27
  - 3.3.4 Group IV - Equivalence Postulates 30
- 3.4 Summary 33

4 QUANTIFYING DISAGREEMENT BETWEEN LABELLINGS: PRODUCT DISTANCE 35
- 4.1 Introduction 35
- 4.2 Full Sum Method 36
- 4.3 Critical Set Based Methods 39
- 4.4 Irresolute Product Measures 47
- 4.5 Issue-based Measures 49
- 4.6 Conclusion and Related Work 54
  - 4.6.1 Conclusions 54
  - 4.6.2 Related Work 56

5 USING DISTANCES FOR AGGREGATION IN ABSTRACT ARGUMENTATION 59
- 5.1 Introduction 59
- 5.2 Aggregation Preliminaries 60
- 5.3 Miller and Osherson Aggregation Methods. 60
  - 5.3.1 Initial Aggregation Methods 62

5.3.2 Extension to the Set of All Labellings 63
5.3.3 Example of the MO Methods 65
5.4 Agenda Manipulation 66
5.5 Down-admissible and Up-complete as a MO Method 67
5.6 Conclusions 69
6 A PERSUASION DIALOGUE FOR GROUNDED SEMANTICS 71
6.1 Introduction 71
6.2 A Dialogue Game for Grounded Semantics 72
6.3 Discussion 84
7 IMPLEMENTING CRASH-RESISTANCE AND NON-INTERFERENCE 87
7.1 Introduction 87
7.2 Preliminaries 90
7.2.1 The Three-Step Argumentation Process 90
7.2.2 Rationality Postulates 94
7.2.3 The *ASPIC Lite* System 96
7.3 Problems and Postulates 98
7.4 Solution 104
7.5 Related Work 119
7.6 Summary and Future Work 120
8 CONCLUSIONS 125

BIBLIOGRAPHY 131

# LIST OF FIGURES


Figure 1 Simple argumentation framework 2
Figure 2 An argumentation framework with four complete labellings. 3
Figure 3 Argumentation for inference 5
Figure 4 An argumentation framework with all its complete labellings ($L_1$-$L_6$), and three other labellings ($N_1$-$N_3$). 18
Figure 5 An argumentation framework with all its six complete labellings. 22
Figure 6 Two pairs of mutually attacking arguments 25
Figure 7 AF with two independent parts. 27
Figure 8 AF with semantically but not syntactically independent partition 29
Figure 9 2-loop $\mathcal{A}_1$ and its extension $\mathcal{A}_2$ with peripheral issues $c, d, e$ 32
Figure 10 The dependencies between the postulates. 1) For $d$ satisfying **(LAB$_{Sem}$)**, 2) For *Sem* satisfying Independence principle, 3) For *Sem* satisfying Language Independence principle. 34
Figure 11 Three possible complete labellings $L_1, L_2$ and $L_3$ 37
Figure 12 Example showing opposite results of Hd and rHd. 38
Figure 13 The AF $\mathcal{A}_1$ and its four possible complete labellings $L_1$-$L_4$ 46
Figure 14 The AF for which all minimal critical set distance measures fail **(SQDA)** 46
Figure 15 Example showing that *mincd* fails the triangle inequality 47
Figure 16 AF with two independent parts. 48
Figure 17 Source labelling $S$ and 2 target labellings $T_1$, $T_2$ 50
Figure 18 Two aggregation frameworks. 64
Figure 19 Aggregation frameworks extended with dependant chains. 66
Figure 20 An Argumentation framework with simple reinstatement 73
Figure 21 Two paths coming from the same argument 74
Figure 22 Two arguments attacking each other 75
Figure 23 Argumentation for inference 88

Figure 24 Arguments $J_1$ and $M_1$ “contaminate” argument $W_1$. 99
Figure 25 Switching from grounded to preferred semantics still does not always provide a solution. 100
Figure 26 The argumentation framework $\mathcal{A}_2$ built from $\mathcal{B}_2$. 103
Figure 27 Repaired argumentation framework 107
Figure 28 Different types of arguments produced from a union of two syntactically disjoint knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$ 112
Figure 29 Deleting self-defeating arguments from the AF of Figure 25 119
Figure 30 The argumentation framework of Example 163 120
Figure 31 The argumentation frameworks illustrating problems with integration of preferences: a) ASPICLite, b) ASPICLite with Last Link, c) inconsistency-cleaned version of ASPICLite with Last Link 121

# LIST OF TABLES


Table 1 The overview of the postulates 33

Table 2 Restrictions placed by Metric and Intuition-based postulates 55

Table 3 Restrictions placed by Compositionality and Equivalence 55

Table 4 Summary of Properties of proposed Distance Measures. All included distances satisfy **(REF)**,**(DD)**,**(SYM)** and **(TRI)**. 57

Table 5 Distance between labellings by $\mathtt{eid}^{\mathit{diff}}$ with refined Hamming distance $\mathit{diff}_{rh}$ over the labels of $\mathcal{A}_2$ from Figure 18. 64

Table 6 Aggregation of the profile $\mathbf{L} = (\mathtt{l}_4 : \mathtt{uuoio}, \mathtt{l}_5 : \mathtt{uuuoi}, \mathtt{l}_5 : \mathtt{uuuoi})$ - outcomes for different MO aggregation methods used with: extended issue-based labelling distance ($\mathtt{rEIBd}$) (left column), and sum over all arguments ($\mathtt{rHd}$) (right column). In both columns refined hamming distance over the labels and credulous initial operator is used. 65

Table 7 Two profiles $(\mathtt{l}_5, \mathtt{l}_9)$ and $(\mathtt{l}_4, \mathtt{l}_5)$ for which $\mathtt{cio}$ violates collective rationality and is repaired with down-admissible up-complete procedures. 68

# ACRONYMS

**NMR** Non-monotonic Reasoning

**AI** Artificial Intelligence

**MAS** Multi Agent Systems

**JA** Judgement Aggregation

**KR** Knowledge Representation

# 1 INTRODUCTION

In everyday life humans use both deductive and defeasible reasoning. In contrast to deductive reasoning where the truth of a conclusion is guaranteed upon the truth of premises, defeasible conclusions are likely but not necessarily true and can be withdrawn upon additional information. For example 'Party is on Wednesday' can be concluded from 'John said that party is on Wednesday' and retracted upon information 'John is joking'. 'It rained' can be concluded upon observation 'Street is wet' and retracted upon additional information that 'Street was washed by public service'. The reader can find more examples of defeasibe reasoning in the overview by Koons (2014).

Defeasible reasoning is studied in philosophy and artificial intelligence (AI). Today with the omnipresence of computational power it has become of practical importance to implement this kind of reasoning. Examples range from simple plugins asking whether we forgot to attach a file to an email mentioning an attachment, to complex expert systems supporting medical diagnosis. While simple cases like writing assistance can be implemented directly, complex systems require a systematic approach. One of the ways to implement defeasible reasoning is to use *non-monotonic logic*. Many formal systems for non-monotonic logic have been developed. For an overview of approaches and formalisms we refer the reader to Strasser and Antonelli (2014). In 1995 Dung introduced his theory of *abstract argumentation* and demonstrated it can model popular non-monotonic formalisms like Reiter's default logic (1980), Pollock's defeasible logic (1987) and logic programming Gelfond and Lifschitz (1988, 1991), Dung (1995). Subsequently his theory was extended in several ways and remains a popular topic of research in AI.

This thesis consists of three loosely connected parts in which we study Dung's abstract argumentation theory:

1. Quantifying disagreement between argument labellings,
2. Developing persuasion dialogue for grounded semantics,
3. Implementing the postulates of Crash-resistance and Non-interference.

Before providing more details in Section 1.3, there is a need for some background information.

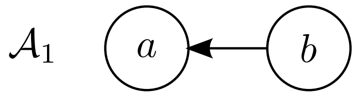



Figure 1: Simple argumentation framework

## 1.1 DUNG'S ABSTRACT ARGUMENTATION

Nowadays, much research on the topic of argumentation is based on the abstract argumentation theory of Dung (1995). The central concept in this theory is that of an *argumentation framework* (AF), which is essentially a directed graph in which the arguments are represented as nodes and the defeat (attack) relation is represented by the arrows. The theory abstracts away the content of arguments and based solely on the attack structure between them addresses the question which arguments to accept. The selection of the argument is non-monotonic, since adding more arguments to the framework possibly yields fewer accepted arguments. It has been shown to capture the non-monotonic part of reasoning modelled by other formalisms. In the rest of this section we introduce informally concepts used in the theory.

Consider the following exchange of arguments:

- $a$: Global warming is mainly caused by volcanic activity, according to research by expert X.
- $b$: Research of expert X is financed by the oil industry which has financial interest in the results, thus the research is not credible.

After abstracting away the content of arguments the above example is modelled by an argumentation framework $\mathcal{A}_1$ depicted in the Figure 1 where argument $b$ attacks argument $a$.

Given such a graph, the remaining task is to decide which arguments to accept. In the above situation assuming no more arguments can be created, one can reason as follows: since there are no reasons to reject the argument $b$ I accept it and I reject the argument $a$ since it is attacked by $b$ which I accepted. Although in this case the answer is straightforward there can be more than one reasonable answer. Consider three arguments about global warming, each one grounded in some scientific evidence, with the following associated conclusions:

- $a$: Global warming is mainly caused by volcanic activity.
- $b$: Global warming is mainly caused by natural variation in solar radiation.
- $c$: Global warming is a human-induced phenomenon.

This situation is modelled in argumentation framework $\mathcal{A}_2$ in Figure 2. Clearly, it is not possible to subscribe to both arguments $a$

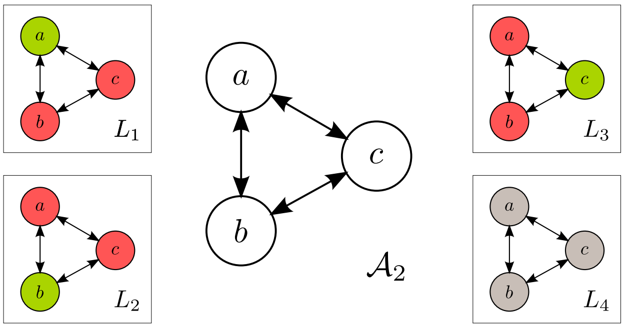



Figure 2: An argumentation framework with four complete labellings.

and b, since they attribute global warming to different major causes. Moreover, both these arguments attack argument c, which attributes global warming to human activity. In this situation at least four positions seem possible. One may accept one of the three arguments and reject the other two or one may abstain from taking any decision and stay undecided about all arguments.

Answering the question which arguments can be accepted corresponds to defining an *argumentation semantics*. In this thesis we express semantics in two ways. Extension-based semantics for each argumentation framework returns the set of extensions, where an extension is a set of arguments which can be accepted. Labelling-based semantics returns for each argumentation framework the set of labellings. A labelling of an argumentation framework is a function assigning to each argument one of the labels `in`, `out` or `undec` which correspond to acceptance of an argument, rejection of an argument or abstaining from taking a position about an argument.

Various argumentation semantics have been formulated in this respect, and in Chapter 2 we describe some of the mainstream approaches. Here we just state intuitively the main criteria on which the semantics are based. The first criterion is that of *conflict freeness* which says that if there is attack between two arguments, then they cannot be accepted together. The second criterion is that of *defence* which says that accepted arguments should defend themselves, i.e. for each argument a which attacks an accepted argument, there exists an accepted argument which attacks a. A conflict-free set of arguments which defends itself is called *admissible*. The third criterion is that of *reinstatement* which says that accepted arguments should include all arguments they defend. Those criteria can be expressed in terms of labellings in the following way:

1. if an argument is labelled `in`, then all its attackers are labelled `out`.

2. if an argument is labelled `out`, then it has an attacker that is labelled `in`.

3. if an argument is labelled `undec`, then neither all its attackers are labelled `out` and nor does it has an attacker that is labelled `in`.

In the last example depicted in Figure 2 we have exactly four labellings $L_1 - L_4$ satisfying the above conditions. We depict them by colouring the nodes labelled `in`, `out`, `undec` with green, red and grey colour respectively. When the order of arguments is clear we denote a labelling by a string of labels assigned to consecutive arguments. For example $L_1$: `ioo` denotes labelling $L_1 = \{(a, \texttt{in}), (b, \texttt{out}), (c, \texttt{out})\}$.

The semantics using those rules (and possibly some additional ones) are called *complete-based*. The *complete* semantics does not add any additional rules, the *grounded* semantics requires additionally that the set of accepted arguments be minimal while the *preferred* semantics requires the set of accepted arguments be maximal. In Figure 2 all labellings are complete, labellings $L_1 - L_3$ are preferred, labelling $L_4$ is the grounded labelling. Accordingly $\{a\}, \{b\}, \{c\}$, and $\emptyset$ are corresponding preferred and grounded extensions.

To apply the theory we need to instantiate it, i.e. decide what an argumentation framework represents and how to interpret the meaning of acceptance. Although resemblance to human reasoning is an advantage of Dung's theory, it is a metaphor. Argumentation vocabulary and our natural text examples can be misleading here. In his original paper Dung (1995) gives five examples of instantiation. Three of them model non-monotonic logic formalisms, where arguments are generated from a set of logical formulas and accepted arguments support an inferred formula. We describe this kind of instantiation in the next section. N-person games model reasoning about coalition forming, arguments represent the possible coalitions and the attack relation represents a dominance relation. The stable marriage problem is addressed in a similar way; accepted arguments correspond to stably formed couples. Each of those instantiations combined different argumentation framework construction, argumentation semantics and a way to interpret accepted arguments. It is not trivial to connect these three parts.

Dung's abstract theory was developed in several ways:

ABSTRACT On the abstract level more semantics were developed, argumentation frameworks were enriched with other types of relations like support, attacks on attacks, arguments were given more attributes like a supporting value, acceptance was extended to multiple labels.

INSTANTIATION New instantiations were proposed, logical and nonlogical.

We move to the use of Dung in modelling of formalisms for nonmonotonic reasoning.

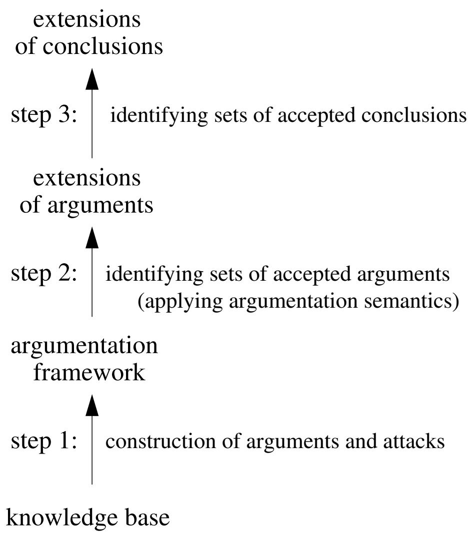



Figure 3: Argumentation for inference

## 1.2 DUNG FOR NON-MONOTONIC INFERENCE

Argumentation as a tool for practical reasoning was already studied by Aristotle. In modern times the field of defeasible argumentation can be traced back to the work of Pollock (1992, 1995), Vreeswijk (1993, 1997), and Simari and Loui (1992). The idea is that (non-monotonic) reasoning can be performed by constructing and evaluating arguments, which are composed of a number of reasons for the validity of a claim. Arguments distinguish themselves from proofs by the fact that they are defeasible, that is, the validity of their conclusions can be disputed by other arguments. Whether a claim can be accepted therefore depends not only on the existence of an argument that supports this claim, but also on the existence of possible counter-arguments, that can themselves be defeated by counter-arguments, etc.

Dung's argumentation theory contributes to this line of research. If one wants to use it for the purpose of (non-monotonic) entailment, one can distinguish three steps (see Figure 3). First of all, one would use an underlying knowledge base to generate a set of arguments and determine in which ways these arguments attack each other (step 1). The result is an argumentation framework, represented as a directed graph in which the internal structure of the arguments, as well as the nature of the defeat relation, has been abstracted away. An argumentation semantics can be used to determine the sets of arguments that can be accepted (step 2). After the set(s) of accepted arguments have been identified, one then has to identify the set(s) of accepted conclusions (step 3), for which various approaches exist.

As illustrated in Figure 3, the argumentation approach provides a graph based way of performing non-monotonic reasoning. Dung's idea is to isolate the evaluation of arguments in step 2, which captures the non-monotonic part of the reasoning process, using an argumentation framework as an interface with the rest of the process. This isolation proved useful for comparison of different systems.

Considering Dung's argumentation in the context of non-monotonic reasoning several questions arise. From the point of view of a designer of logical formalisms we can ask: Is instantiating of Dung's abstract argumentation frameworks a good methodology to develop new formalisms? Adding the additional steps potentially restricts us. There are many formalisms already available. How do formalisms defined as instantiations of Dung's theory perform compared to others? Does it make sense to express already existing formalisms equivalently in Dung's framework? What are the properties that formalism in Dung's form highlight? The last question is especially important from an application point of view. Which of the two equivalent forms to choose? Answers to those questions are not clear. In this thesis we make a small step to get closer to them.

## 1.3 RESEARCH QUESTIONS

### 1.3.1 *Quantifying Disagreement*

The presence of multiple reasonable positions raises a fundamental question:

> **RQ 1:** How can we measure a distance between viewpoints represented by labellings in argumentation?

This question is relevant to two fundamental problems. The first problem is *argument-based belief revision*. Suppose a diplomat receives instructions to switch his position on one particular argument . To maintain a consistent viewpoint, the diplomat must revise his evaluation of other related arguments. Faced with multiple possibilities, the diplomat may wish to choose the one that differs the least from his initial position (e.g. to maintain credibility).

The issue of distance is also relevant to the problem of *judgement aggregation* over how a given set of arguments should be evaluated collectively by a group of agents with different opinions Caminada and Pigozzi (2011); Caminada et al. (2011); Rahwan and Tohmé (2010). For instance it is very well possible that the members of a jury in a criminal trial all share the same information on the case (and hence have the same argumentation framework) but still have different opinions on what the verdict should be. Hence, these differences of opinion are consequences not of differences in the knowledge base but of the nature of non-monotonic reasoning, which allows for various reasonable positions (extensions). In the context of judgement aggregation one may examine the extent to which the collective position differs from the various positions of the individual participants. Ideally, one would like to have a collective position that is closest to the collection of individual positions, for example such that the sum of its distance to each individual position is minimal.

The success criterion of RQ1 is to define distance measures which takes into account the argumentation semantics. Since Dung's argumentation framework can be used with different semantics like admissible, grounded, complete which give a meaning to the particular labelling we should expect the distance between two complete labelling may differ from two admissible labellings. There exist general distance measures like Hamming distance, but they treat labellings as a vectors of labels ignoring what these vectors represent.

Our methodology is top-down followed by bottom-up. First we list different possible ways the semantics can influence a distance measure as a postulates and study the dependencies between them. This leads to the first subquestion:

> **RQ 1.1:** What are desirable properties of distance measures for labellings?

The argumentation semantics assigns a set of labellings to the given argumentation framework. We would like to capture the influence of semantics on distance between labellings of one framework and also across different frameworks.

Then we state the following subquestion:

> **RQ 1.2:** Are those postulates jointly consistent?

To address the above subquestion we use topological constructions to actually build the distances satisfying the postulates.

### 1.3.2 *Judgement Aggregation in Abstract Argumentation*

Individuals presented with the same set of conflicting arguments might take different rational positions. In such a situation one often faces the following problem: How to aggregate positions of different agents over sets of arguments into a collective one? This problem has been explored in a number of recent papers Booth et al. (2014); Caminada and Pigozzi (2011); Rahwan and Tohmé (2010) which employ techniques from *judgement aggregation* (JA) List and Puppe (2009) to the problem of aggregating 3-valued argument labellings.

> **RQ 2:** How to use distance for aggregation of judgements represented by labellings?

The success criterion for RQ2 is to define aggregation operators for abstract argumentation.

The existing works mentioned above have shown that, as with classical judgement aggregation, it is not possible to define general aggregation operators that satisfy a number of seemingly mild constraints while ensuring collective rationality of the outcome. One way of getting around this problem is to first use an *initial* aggregation operator, which intuitively can be thought of as a *gold standard* operator that satisfies a number of basic postulates, without always yielding collectively rational results and then to *repair* the result of this operator in the cases when it does not give a collectively rational outcome. This leads to the following question.

> **RQ 2.1:** How can we 'repair' the collective outcome when it is not rational?

In the argumentation setting, Caminada and Pigozzi suggested one way to carry out such a repair, using what they called the *down-admissible* and *up-complete* procedures (2011). In the JA setting, another way to carry out such a repair is to use one of the *distance-based solution methods* that were studied by Miller and Osherson (hereafter MO) (2009) within the framework of binary judgement aggregation. As the name suggests these methods depend on a provided notion of distance measure between binary judgement sets.

### 1.3.3 *Grounded Persuasion Game*

The field of formal argumentation can be seen as consisting of two main lines of research. One line of research is concerned with the dialectical process of two or more players who are involved in a discussion. This kind of argumentation, referred to as *dialogue theory* in the ASPIC project ASPIC-consortium (2005), can be traced back to the work of Hamblin (1970; 1971) and Mackenzie (1979; 1990). A different line of research is concerned with arguments as a basis for non-monotonic inference. The idea is that (non-monotonic) reasoning can be performed by constructing and evaluating arguments, which are composed of a number of reasons that collectively support a particular claim. This line of research can be traced back to the work of Pollock (1992; 1995), Vreeswijk (1993; 1997) and Simari and Loui (1992), and has culminated with the work of Dung (1995), which serves as the basis of much of today's argumentation research.

One particular question one may ask is to what extent it is possible to create links between these two lines of research. One particular way of doing so would be to have an argument accepted (under a particular Dung-style semantics) iff it can be defended in a particular type of formal dialogue. In previous work, Caminada (2010) observed that (credulous) preferred semantics can be reinterpreted as a particular type of Socratic dialogue. That is, an argument is in at least one preferred extension iff the proponent is able to successfully defend the argument in the associated Socratic discussion game, against a maximally sceptical opponent. We follow this line of research by addressing the following question.

> **RQ 3:** What type of dialogue can be associated with grounded semantics?

One of the aims of our work is to contribute to a conceptual basis for (abstract) argumentation theory. Whereas, for instance, classical logic is based on the notion of truth, it is not immediately obvious where a notion like truth would fit in when it comes to (abstact) argumentation research. Still, one would like to determine what the various argumentation semantics actually constitute to. An answer like "preferred semantics is about the maximal conflict-free fixpoints, whereas grounded semantics is about the minimal conflict-free fixpoint" might be technically correct, but is still conceptually somewhat unsatisfying. We believe that formal dialogue can serve as a conceptual basis for (abstract) argumentation theory. The idea is that one infers not so much what is *true*, as is the case in classical logic, but what can be *defended in rational discussion*. In particular, our aim is to show that different argumentation semantics correspond with different *types* of rational discussion.

### 1.3.4 *Implementing Crash Resistance*

The ASPIC+ Prakken (2010); Modgil and Prakken (2013) is a state-of-the-art framework for specifying argumentation based systems for rule-based defeasible reasoning. The specification includes a set of strict and defeasible rules over some language closed under negation. If we consider Dung's style argumentation as a methodology to develop a non-monotonic logic formalism it is reasonable to ask how this formalism performs compared to others and whether it is easy to implement new features. Caminada and Amgoud (2007) provided some answers to this question introducing three postulates for argument-based entailment: *Direct Consistency*, *Indirect Consistency* and *Closure*, and noticing that they should not be taken for granted. Unrestricted instantiation of ASPIC+ fails them unless the set of strict rules needs to be closed under contraposition or transposition as shown by Caminada and Amgoud (2007); Modgil and Prakken (2013); Prakken (2010). Classical logic is closed by negation and its entailment is closed under contraposition or transposition. Moreover, entailment in classical logic models deductive reasoning which is similar to the reasoning represented by strict rules. The question arises:

> **RQ 4:** What are the consequences of generating strict rules by the entailment of classical logic?

The main concern here is to satisfy postulates of *Crash-resistance* and *Non-interference* Caminada et al. (2012) which deals with the principle of explosion - the common problem in classical logic entailment which entails any formula from contradiction. We follow the methodology sketched by Caminada (2005), i.e. we remove inconsistent arguments. This leads to the subquestion:

> **RQ 4.1:** Is it feasible to implement *Crash-resistance* and *Non-interference* in Argumentation-based logical formalism by removing inconsistent arguments?

The success criteria is to demonstrate the implementation or give an example showing that it is impossible. By feasible we mean implementation which extends the previously obtained results.

## 1.4 RELEVANCE

This thesis is about formal argumentation, a domain relevant to Artificial Intelligence (AI) which is part of Computer Science (CS). Argumentation play several roles in AI and wider in CS.

NON-MONOTONIC REASONING (NMR) NMR is an area of AI which aims at modelling natural reasoning. In his seminal work Dung (1995) showed that reasoning as modelled by several of non-monotonic formalisms like Reiter's default logic (1980), Pollock's defeasible logic (1987), and logic programming (1988) can be formulated as a three step process of constructing arguments, evaluating them and taking their conclusion. A recent examples in this line of research include Modgil and Prakken (2013); Prakken (2010) (ASPIC+ framework) and Besnard and Hunter (2001).

KNOWLEDGE REPRESENTATION (KR) This is a wide area studying how to represent knowledge. In particular, study of arguments which are a natural means to exchange information between humans. In the above-mentioned systems arguments are built from the knowledge-base expressed in artificial language which are potentially easier to understand by humans. Another source of arguments is the growing body of tools to analyse text Bex et al. (2013) which enables application of computer tools.

MULTIAGENT-SYSTEMS (MAS) MAS is a field studying systems of intelligent autonomous agents (human and artificial) which interact within an environment. On the one hand, problems in this context are often similar to the one encountered by groups of people. Endriss (2014) argues that resource allocation, voting and judgement aggregation, problems originally studied in Social Choice Theory form a foundation of MAS. On the other hand, the internet provides an environment to build variety of systems for human agents. An example where argumentation plays a role are systems supporting humans in performing large-scale online debates. See the Arguing Web 2.0[1] workshop for some current work in that direction.

HUMAN-COMPUTER INTERACTION (HCI) With the increase of complexity of computer systems, the communication and interaction with humans become an important issue. It is studied by a recent interdisciplinary field of Human-computer Interaction. One of the assets of argumentation-based systems important in that context is their resemblance to human reasoning. An example of the research in that domain is the SAsSy project conducted at University of Aberdeen, Tintarev et al. (2013).

For a detailed look at the influence of argumentation we refer the reader to Modgil et al. (2013).

Work in this thesis contributes to several points of the map sketched above. Distances developed in Chapter 3 and 4 are subsequently applied to judgement aggregation in Chapter 5 thus contributing to MAS. Distances are used in many domains of computer science. Our work could be potentially applied to clustering which can be used for visualisation which is part of HCI. We also uncover an interesting new problem in graph theory, namely the one how to measure the distance between two graph colourings. The Grounded Persuasion Game developed in Chapter 6 contributes to HCI in two ways. First, linking grounded semantics with persuasion dialogue contributes to better understanding what makes argumentation "human-friendly". Second, it can serve as a basis for a development of dialogue-based interface for expert system. In Chapter 7 we contribute to NMR by investigating postulates desirable for the reasoning process in the context of argumentation. We demonstrate that those postulates are often violated and we propose the ASPICLite formalism satisfying them.

## 1.5 THESIS OVERVIEW

This thesis will be divided into chapters as follows:

- Chapter 2: Preliminaries

  We recall basic definitions of Dung's Abstract Argumentation Dung (1995) and extension-based and labelling-based semantics.

- Chapter 3: Quantifying Disagreement between Labellings: Postulates

  We identify different properties, which the *labelling distance method* should satisfy. We divide them into four groups: metric postulates, betweenness and qualitative distance, compositionality

[1] http://www.sintelnet.eu/content/arguing-web-20-0

and equivalence postulates. We investigate the relation between them.

- Chapter 4: Quantifying Disagreement between Labellings: Product Distance

  We define the family of distance methods which we call product distance methods. These are parametrised by selection function selecting a set of 'important' arguments and a distance measure between the set of labels. The distance between labellings is a sum of the distances between the labels over the selected arguments. We identify several subfamilies of product distance methods, starting with the *Full Sum* method which selects the set of all arguments and building towards *Issue-based* distance methods, which satisfy different postulates. We propose a distance method from each subfamily.

- Chapter 5: Using Distances for Aggregation in Abstract Argumentation

  Miller and Osherson (2009) proposed a framework for aggregation using distances in Judgement Aggregation. We adapt it to abstract argumentation setting and show how the distance methods defined in the previous chapter can be plugged in to define families of labelling aggregation operators. We thus illustrate the usefulness of these distance measures. We discuss agenda manipulation and link it to the *Indifference to Peripheral Issues* **(IPI)** postulate for distance method that we introduced earlier. Finally, we show that the existing aggregation operators due to Caminada and Pigozzi (2011) fit into this distance-based framework.

- Chapter 6: A Persuasion Dialogue for Grounded Semantics

  We define a dialogue game for grounded semantics which improves the standard grounded game Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009). We define acceptance of an argument in terms of the existence of the winning game while the standard game defines acceptance of an argument in terms of existence of winning strategy.

- Chapter 7: Implementing Crash-resistance and Non-interference in Logic-based Argumentation

  We discuss the abstract argumentation as used in the proccess on non-monotonic reasoning. We describe the postulates of Non-interference and Crash-resistance and show how these are violated by formalisms like ASPICLite, OSCAR Pollock (1995) and ASPIC Prakken (2010). Then we provide a general solution to satisfy both the postulates introduced in Caminada and Amgoud (2005, 2007) (Direct Consistency, Indirect Consistency and

Closure) and the additional postulates examined in the current work (Non-interference and Crash-resistance Caminada et al. (2012)).

- Chapter 8: Conclusions

  We summarize the results and list some ideas we would like to develop in the future.

The work presented in Chapter 3 and 4 is based in part on joint work with Richard Booth, Martin Caminada and Iyad Rhawan (2012). The work in Chapter 5 is based on joint work with Richard Booth (2014). The work in Chapter 6 is based on joint work with Martin Caminada (2012a; 2012b). The work in Chapter 7 is based on joint work with Yining Wu (2014).

# 2 PRELIMINARIES

## 2.1 ARGUMENTATION FRAMEWORK

We use the familiar setting of *abstract argumentation* Dung (1995). We start by assuming a countably infinite set $\mathcal{U}$ of argument names, from which all possible argumentation frameworks are built.

**Definition 1.** *An* argumentation framework *(AF for short)* $\mathcal{A} = (Args, \rightharpoonup)$ *is a pair consisting of a finite set* $Args \subseteq \mathcal{U}$ *of arguments and an attack relation* $\rightharpoonup \subseteq Args \times Args$. *We also use* $Args_{\mathcal{A}}$ *and* $\rightharpoonup_{\mathcal{A}}$ *to denote the arguments and attack relation of a given AF* $\mathcal{A}$.

We say that argument $a$ *attacks* argument $b$ iff $(a, b) \in \rightharpoonup$ and we write it as $a \rightharpoonup b$. An AF is a directed graph in which the arguments are represented as nodes and the attack relation is represented as arrows.

Sometimes we are interested in an argumentation framework restricted to the set of arguments.

**Definition 2.** *Let* $\mathcal{A} = (Args, \rightharpoonup)$ *be an AF and* $\mathcal{A}r \subset Args$. *The* restriction of $\mathcal{A}$ to $\mathcal{A}r$ *is a framework* $\mathcal{A}_{\downarrow \mathcal{A}r} = (Args \cap \mathcal{A}r, \rightharpoonup \cap \mathcal{A}r^2)$.

## 2.2 SEMANTICS

The major question in abstract argumentation is which arguments to accept given an AF encoding all the conflicts between them. It can be defined in several ways including extensions, labellings and dialogue games. We describe the first two here. The last one will be given in Chapter 6.

### 2.2.1 *Extension-based Semantics*

In essence, extension-based semantics is a function which for each AF returns a set of extensions. An extension is a set of arguments which can be accepted at the same time.

**Definition 3.** *An* extension-based semantics *is a function Sem assigning to each AF* $\mathcal{A}$ *a set of extensions* $Sem(\mathcal{A}) \subseteq 2^{Args}$.

The most common criteria for defining semantics are conflict-freeness and defence.

**Definition 4** (conflict-free / defence).

*Let $\mathcal{A} = (Args, \rightharpoonup)$ be an argumentation framework, $a \in Args$ and $\mathcal{A}r \subseteq Args$. We define*

$$a^+ \stackrel{\text{def}}{=} \{b \in Args \mid a \rightharpoonup b\},$$
$$a^- \stackrel{\text{def}}{=} \{b \in Args \mid b \rightharpoonup a\}.$$

*We extend this definitions to the set of arguments $\mathcal{A}r$*

$$\mathcal{A}r^+ \stackrel{\text{def}}{=} \{b \in Args \mid a \rightharpoonup b \textit{ for some } a \in \mathcal{A}r\},$$
$$\mathcal{A}r^- \stackrel{\text{def}}{=} \{b \in Args \mid b \rightharpoonup a \textit{ for some } a \in \mathcal{A}r\}.$$

$\mathcal{A}r$ *is* conflict-free *iff* $\mathcal{A}r \cap \mathcal{A}r^+ = \emptyset$. $\mathcal{A}r$ defends *an argument* $a$ *iff* $a^- \subseteq \mathcal{A}r^+$. *We define function* $F_{\mathcal{A}} : 2^{Args} \rightarrow 2^{Args}$ *as*

$$F_{\mathcal{A}}(\mathcal{A}r) = \{a \in Args \mid a \textit{ is defended by } \mathcal{A}r\}.$$

When only one argumentation framework is concerned, $F$ is used as the shortening of $F_{\mathcal{A}}$.

**Definition 5** (acceptability semantics)**.** *Let $(Args, \rightharpoonup)$ be an argumentation framework. A conflict-free set $\mathcal{A}r \subseteq Args$ is called*

- *an* admissible set *iff* $\mathcal{A}r \subseteq F(\mathcal{A}r)$.
- *a* complete extension *iff* $\mathcal{A}r = F(\mathcal{A}r)$.
- *a* grounded extension *iff $\mathcal{A}r$ is a minimal complete extension.*
- *a* preferred extension *iff $\mathcal{A}r$ is a maximal complete extension.*
- *a* stable extension *iff $\mathcal{A}r$ is a complete extension that defeats every argument in $Args \backslash \mathcal{A}r$.*
- *a* semi-stable extension *iff $\mathcal{A}r$ is a complete extension such that $\mathcal{A}r \cup \mathcal{A}r^+$ is maximal.*

### 2.2.2 *Labelling-based Semantics*

In the current work, we focus on the approach of Caminada (2006a); Caminada and Gabbay (2009) in which the semantics of abstract argumentation is expressed in terms of *argument labellings*. The idea is to distinguish between the arguments that one accepts (that are labelled `in`), the arguments that one rejects (that are labelled `out`) and the arguments which one abstains from having an opinion about (that are labelled `undec` for "undecided").

**Definition 6.** *Let $\mathcal{A} = (Args, \rightharpoonup)$ be an AF. An $\mathcal{A}$-*labelling *is a function* $L\colon Args \rightarrow \{\texttt{in}, \texttt{out}, \texttt{undec}\}$*. The set of all $\mathcal{A}$-labellings is denoted by $Labs(\mathcal{A})$. Given $\mathcal{A}r \subseteq Args$ we denote by $L[\mathcal{A}r]$ the restriction of $L$ to $\mathcal{A}r$.*

For notational purposes it is useful to define a unary "negation" operator on the set of labels by $\neg\mathtt{in} = \mathtt{out}$, $\neg\mathtt{out} = \mathtt{in}$ and $\neg\mathtt{undec} = \mathtt{undec}$.

Since a labelling is a function, it can be represented as a set of pairs, each consisting of an argument and a label (in, out, or undec). In addition, if L is a (partial) labelling, then we write $\mathtt{in}(L)$ for $\{a \mid L(a) = \mathtt{in}\}$, $\mathtt{out}(L)$ for $\{a \mid L(a) = \mathtt{out}\}$ and $\mathtt{undec}(L)$ for $\{a \mid L(a) = \mathtt{undec}\}$. Since a labelling can be seen as a partition of the set of arguments in the AF, we will sometimes write a labelling L as a triple $(\mathtt{in}(L), \mathtt{out}(L), \mathtt{undec}(L))$.

Of course a *rational* labelling should somehow respect the attack relation.

**Definition 7.** *A* labelling based semantics *is a function Sem assigning to each AF $\mathcal{A}$ a set of labellings Sem($\mathcal{A}$) $\subseteq \mathtt{Labs}(\mathcal{A})$.*

The most common example is the complete semantics.

**Definition 8.** *Let $\mathcal{A}$ be an AF and $L \in \mathtt{Labs}(\mathcal{A})$. For all arguments $a \in Args_{\mathcal{A}}$ we say:*

- *$a$ is legally* in *if* $L(a) = \mathtt{in}$ *and* $L(b) = \mathtt{out}$ *for all* $b \in Args_{\mathcal{A}}$ *such that* $b \rightharpoonup_{\mathcal{A}} a$,
- *$a$ is legally* out *if* $L(a) = \mathtt{out}$ *and* $L(b) = \mathtt{in}$ *for some* $b \in Args_{\mathcal{A}}$ *such that* $b \rightharpoonup_{\mathcal{A}} a$,
- *$a$ is legally* undec *if* $L(a) = \mathtt{undec}$ *and there is no* $b \in Args_{\mathcal{A}}$ *such that* $b \rightharpoonup_{\mathcal{A}} a$ *and* $L(b) = \mathtt{in}$, *and there exists* $c \in Args_{\mathcal{A}}$ *s.t.* $c \rightharpoonup_{\mathcal{A}} a$ *and* $L(c) = \mathtt{undec}$.

*Arguments which are labelled* in, out, undec *but are not legally* in, out, undec *we call illegally* in, out, undec *respectively.*

**Definition 9.** *Let $\mathcal{A}$ be an AF and $L \in \mathtt{Labs}(\mathcal{A})$. L is*

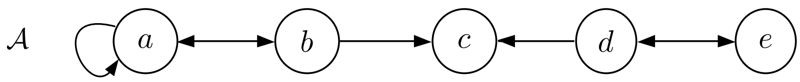
- *an* admissible labelling *iff it has no illegally* in *and no illegally* out *arguments.*
- *a* complete labelling *iff it has no illegally* in *and no illegally* out *and no illegally* undec *arguments. We denote the set of complete $\mathcal{A}$-labellings by* $\mathtt{Comp}(\mathcal{A})$.
- *a* grounded labelling *iff* L *is a complete labelling with minimal* $\mathtt{in}(L)$.
- *a* preferred labelling *iff* L *is a complete labelling with maximal* $\mathtt{in}(L)$.
- *a* stable labelling *iff* L *is a complete labelling with* $\mathtt{undec}(L) = \emptyset$.
- *a* semi-stable labelling *iff* L *is a complete labelling such that* $\mathtt{in}(L) \cup \mathtt{out}(L)$ *is maximal.*

Note that complete semantics is used as a basic criterion for other semantics. We call a semantics *complete-based* if its labellings are complete. The grounded, preferred, stable and semi-stable semantics as well as the complete semantics itself are complete-based.

As stated in Caminada (2006a); Caminada and Gabbay (2009), complete labellings coincide with complete extensions in the sense of Dung (1995). Moreover, the relationship between them is one-to-one. In essence, a complete extension is simply the in-labelled part of a complete labelling Caminada (2006a); Caminada and Gabbay (2009). Based on that, the correspondence between grounded, preferred, stable and semi-stable labellings and extensions is straightforward.

The set of complete $\mathcal{A}$-labellings is non-empty and usually contains more then one labelling.

In the rest of this work we identify rational $\mathcal{A}$-labellings with complete $\mathcal{A}$-labellings. This is because they form the basis for other semantics such as preferred, stable, semi-stable, etc (see Caminada and Gabbay (2009)). This choice is also in line with other works on aggregation in argumentation Booth et al. (2014); Caminada and Pigozzi (2011); Rahwan and Tohmé (2010). However, the results which do not depend on any particular semantics will assume some fixed *Sem*.




| *abcde* | *abcde* | *abcde* |
|---|---|---|
| $L_1$: oioio | $L_4$: uuoio | $N_1$: uuouu |
| $L_2$: oiooi | $L_5$: uuuoi | $N_2$: iouoi |
| $L_3$: oiouu | $L_6$: uuuuu | $N_3$: uuioi |

Figure 4: An argumentation framework with all its complete labellings ($L_1$-$L_6$), and three other labellings ($N_1$-$N_3$).

**Example 10.** *Figure 4 depicts argumentation framework $\mathcal{A}$ and its labellings. We represent labellings by a string of letters* i, u *and* o *corresponding to* in, undec *and* out *respectively. AF $\mathcal{A}$ has 6 complete labellings $L_1$-$L_6$. Between them labelling $L_6$ is a grounded labelling and $\emptyset$ is a corresponding grounded extension. The labellings $L_1$ and $L_2$ are stable, preferred and semi-stable. The sets $\{b, d\}$ and $\{b, e\}$ are the corresponding extensions.*

*The labellings $N_1$-$N_3$ are not complete because the argument* c *is respectively illegally* out, *illegally* undec *and illegally* in.

# 3 QUANTIFYING DISAGREEMENT BETWEEN LABELLINGS: POSTULATES

## 3.1 INTRODUCTION

Given a conflicting logical theory, an agent is faced with the problem of deciding what it could reasonably believe. The presence of multiple reasonable positions raises a fundamental question: *how different are two given evaluations of a conflicting logical theory?* We attempt to answer this question in the context of abstract argumentation theory.

This question is relevant to two fundamental problems. The first problem is *argument-based belief revision*. Suppose a diplomat receives instructions to switch his position on one particular argument. To maintain a consistent viewpoint, the diplomat must revise his evaluation of other related arguments. Faced with multiple possibilities, the diplomat may wish to choose the one that differs the least from his initial position (e.g. to maintain credibility).

The issue of distance is also relevant to the problem of *judgement aggregation* over how a given set of arguments should be evaluated collectively by a group of agents with different opinions, Caminada and Pigozzi (2011); Caminada et al. (2011); Rahwan and Tohmé (2010). For instance it is very well possible that the members of a jury in a criminal trial all share the same information on the case (and hence have the same argumentation framework) but still have different opinions on what the verdict should be. Hence, these differences of opinion are consequences not of differences in the knowledge base but of the nature of nonmonotonic reasoning, which allows for various reasonable positions (extensions). In the context of judgement aggregation one may examine the extent to which the collective position differs from the various positions of the individual participants. Ideally, one would like to have a collective position that is closest to the collection of individual positions, for example such that the sum of its distance to each individual position is minimal.

In this chapter, we examine a number of possible candidates for measuring the *distance* between different labellings (evaluations) of an argumentation graph. Our work advances the state-of-the-art in argument-based reasoning in three ways: firstly we provide the first systematic investigation of quantifying the distance between two evaluations of an argument graph; secondly We examine a number of intuitive measures and show that they fail to satisfy basic desirable postulates; finally we come up with a measure that satisfies them all. In addition to providing many answers, our work also raises many interesting questions to the community at the intersection between argumentation and social choice.

## 3.2 DISTANCE BETWEEN LABELLINGS

The problem we are addressing is the following:

> Given an AF $\mathcal{A}$, and given two *Sem*-labellings S (the *source* labelling) and T (the *target* labelling) over $\mathcal{A}$, how can we quantify the *distance* from S to T, denoted $d(S, T)$?

Of course we do not just want a method which applies to only one AF, we want a method to be able to do this for *any* given $\mathcal{A}$.

**Definition 11.** *Let $\mathcal{A}$ be an AF and Sem some fixed semantics. A Sem-*labelling distance measure (for AF $\mathcal{A}$) *is a function* $d_{\mathcal{A}} : Sem(\mathcal{A}) \times Sem(\mathcal{A}) \to \mathbb{R}$. *A Sem-*labelling distance method $d$ *is a function that associates a distance measure* $d_{\mathcal{A}}$ *to each AF $\mathcal{A}$.*

Sometimes we will leave out *Sem* from *Sem-labelling distance method* and write simply *labelling distance method* instead but we always assume for each AF, a designated set of labellings, which represent agents' reasonable positions, returned by some semantics *Sem*. In Chapter 5 we consider the case which requires us to extend our distances to the set of all labellings, therefore we define it as such already here. It is possible that such extension can be different for different semantics. In fact we deem it desirable to include the information contained in the semantics into definition of the distance methods.

In the rest of this chapter we identify different properties, which should be satisfied by *labelling distance methods* solving the problem stated above. In the next chapter we analyse the family of product distance methods and identify the criteria under which proposed properties hold. Finally, in Chapter 5 we apply such constructed distances to the problem of judgement aggregation.

## 3.3 POSTULATES FOR DISTANCE METHODS

We split the proposed postulates in four different groups. Group I and II contain postulates for distance measure rather than distance method. We assume universal quantification i.e. distance method $d$ satisfy **(x)** if and only if distance measure $d_{\mathcal{A}}$ assigned by $d$ satisfy **(x)** for all AFs $\mathcal{A}$. Since $\mathcal{A}$ is fixed in these groups we will skip index and write $d$ instead of $d_{\mathcal{A}}$. Group III and IV contain postulates of distance methods relating distance measures assigned to different AFs (and properly indexed).

### 3.3.1 *Group I - Metric Postulates*

In mathematics, when formalising the notion of distance it is common to require that $d$ be a *metric*.

**Definition 12.** *A* metric *on a set* $X$ *is a function* $d : X \times X \to \mathbb{R}^+$ *such that for all* $x, y, z \in X$*:*

| | | |
|---|---|---|
| **(REF)** | $d(x, x) = 0$ | *(Reflexivity)* |
| **(DD)** | *if* $d(x, y) > 0$ *then* $x \neq y$, | *(Dissimilarity of the Diverse)* |
| **(SYM)** | $d(x, y) = d(y, x)$, | *(Symmetry)* |
| **(TRI)** | $d(x, z) \leqslant d(x, y) + d(y, z)$. | *(Triangle inequality)* |

*If* $d$ *satisfies all the above except, possibly,* **(DD)**, *then it is called a* pseudo-metric. *If it satisfies all except, possibly,* **(TRI)** *then it is a* semi-metric.

It can be discussed if distance between agents' positions should be a metric. Especially symmetry and triangle inequality can be questioned. For example, regarding symmetry convincing a sceptical agent to accept or reject an argument can be much harder than raising his doubts. It is not clear how to interpret triangle inequality, which is a spacial property, in case of labelling. Nevertheless we investigate those properties as an important reference without claiming that they should be satisfied.

On the other hand, the above conditions can be easily satisfied. One could possibly take *discrete metric*

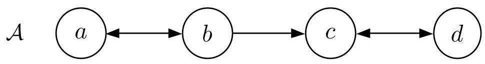
$$DM(x, y) = \begin{cases} 0 & x = y, \\ 1 & x \neq y \end{cases}$$

which works for any set of elements. Obviously we are interested in something more specific to the set of labellings, structure of AF and used semantics.

### 3.3.2 *Group II - Intuition-based Postulates*

Someone who agrees that *discrete metric* is not enough should be able to answer the following question. For which labellings $S, T_1, T_2$ we would like to have $d(S, T_1) < d(S, T_2)$? In this subsection we formalise two intuitions such an answer can be based on. Let us start with an example.

Imagine a problem described by the AF $\mathcal{A}$ in Figure 5 with $Sem = Comp$ and an agent switching its position represented by labelling $L_1$ to the position represented by labelling $L_3$. Now consider the labelling $L_2$. In this situation it may seem reasonable to have $d(L_1, L_2) < d(L_1, L_3)$ for two reasons:




| abcd | abcd |
|---|---|
| $L_1$: ioio | $L_4$: iouu |
| $L_2$: iooi | $L_5$: uuuu |
| $L_3$: oioi | $L_6$: uuoi |

Figure 5: An argumentation framework with all its six complete labellings.

CHANGE Agent updates evaluation of all the four arguments. The position represented by $L_2$ looks like a middle step in which agent has already changed his opinion about arguments $c, d$ but still has not revised arguments $a, b$. Since $L_2$ is between $L_1$ and $L_3$ changing from $L_1$ to $L_3$ requires two steps when changing from $L_1$ to $L_2$ requires just the first one. Although steps can be of a different sizes we concentrate here on the process. Agent changes its position as a whole and what matters are the positions it needs to pass 'on the way' which depend on the set of feasible labellings. This set in turn may differ for different semantics and/or AFs.

DISAGREEMENT The disagreement between $L_1$ and $L_2$ consist of two arguments $c, d$ while the disagreement between $L_1$ and $L_3$ consists of four arguments $a, b, c, d$. The second one strictly contains the first one therefore is bigger. We define here what is the disagreement between two labellings and try to quantify its size. The disagreement depends solely on the labellings.

We investigate both intuitions, which are slightly different and connect with different expectations, and later on give our proposal for some concrete definitions.

For now, assume one can distinguish some labellings as being between the others and define it as a ternary relation. Let us denote the fact that labelling $L_2$ is between labellings $L_1$ and $L_3$ by $L_1 \triangleright L_2 \triangleleft L_3$. If moreover $L_2 \neq L_1$ and $L_2 \neq L_3$ we will write $L_1 \triangleright\triangleright L_2 \triangleleft L_3$ and $L_1 \triangleright L_2 \triangleleft\triangleleft L_3$ respectively. Regardless of the particular choice of betweenness relation $\triangleright$, the following might seem to be a reasonable requirement on a distance function $d$:

**(BTM$_\triangleright$)** If $L_1 \triangleright L_2 \triangleleft\triangleleft L_3$ then $d(L_1, L_2) < d(L_1, L_3)$

(Betweenness monotonicity)

The above postulate is saying that if the labelling $L_2$ is strictly between labellings $L_1$ and $L_3$, then the distance between labellings $L_1$ and $L_3$ should be strictly bigger than the distance between labellings $L_1$ and $L_2$.

Similarly, assume one can define disagreement between labellings $\ominus$ and partially order it. Let us denote the fact that the disagreement between labellings $L_1$ and $L_2$ is smaller than or equal to the disagreement between labellings $L_3$ and $L_4$ by $(L_1 \ominus L_2) \subseteq (L_3 \ominus L_4)$. For instance in the above example the qualitative distance is simply the set of differently labelled arguments ordered by set inclusion ($\{c, d\} \subseteq \{a, b, c, d\}$). Again, regardless of the particular choice of qualitative distance $\ominus$, it might seem to be a reasonable requirement on a distance function $d$ to respect this partial order.

**(QDA$_\ominus$)** If $(L_1 \ominus L_2) \subseteq (L_3 \ominus L_4)$ then $d(L_1, L_2) \leqslant d(L_3, L_4)$
(Qualitative Distance Alignment)

**(SQDA$_\ominus$)** If $(L_1 \ominus L_2) \subset (L_3 \ominus L_4)$ then $d(L_1, L_2) < d(L_3, L_4)$
(Strict Qualitative Distance Alignment)

**(QDA)** implies that the distance between the labellings with the same qualitative distance needs to be the same.

**(EQDA$_\ominus$)** If $(L_1 \ominus L_2) = (L_3 \ominus L_4)$ then $d(L_1, L_2) = d(L_3, L_4)$

**(SQDA)** postulates strict increase in qualitative distance should entail strict increase in quantitative distance.

In the rest of the subsection we will propose two concrete betweenness relation and two qualitative distances, and study the link between them.

*Examples of Betweenness*

**Definition 13** (Simple Betweenness). *Let $\mathcal{A}$ be an AF. We define* simple betweenness *relation for any* $L_1, L_2, L_3 \in$ *Labs*$(\mathcal{A})$ *as follows*

$$L_1 \triangleright L_2 \triangleleft L_3 \textit{ iff } \forall a \in \textit{Args}_{\mathcal{A}} \, (L_2(a) = L_1(a) \vee L_2(a) = L_3(a)) .$$

$L_1 \triangleright L_2 \triangleleft L_3$ means that $L_2$ labels every argument either the same as $L_1$ or the same as $L_3$. In other words if agent switches from the labelling $L_1$ into $L_3$ every argument that $L_2$ labels differently from $L_1$, is labelled equally differently by $L_3$. Thus $L_3$ differs from $L_1$ at least as much as $L_2$ does.

We can check that for AF $\mathcal{A}$ from Figure 5 we have $L_1 \triangleright L_2 \triangleleft L_3$. We can also check that we do not have $L_1 \triangleright L_4 \triangleleft L_2$. Is this latter one natural? It may seem unnatural because one may see `undec` label as a value between `in` and `out`. If we commit to this special status of `undec` label we shall refine the previous betweenness relation as follows.

**Definition 14** (Refined Betweenness). *Let $\mathcal{A}$ be an AF. We define* refined betweenness *relation for any* $L_1, L_2, L_3 \in$ *Labs*$(\mathcal{A})$ *as follows*

$$L_1 \blacktriangleright L_2 \blacktriangleleft L_3 \textit{ iff } \forall a \in \textit{Args}_{\mathcal{A}} \begin{pmatrix} L_2(a) = L_1(a) \vee L_2(a) = L_3(a) \\ \vee [L_2(a) = \texttt{undec} \wedge L_1(a) \neq L_3(a)] \end{pmatrix} .$$

$L_1 \blacktriangleright L_2 \blacktriangleleft L_3$ is merely expressing that, for all $a \in$ *Args*, $L_2(a)$ lies on a path *between* $L_1(a)$ and $L_3(a)$, assuming the natural neighbourhood graph `in` – `undec` – `out` over the labels.

The above betweenness relations lead to the two versions of betweenness monotonicity postulate - **($BTM_{\triangleright}$)** and **($BTM_{\blacktriangleright}$)**. In our previous work we studied them as two separate postulates *Disagreement monotonicity* and *Betweenness monotonicity* respectively, Booth et al. (2012).

We finish by an observation.

**Observation 15.** *Since any labellings which are between as defined by simple betweenness are also between as defined by refined betweenness, it holds that* **($BTM_{\blacktriangleright}$)** *implies* **($BTM_{\triangleright}$)***.*

*Examples of Qualitative Distances*

To define notion of qualitative distance it is useful to define types of conflicts.

**Definition 16** (Types of conflict)**.** *Let $\mathcal{A}$ be an AF. For any $L_1, L_2 \in Labs(\mathcal{A})$ we define the following sets:*

CONFLICTS $C(L_1, L_2) = \{a \in Args_{\mathcal{A}} \mid L_1(a) \neq L_2(a)\}$,

HARD CONFLICTS $H(L_1, L_2) = \{a \in Args_{\mathcal{A}} \mid L_1(a) = \neg L_2(a) \neq \texttt{undec}\}$,

SOFT CONFLICTS $S(L_1, L_2) = C(L_1, L_2) \setminus H(L_1, L_2)$.

*We call members of those sets* conflicts, hard conflicts *and* soft conflicts *between labellings $L_1$ and $L_2$. For $a \in C(L_1, L_2)$ (respectively $H(L_1, L_2)$ and $S(L_1, L_2)$) we say also there is a conflict (respectively hard conflict and soft conflict) on argument $a$ (between labellings $L_1$ and $L_2$).*

Note that by definition the set of conflicts $C(L_1, L_2)$ is partitioned into soft and hard conflicts.

We define two types of qualitative distance.

**Definition 17** (Hamming set, Refined Hamming pair)**.** *Let $L_1, L_2$ be two labellings. We define two qualitative distances:*

HAMMING SET $(L_1 \ominus_{HS} L_2) = C(L_1, L_2)$,

REFINED HAMMING PAIR $(L_1 \ominus_{RHP} L_2) = (C(L_1, L_2), H(L_1, L_2))$,

*with the natural order given by set inclusion with the bottom element $\emptyset$ (Hamming Set) and set inclusion extended to pairs in the following way $(A, B) \subseteq (X, Y)$ iff $A \subseteq X$ and $B \subseteq Y$ and $(A, B) \subsetneq (X, Y)$ iff $(A, B) \subseteq (X, Y)$ and $A \subsetneq X$ or $B \subsetneq Y$, with the bottom element $(\emptyset, \emptyset)$ (Refined Hamming Pair).*

The Hamming Set defines disagreement between two labellings as a set of arguments which are labelled differently. The Refined Hamming pair additionally distinguishes hard and soft conflicts and imposes that hard conflict is stronger than soft conflict.

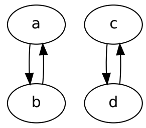



Figure 6: Two pairs of mutually attacking arguments

The two qualitative distances leads to the postulates **(QDA$_{\ominus}$)** and **(SQDA$_{\ominus}$)** in two variants. What is the relation between those variants? Let us investigate the partial orderings between $Labs(\mathcal{A})^2$ induced by partial ordering between qualitative distances. The partial order induced by Hamming set extends the one induced by Refined Hamming Pair, i.e., $(L_1 \ominus_{RHP} L_2) \subseteq (L_3 \ominus_{RHP} L_4)$ implies $(L_1 \ominus_{HS} L_2) \subseteq (L_3 \ominus_{HS} L_4)$. It is because Hamming set compares just the sets of conflicting arguments while Refined Hamming Pair compares additionally sets of hard conflicts. It means that the scope of **(QDA$_{\ominus_{HS}}$)** is bigger than **(QDA$_{\ominus_{RHP}}$)** and therefore **(QDA$_{\ominus_{HS}}$)** implies **(QDA$_{\ominus_{RHP}}$)**.

**Proposition 18.** *Any distance method* $d$ *which satisfies* **(QDA$_{\ominus_{HS}}$)**, *satisfies* **(QDA$_{\ominus_{RHP}}$)**.

*Proof.* It follows from the fact that $(L_1 \ominus_{RHP} L_2) \subseteq (L_3 \ominus_{RHP} L_4)$ implies $(L_1 \ominus_{HS} L_2) \subseteq (L_3 \ominus_{HS} L_4)$ for all labellings $L_1, ..., L_4$. □

The following example shows that neither **(SQDA$_{\ominus_{HS}}$)** implies **(SQDA$_{\ominus_{RHP}}$)** nor the reverse. Moreover **(SQDA$_{\ominus_{RHP}}$)** is incompatible with **(QDA$_{\ominus_{HS}}$)**.

**Example 19.** *Let* $\mathcal{A} = (\{a, b, c, d\}, \{(a, b), (b, a), (c, d), (d, c)\})$ *containing two pairs of mutually attacking arguments depicted in Figure 6. Consider complete labellings* $L_1$: `ioio`, $L_2$: `oiio`, $L_3$: `uuuu`, $L_4$: `oiuu`.

*First notice that* $(L_1 \ominus_{HS} L_2) = \{a, b\} \subsetneq \{a, b, c, d\} = (L_1 \ominus_{HS} L_3)$ *but* $(L_1 \ominus_{RHP} L_2) = (\{a, b\}, \{a, b\}) \not\subseteq (\{a, b, c, d\}, \emptyset) = (L_1 \ominus_{HP} L_3)$. *In this case the set of conflicting arguments increased but 'strength' of the conflict decreased which makes it incomparable when Refined Hamming Pair is used but it is comparable when Hamming set, which ignores the 'strength' of the conflict, is used. Hence, in this case, satisfaction of* **(SQDA$_{\ominus_{HS}}$)** *forces any distance method* $d$ *to increase but satisfaction of* **(SQDA$_{\ominus_{RHP}}$)** *does not pose this restriction.*

*Second notice that* $(L_1 \ominus_{RHP} L_3) = (\{a, b, c, d\}, \emptyset) \subsetneq (\{a, b, c, d\}, \{a, b\}) = (L_1 \ominus_{RHP} L_4)$ *but* $(L_1 \ominus_{HS} L_3) = \{a, b, c, d\} = (L_1 \ominus_{HS} L_4)$. *In this case a set of conflicting arguments stays the same but 'strength' of the conflict increases. In this case satisfaction of* **(SQDA$_{\ominus_{RHP}}$)** *forces any distance method* $d$ *to increase, but satisfaction of* **(SQDA$_{\ominus_{HS}}$)** *allows also to stay the same. Moreover satisfaction of* **(QDA$_{\ominus_{HS}}$)** *forces* $d$ *to stay the same which cannot be at the same time.*

We switch now to the question what is the relation between different variants of Betweenness monotonicity and Strict Qualitative Distance Alignment.

**Lemma 20.** *Any distance method* $d$ *which satisfies* **(SQDA$_{\ominus_{HS}}$)** *satisfies* **(BTM$_{\triangleright}$)**. *Any distance method* $d$ *which satisfies* **(SQDA$_{\ominus_{RHP}}$)** *satisfies* **(BTM$_{\blacktriangleright}$)** *(and hence* **(BTM$_{\triangleright}$)** *Observation 15).*

*Proof.* It follows from the fact that $L_1 \triangleright L_2 \triangleleft\triangleleft L_3$ and $L_1 \blacktriangleright L_2 \blacktriangleleft\blacktriangleleft L_3$ implies $(L_1 \ominus_{HS} L_2) \subset (L_1 \ominus_{HS} L_3)$ and $(L_1 \ominus_{RHP} L_2) \subset (L_1 \ominus_{RHP} L_3)$ respectively.

Indeed, consider $a \in (L_1 \ominus_{HS} L_2)$ then $L_1(a) \neq L_2(a)$ and by definition of $\triangleright$ we have $L_2(a) = L_3(a)$ so $L_1(a) \neq L_3(a)$ and so $a \in (L_1 \ominus_{HS} L_3)$. It cannot be the case $(L_1 \ominus_{HS} L_2) = (L_1 \ominus_{HS} L_3)$ since $L_2 \neq L_3$.

Similarly, consider $a \in H(L_1, L_2)$, then $L_1(a) \neq L_2(a)$ and $L_2(a) \neq \mathtt{undec}$. By definition of $\blacktriangleright$ we have $L_2(a) = L_3(a)$ so $a \in H(L_1, L_3)$. Now, consider $a \in S(L_1, L_2)$, then $L_1(a) \neq L_2(a)$ and one of the two cases holds:

CASE 1: $L_1(a) = \mathtt{undec}, L_2(a) \neq \mathtt{undec}$
Then $L_2(a) = L_3(a)$ and $a \in S(L_1, L_3)$.

CASE 2: $L_1(a) \neq \mathtt{undec}, L_2(a) = \mathtt{undec}$
Then $L_1(a) \neq L_3(a)$ so $a \in S(L_1, L_3)$ (and possibly also $a \in H(L_1, L_3)$).

It cannot be the case $(L_1 \ominus_{RHP} L_2) = (L_1 \ominus_{RHP} L_3)$ since $L_2 \neq L_3$. □

We should note that **(BTM$_{\blacktriangleright}$)** is not implied by **(SQDA$_{\ominus_{HS}}$)**. Consider a 2-loop. we have $\mathtt{io} \blacktriangleright \mathtt{uu} \blacktriangleleft\blacktriangleleft \mathtt{oi}$ so **(BTM$_{\blacktriangleright}$)** requires $d(\mathtt{io}, \mathtt{uu}) < d(\mathtt{io}, \mathtt{oi})$ which is not required by **(SQDA$_{\ominus_{HS}}$)** since $(\mathtt{io} \ominus_{HS} \mathtt{uu}) = (\mathtt{io} \ominus_{HS} \mathtt{oi})$.

The above Lemma 20 shows that **(SQDA)** is at least as strong as the corresponding **(BTM)**. In fact it is strictly stronger because of two important differences.

First, Betweenness Monotonicity talks about the distance of two target labellings from a common source labelling, while Qualitative Distance Alignment constrains distances between pairs of different labellings. This difference will be exploited in Example 64 in the next chapter.

Second, for any three labellings $L_1, L_2, L_3$ we have that $L_1 \triangleright L_2 \triangleleft\triangleleft L_3$ (respectively $L_1 \blacktriangleright L_2 \blacktriangleleft\blacktriangleleft L_3$) implies $(L_1 \ominus_{HP} L_2) \subsetneq (L_1 \ominus_{HP} L_3)$ (respectively $(L_1 \ominus_{RHP} L_2) \subsetneq (L_1 \ominus_{RHP} L_3)$) (the proof of the Lemma 20 is based on that fact) but the reverse is not true which is illustrated in the following example.

**Example 21.** *Let us recall the framework from Example 19 consisting of two pairs of mutually attacking arguments depicted in Figure 6.*

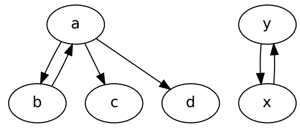



Figure 7: AF with two independent parts.

*Consider complete labellings* $L_1$ : `uuoi`, $L_2$ : `oioi`, $L_3$ : `ioio`. *We have* $(L_1 \ominus L_2) \subsetneq (L_1 \ominus L_3)$ *for* $\ominus \in \{\ominus_{HS}, \ominus_{RHP}\}$. *Therefore, both versions of* **(SQDA)** *require* $d(L_1, L_2) < d(L_1, L_3)$ *which is not required by either of* **(BTM)** *postulates because it is not the case that* $L_1 \blacktriangleright L_2 \blacktriangleleft\blacktriangleleft L_3$ *and hence also not* $L_1 \triangleright L_2 \triangleleft\triangleleft L_3$.

We finish this subsection with an observation linking the postulates from this subsection with the metric postulates from the previous subsection.

**Observation 22.** *Any variant of* **(SQDA)**, **(BTM)** *considered in this subsection together with* **(REF)** *implies* **(DD)**.

This is because for any labellings $L_1, L_2, L_3$, $L_1 \neq L_2$ it holds that $L_1 \triangleright L_1 \triangleleft\triangleleft L_2$ and $(L_1 \ominus L_1) \subsetneq (L_2 \ominus L_3)$.

### 3.3.3 *Group III - Compositionality Postulates*

In the previous subsection we proposed the postulates of distance measure expressing some possible intuitions that particular distances should be bigger than other. They rule out some simple metrics like discrete metric. Another motivation for using more sophisticated measures is provided by postulates of distance method formulated in this subsection. The idea is that for AFs composed from two independent parts, the distance between the labellings of the whole should be determined by the distances between corresponding labellings of the parts. We will consider here taking sum of the distances (although max or weighted sum is another possibility) and restriction of the labelling to the part. What we need is to specify the independence relations. Let us start with an example explaining that idea.

**Example 23.** *Consider argumentation framework* $\mathcal{A}$ *depicted in Figure 7* ($\mathcal{A} = \{\{a, b, c, d, x, y\}, \{(a, b)(b, a)(a, c)(a, d), (x, y), (y, x)\}\}$). *It can be divided into two independent parts* $\mathcal{A}_2 = \{\{x, y\}, \{(x, y), (y, x)\}\}$ *and* $\mathcal{A}_1 = \{\{a, b, c, d\}, \{(a, b)(b, a)(a, c)(a, d)\}\}$. *We consider them independent for two reasons. One, considering AF structure,* $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *are two disconnected components of* $\mathcal{A}$. *Two, taking the complete labellings as the set of feasible labellings, any labelling* $L \in Comp(\mathcal{A})$ *can be decomposed into union of* $L_1 \cup L_2$ *where* $L_1 \in Comp(\mathcal{A}_1)$,$L_2 \in Comp(\mathcal{A}_2)$.*Furthermore any union*

$L_1 \cup L_2$ *forms a labelling of* $\mathcal{A}$. *Therefore we can expect* $d_{\mathcal{A}}(L_1 \cup L_2, L_3 \cup L_4) = d_{\mathcal{A}_1}(L_1, L_3) + d_{\mathcal{A}_2}(L_2, L_4)$.

First, we define a partition into independent sets of arguments.

**Definition 24** (Independent partition)**.** *Let* $\mathcal{A} = (Args, \rightharpoonup)$ *be an AF and* $A_1, A_2$ *a partition of its arguments, i.e.* $A_1 \cup A_2 = Args, A_1 \cap A_2 = \emptyset$. *The partition* $A_1, A_2$ *is:*

SYNTACTICALLY INDEPENDENT *iff* $A_1^+ \cap A_2 = A_1 \cap A_2^+ = \emptyset$,

SEMANTICALLY INDEPENDENT *under semantics Sem iff* $Sem(\mathcal{A}) = Sem(\mathcal{A}_{\downharpoonright A_1}) \otimes Sem(\mathcal{A}_{\downharpoonright A_2})$,

*where* $\otimes$ *is defined for any two sets of labellings* $\mathtt{Labs}_1, \mathtt{Labs}_2$ *with disjoint domains as follows*

$$\mathtt{Labs}_1 \otimes \mathtt{Labs}_2 = \{L_1 \cup L_2 \mid L_1 \in \mathtt{Labs}_1, L_2 \in \mathtt{Labs}_2\}.$$

The syntactic independence is defined purely by graph topology while the semantics independence is defined for a specific semantics. In general the following principle holds for the semantics we consider in this thesis:

**Definition 25** (Independence principle)**.** *The labelling based semantics Sem satisfies the* independence *principle if and only if syntactic independence implies semantics independence under semantics Sem.*

*Draft.* We check the implication for the main semantics. (admissible / complete) From syntactic independence follows that arguments neighbourhoods are preserved. It means that conflict free arguments stay conflict free and legally labelled arguments stay legal. (stable) Additionally, if there are no `undec` labelled arguments in the whole then there are no `undec` labelled arguments in the parts and vice versa. (preferred / semi-stable and other set minimisation / maximisation semantics) Proof by contradiction, assume the set is not maximal and show it leads to the labelling that is not maximal in the part / whole. □

Note that Independence principle is independent from Directionality principle, Baroni et al. (2011). Independence is not stronger since we have just shown that stable semantics satisfy it, while it does not satisfy Directionality. It is also not weaker what can be seen in the following 'synthetic' example. Let $A_1, A_2$ syntactically independent partition of $\mathcal{A}$ and semantic *Sem* such that $Sem(\mathcal{A}_{\downharpoonright A_1}) = \{L_1^1, L_2^1\}, Sem(\mathcal{A}_{\downharpoonright A_2}) = \{L_1^2, L_2^2\}, Sem(\mathcal{A}) = \{L_1^1 \cup L_1^2, L_2^1 \cup L_1^2, L_2^1 \cup L_2^2\}$. We have $Sem(\mathcal{A}_{\downharpoonright A_1}) = Sem(\mathcal{A})[A_1]$ and $Sem(\mathcal{A}_{\downharpoonright A_2}) = Sem(\mathcal{A})[A_2]$, therefore *Sem* satisfy Directionality principle, but doesn't satisfy Independence.

Semantics independence under semantics *Sem* usually doesn't imply syntactic independence as seen in the Example 26.

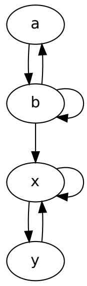



Figure 8: AF with semantically but not syntactically independent partition

**Example 26.** *Consider argumentation framework $\mathcal{A}$ depicted in Figure 8 ($\mathcal{A} = (\{a, b, x, y\}, \{(a, b), (b, a), (b, b), (b, x), (x, x), (x, y), (y, x)\})$). For the partition of arguments into $A = \{a, b\}$ and $B = \{x, y\}$ we have $\mathrm{Comp}\mathcal{A} = \mathrm{Comp}\mathcal{A}_{\downarrow A} \otimes \mathrm{Comp}\mathcal{A}_{\downarrow B}$ but $\mathcal{A}_{\downarrow A}$ and $\mathcal{A}_{\downarrow B}$ are not the disconnected components of $\mathcal{A}$ therefore $A, B$ are semantically independent under complete semantics (and more generally under complete based semantics) but are not syntactically independent.*

The two types of independence lead to the following two postulates.

**(COM)** if $A, B$ is a syntactically independent partition of $\mathcal{A}$ then
$$d_{\mathcal{A}}(L_1, L_2) = d_{\mathcal{A}_{\downarrow A}}(L_1[A], L_2[A]) + d_{\mathcal{A}_{\downarrow B}}(L_1[B], L_2[B]).$$
(Syntactic Compositionality)

**(COM$_{Sem}$)** if A, B is semantically independent partition of $\mathcal{A}$ under semantics *Sem* then
$$d_{\mathcal{A}}(L_1, L_2) = d_{\mathcal{A}_{\downarrow A}}(L_1[A], L_2[A]) + d_{\mathcal{A}_{\downarrow B}}(L_1[B], L_2[B]).$$
(Semantic Compositionality)

The first postulate, although it doesn't mention any particular semantics, requires that $Sem(\mathcal{A})[A] \subseteq Sem(\mathcal{A}_{\downarrow A})$. This is a weaker condition than the non-interference, Baroni et al. (2011) which requires that $Sem(\mathcal{A})[A] = Sem(\mathcal{A}_{\downarrow A})$. Our condition holds even for the stable semantics, the only semantics among popular semantics considered by Baroni et al. which fails non-interference.

Since all semantics we are interested in satisfy independence, the second postulate implies the first one. In the next subsection we introduce **(LAB$_{Sem}$)** postulate and prove (Lemma 33) that for all distance methods satisfying **(LAB$_{Sem}$)** both postulates are equivalent.

### 3.3.4 *Group IV - Equivalence Postulates*

The postulates in the two previous subsections describes the cases when distance between labellings should differ. The postulates in this subsection set some limits to that. As in the previous subsection the postulates in this subsection relate distance between labellings over different argumentation frameworks, so are again a properties of the *distance method*, i.e., the mapping $\mathcal{A} \mapsto d_{\mathcal{A}}$ (Definition 11).

The next distance properties we propose come from symmetry considerations. The idea is that applying the distance measure over AFs which are in some sense *equivalent* should yield equivalent results. In the context of argumentation semantics, such a property has been referred to as the language independence principle by Baroni and Giacomin (2007). We are interested in describing a similar property in the context of distance measures.

We begin with the common idea of graph-isomorphism, applied to argumentation frameworks.

**Definition 27.** *Let* $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ *and* $\mathcal{A}_2 = (Args_2, \rightharpoonup_2)$ *be two AFs. An* isomorphism from $\mathcal{A}_1$ to $\mathcal{A}_2$ *is any bijection* $g : Args_1 \to Args_2$ *such that, for all* $a, b \in Args_1$*,* $a \rightharpoonup_1 b$ *iff* $g(a) \rightharpoonup_2 g(b)$*. In the special case when* $\mathcal{A}_1 = \mathcal{A}_2$ *we call* $g$ *an* automorphism.

So basically an isomorphism just changes the names of arguments – or in the case of automorphism permutes them – while preserving the attack structure. Of course if $g$ is an isomorphism from $\mathcal{A}_1$ to $\mathcal{A}_2$ then $g^{-1}$ is an isomorphism from $\mathcal{A}_2$ to $\mathcal{A}_1$.

Any function (not just isomorphism) between arguments corresponds to a function between labellings.

**Definition 28** (Pullback labelling)**.** *Let* $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ *and* $\mathcal{A}_2 = (Args_2, \rightharpoonup_2)$ *be two AFs. For any mapping between the arguments* $f \colon Args_1 \to Args_2$ *we define* pullback *mapping between labellings* $\overline{f} \colon Labs(\mathcal{A}_2) \to Labs(\mathcal{A}_1)$ *given by formula* $[\overline{f}(L)](a) = L(f(a))$*.*

If $g$ is an isomorphism from $\mathcal{A}_1$ to $\mathcal{A}_2$ then we will abuse the notation extending $g$ to the set of labellings writing $g$ instead of $\overline{g^{-1}}$.

The following property says that the distance should be the same for isomorphic AFs. This is in line with the intuition that an argument is characterised completely by its interactions with the other arguments.

**(ISO)** If $g$ is an isomorphism from $\mathcal{A}_1$ to $\mathcal{A}_2$ then
$$d_{\mathcal{A}_1}(S, T) = d_{\mathcal{A}_2}(g(S), g(T))$$

In the case of automorphism we get the special case:

**(AUTO)** If $g$ is an automorphism on $\mathcal{A}$ then
$$d_{\mathcal{A}}(S, T) = d_{\mathcal{A}}(g(S), g(T))$$

All known semantics satisfy the language independence principle Baroni et al. (2011) which can be reformulated as follows:

**Definition 29** (Language Independence)**.** *Let* $g$ *be an isomorphism from* $\mathcal{A}_1$ *to* $\mathcal{A}_2$*. The semantics Sem satisfies* language independence *iff* $g(Sem(\mathcal{A}_1)) = Sem(\mathcal{A}_2)$*.*

The other way to define equivalence is in terms of labellings.

**Definition 30.** *Two AFs* $\mathcal{A}_1, \mathcal{A}_2$ *are* labelling equivalent *(under semantics Sem) via bijection* $g\colon Args_{\mathcal{A}_1} \to Args_{\mathcal{A}_2}$ *if* $g$ *extended to the set of labellings is a bijection between* $Sem(\mathcal{A}_1)$ *and* $Sem(\mathcal{A}_2)$*.*

$\mathcal{A}_1$ and $\mathcal{A}_2$ are labelling equivalent under *Sem* means essentially that they have the same set of *Sem*-labellings (up to possibly some renaming of the arguments).

**Example 31.** *Consider* $\mathcal{A}_1 = (\{a, b, c\}, \{(a, b), (b, c), (b, a)\})$ *and* $\mathcal{A}_2 = (\{a, b, c\}, \{(a, b), (b, c), (c, b), (b, a)\})$*. Complete semantics prescribes to them the same set of labellings* $Comp(\mathcal{A}_1) = \{\mathtt{ioi}, \mathtt{uuu}, \mathtt{oio}\} = Comp(\mathcal{A}_2)$ *therefore they are labelling equivalent under Comp (via identity) but they are not isomorphic.*

The following property says that the distance should be the same for AFs with the same set of labellings.

**(LAB$_{Sem}$)** If AFs $Args_1$, $Args_2$ are labelling equivalent via $g$ (under semantics *Sem*) then

$$d_{\mathcal{A}_1}(S, T) = d_{\mathcal{A}_2}(g(S), g(T)) \qquad \text{(Labelling Equivalence)}$$

**Example 32.** *Consider labelling equivalent under Comp* $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *from the previous example. Any distance method* $d$ *satisfying* **(LAB$_{Comp}$)** *needs to put* $d_{\mathcal{A}_1} = d_{\mathcal{A}_2}$*, e.g., we need* $d_{\mathcal{A}_1}(\mathtt{ioi}, \mathtt{uuu}) = d_{\mathcal{A}_2}(id(\mathtt{ioi}), id(\mathtt{uuu})) = d_{\mathcal{A}_2}(\mathtt{ioi}, \mathtt{uuu})$*.*

Note that for any semantics *Sem* which satisfies language independence, isomorphic AFs are also labelling equivalent, therefore for such semantics **(Lab$_{Sem}$)** implies **(ISO)**.

This postulate requires distance method to compute distance measure for a particular AF based only on the set of feasible labellings. In particular a distance method satisfying **(Lab$_{Sem}$)** ignores information about the structure of AF which was not extracted by a semantics. Under this condition the two compositionality postulates from the previous subsection are equivalent.

**Proposition 33.** *Let* $d$ *be a distance method satisfying* **(LAB$_{Sem}$)** *under the semantics Sem which satisfies independence. Then* $d$ *satisfies* **(COM)** *iff* $d$ *satisfies* **(COM$_{Sem}$)***.*

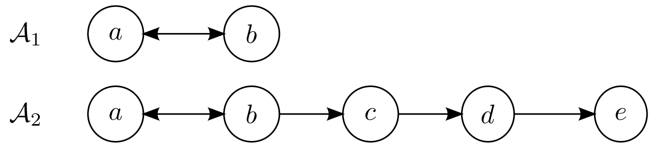



Figure 9: 2-loop $\mathcal{A}_1$ and its extension $\mathcal{A}_2$ with peripheral issues $c, d, e$

*Proof.* Since **(COM$_{Sem}$)** is stronger for semantics *Sem* which satisfies independence, we only need to show that **(COM)** implies **(COM$_{Sem}$)**. Let $\mathcal{A}$ be such that there exists semantically independent partition $A, B$ of $Args_{\mathcal{A}}$, i.e., $Sem(\mathcal{A}) = Sem(\mathcal{A}_{\downarrow A}) \otimes Sem(\mathcal{A}_{\downarrow B})$. Since our semantics satisfies independence, $Sem(\mathcal{A}_{\downarrow A}) \otimes Sem(\mathcal{A}_{\downarrow B}) = Sem(\mathcal{A}_{\downarrow A} \cup \mathcal{A}_{\downarrow B})$. Since $d$ satisfies **(LAB$_{Sem}$)**, $d_{\mathcal{A}} = d_{\mathcal{A}_{\downarrow A} \cup \mathcal{A}_{\downarrow B}}$ and because $d$ satisfies **(COM)**, $d_{\mathcal{A}_{\downarrow A} \cup \mathcal{A}_{\downarrow B}} = d_{\mathcal{A}_{\downarrow A}} + d_{\mathcal{A}_{\downarrow B}}$. □

We finish this subsection, by one postulate that will play an important role in Chapter 5 where we show that judgement aggregation based on a distance methods which fail that postulate is prone to agenda manipulation.

It is specific for the complete-based semantics only, i.e. semantics *Sem* which for all AFs $\mathcal{A}$ return complete labellings ($Sem(\mathcal{A}) \subseteq Comp(\mathcal{A})$), e.g. preferred, stable, semi-stable semantics.

Let $\mathcal{A} = (Args, \rightharpoonup)$ be an arbitrary AF and let $\mathcal{A}^+$ be any framework obtained from $\mathcal{A}$ by adding a single new argument $b \notin Args$ along with a **single** attack $a \rightharpoonup b$ from some $a \in Args$. Then for any two complete labellings $S, T$ over the expanded AF $\mathcal{A}^+$ the following should hold for any distance method:

**(IPI)** $\quad d_{\mathcal{A}^+}(S, T) = d_{\mathcal{A}}(S[Args], T[Args])$.

(Indifference to peripheral issues)

Here, in the right-hand side, $S[Args]$ denotes $S$ restricted to the arguments in *Args*, i.e., ignoring $b$ (and similarly for $T[Args]$). It is easy to check that $S[Args], T[Args] \in Sem(A)$. The property essentially says that adding $a \rightharpoonup b$ to $\mathcal{A}$ should not make any difference to the distance, intuitively because the label of $b$ is in any case determined by that of $a$, and so the introduction of $b$ does not change the situation.

**Example 34.** *Consider AF* $\mathcal{A}_1 = (\{a, b\}, \{(a, b), (b, a)\})$ *and* $\mathcal{A}_2 = (\{a, b, c, d, e\}, \{(a, b), (b, a), (b, c), (c, d), (d, e)\})$, *which is a result of three expansions of* $\mathcal{A}_1$*, depicted in Figure 9. We have the following complete labellings* $Comp(\mathcal{A}_1) = \{L_1\colon \texttt{io}, L_2\colon \texttt{uu}, L_3\colon \texttt{oi}\}$ *and* $Comp(\mathcal{A}_2) = \{L'_1\colon \texttt{ioioi}, L'_2\colon \texttt{uuuuu}, L'_3\colon \texttt{oioio}\}$. *Despite the fact that* $\mathcal{A}_2$ *has three more arguments than* $\mathcal{A}_1$ *'behaviour' of both boils down to* $a, b$ *loop. Any distance methods satisfying* **(IPI)** *(used 3 times) needs to put* $d_{\mathcal{A}_1}(L_i, L_j) = d_{\mathcal{A}_2}(L'_i, L'_j)$ *for* $i, j = 1..3$.

**(IPI)** can be considered as equivalence-based postulate stating that since any AF $\mathcal{A}$ and its extension $\mathcal{A}^+$ are in a sense *Sem*-equivalent

| | |
|---|---|
| Metric Postulates | |
| **(REF)** | Reflexivity |
| **(DD)** | Dissimilarity of the Diverse |
| **(SYM)** | Symmetry |
| **(TRI)** | Triangle inequality |
| Intuition-based Postulates | |
| **(BTW$_{\triangleright}$)** | Simple Betweenness Monotonicity |
| **(BTW$_{\blacktriangleright}$)** | Refined Betweenness Monotonicity |
| **(QDA$_{\ominus_{HS}}$)** | Qualitative Distance Alignment by Hamming Set |
| **(SQDA$_{\ominus_{HS}}$)** | Strict Qualitative Distance Alignment by Hamming Set |
| **(QDA$_{\ominus_{RHP}}$)** | Qualitative Distance Alignment by Refined Hamming Pair |
| **(SQDA$_{\ominus_{RHP}}$)** | Strict Qualitative Distance Alignment by Refined Hamming Pair |
| Compositionality Postulates | |
| **(COM)** | Syntactic Compositionality |
| **(COM$_{Sem}$)** | Semantic Compositionality (under semantics *Sem*) |
| Equivalence Postulates | |
| **(AUTO)** | Automorphism Equivalence |
| **(ISO)** | Isomorphism Equivalence |
| **(LAB$_{Sem}$)** | Labelling Equivalence (under semantics *Sem*) |
| **(IPI)** | Indifference to peripheral issues |

Table 1: The overview of the postulates

with the bijection between labellings assigning to $L \in Sem(\mathcal{A}^+)$ its restriction $L[Args] \in Sem(\mathcal{A})$ therefore the distances between the corresponding labellings should be the same.

## 3.4 SUMMARY

In this chapter we proposed several postulates of distance measures (metric postulates, betweenness and qualitative distance) and distance methods (compositionality and equivalence postulates).

Simple discrete metric satisfies metric postulates **(REF, DD, SYM,TRI)**, but fails betweenness **(BTW$_{\triangleright}$, BTW$_{\blacktriangleright}$)**, strong qualitative distance alignment **(SQDA$_{\ominus_{HS}}$, SQDA$_{\ominus_{RHP}}$)** and compositionality **(COM, COM$_{Sem}$)**. Those postulates give motivation for searching for a more detailed measure. The equivalence postulates

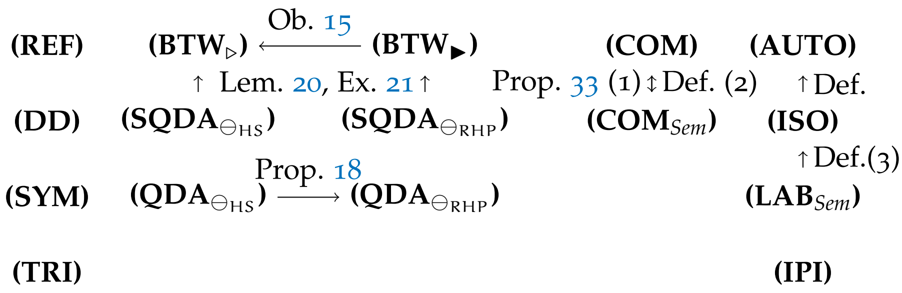



Figure 10: The dependencies between the postulates. 1) For d satisfying **(LAB$_{Sem}$)**, 2) For *Sem* satisfying Independence principle, 3) For *Sem* satisfying Language Independence principle.

**(ISO, AUTO, LAB$_{Sem}$, IPI)** and qualitative distance alignment **(QDA$_{\ominus_{HS}}$, QDA$_{\ominus_{RHP}}$)** set the limit and specify which distances should be the same.

The betweenness postulate **(BTW$_{\triangleright}$)** and qualitative distance postulates **(SQDA$_{\ominus}$, QDA$_{\ominus}$)** formalise some intuition and depend on the provided betweenness relation $\triangleright$ and partially ordered qualitative distance $\ominus$. We proposed two concrete betweenness relations (simple $\triangleright$ /refined $\blacktriangleright$) and two qualitative distances (hamming set $\ominus_{HS}$ and refined hamming pair $\ominus_{RHP}$).

The syntactic compositionality postulate **(COM)**, isomorphism **(ISO)**, and automorphism **(AUTO)** postulates depend only on the structure of the AF, although syntactic compositionality puts some mild assumptions about semantics used. The semantics compositionality postulate **(COM$_{Sem}$)** and labelling equivalence postulate **(LAB$_{Sem}$)** depend on the semantics used. Although betweenness and qualitative distance based postulates **(BTW$_{\triangleright}$, SQDA$_{\ominus}$, QDA$_{\ominus}$)** are defined for all labellings we think they should hold for labellings designated by the semantics used, since their motivation relay on the fact that compared labellings represent agents' reasonable position. It will be important in Chapter 5 where distances between any labellings are considered. The indifference to peripheral issues postulate **(IPI)** is motivated just for complete-based semantics - the semantics we will concentrate on in Chapter 5.

We proposed the Independence principle for semantics and we have investigated dependencies between proposed postulates which are depicted in Figure 10.

In the next chapter we construct some concrete distance methods which satisfy the postulates from this chapter.

# 4 QUANTIFYING DISAGREEMENT BETWEEN LABELLINGS: PRODUCT DISTANCE

## 4.1 INTRODUCTION

In this chapter we investigate a class of distance methods of a specific form. The idea is first to define a distance *diff* over the set $\{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ of *labels* and then define the distance between two $\mathcal{A}$-labellings as a sum over some set of arguments of the distances between labels assigned by those $\mathcal{A}$-labellings to arguments from the set. Formally, all of the distance methods will share the following form.

**Definition 35.** *A* product distance method *is a distance method which is specified by* (i) *a label measure diff, i.e. a function from the set of pairs of labels* $\{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}^2$ *to* $\mathbb{R}$ *and* (ii) *a selection function* $\mathfrak{S}$ *which for each AF* $\mathcal{A}$ *selects a subset* $A \subseteq Args_{\mathcal{A}}$ *of "important" arguments in* $\mathcal{A}$*, and which assigns to AF* $\mathcal{A}$ *a* product distance measure *given by the following equation:*

$$d^{diff}_{\mathfrak{S}(\mathcal{A})}(L_1, L_2) = \sum_{a \in \mathfrak{S}(\mathcal{A})} diff(L_1(a), L_2(a)) \qquad (1)$$

*where* $L_1, L_2 \in Sem(\mathcal{A})$.

This form is a standard way to define a distance over Cartesian product, hence the name. Each argument can be seen as separate dimension over which we measure a disagreement between labellings.

The question we address in this chapter is the following:

> Under which conditions placed on a selection function $\mathfrak{S}$ and a label measure *diff*, does a product distance method satisfy postulates proposed in the previous chapter?

We start from the simplest distance method and tweak the choice of the selection function and label measure so that it satisfies all the postulates.

It turns out that the results in this chapter depend only on a few fundamental requirements on *diff*:

**(ref)** $\textit{diff}(\mathsf{x},\mathsf{x}) = 0$

**(sym)** $\textit{diff}(\mathsf{x},\mathsf{y}) = \textit{diff}(\mathsf{y},\mathsf{x})$

**(dd)** if $\mathsf{y} \neq \mathsf{x}$ then $\textit{diff}(\mathsf{x},\mathsf{y}) > 0$

**(dd-)** $\textit{diff}(\mathtt{in},\mathtt{out}) > 0$

**(tri)** $\textit{diff}(\mathsf{x},\mathsf{z}) \leqslant \textit{diff}(\mathsf{x},\mathsf{y}) + \textit{diff}(\mathsf{y},\mathsf{z})$

**(gcs)** there exists $\mathbf{c} \in \mathbb{R}$ s.t. if $\mathsf{y} \neq \mathsf{x}$ then $\textit{diff}(\mathsf{x},\mathsf{y}) = \mathbf{c}$
(General Conflict Similarity)

**(scs)** $\textit{diff}(\mathtt{in},\mathtt{undec}) = \textit{diff}(\mathtt{out},\mathtt{undec})$ (Soft Conflict Similarity)

**(hcs)** $\textit{diff}(\mathtt{in},\mathtt{out}) \geqslant \max\{\textit{diff}(\mathtt{in},\mathtt{undec}),\textit{diff}(\mathtt{out},\mathtt{undec})\}$
(Hard Conflict Significance)

**(hcs+)** $\textit{diff}(\mathtt{in},\mathtt{out}) > \max\{\textit{diff}(\mathtt{in},\mathtt{undec}),\textit{diff}(\mathtt{out},\mathtt{undec})\}$
(Strong Hard Conflict Significance)

We can recognise the properties of the metric space **(ref, sym, dd, tri)** (written here in lower case to emphasise that they are properties of label measure) and a few additional properties **(gcs, scs, hcs, hcs+, dd-)** describing the relation between different types of conflict.

We will always assume *diff* to satisfy **(ref)** and **(sym)**. We implicitly assume it in other properties.

Note the following dependencies: **(dd-)** is a weakening of **(dd)**, **(hcs+)** is a strenghtening of **(hcs)**, **(gcs)** implies **(scs)**.

Note that the space of labels is small therefore *diff* can be given by a few constants. If *diff* satisfies **(sym)** then it is completely specified by three quantities: $\textit{diff}(\mathtt{in},\mathtt{undec})$, $\textit{diff}(\mathtt{in},\mathtt{out})$ and $\textit{diff}(\mathtt{out},\mathtt{undec})$. Furthermore, if *diff* satisfies **(scs)** then it is completely specified by 2 quantities: $\textit{diff}(\mathtt{in},\mathtt{undec})$ and $\textit{diff}(\mathtt{in},\mathtt{out})$, which may respectively be thought of as the costs attached to a *soft* and *hard* conflict. Further, if *diff* satisfies **(gcs)** then it is described by single constant **c**.

The most obvious choice for a *diff* function is a *discrete* metric, i.e. $\mathrm{DM}(\mathsf{x},\mathsf{y}) = 1$ if $\mathsf{x} \neq \mathsf{y}$, $\mathrm{DM}(\mathsf{x},\mathsf{y}) = 0$ if $\mathsf{x} = \mathsf{y}$ which satisfies **(ref)**, **(dd)**, **(sym)**, **(tri)**, **(gcs)**, **(hsc)**.

## 4.2 FULL SUM METHOD

The most obvious choice for $\mathfrak{S}$ is a function returning the set of all arguments in the framework. We will fix it for now and concentrate only on the *diff* function.

**Definition 36.** *A* full sum method *is a product distance method with* $\mathfrak{S}(\mathcal{A}) = \textit{Args}_{\mathcal{A}}$ *for all* $\mathcal{A}$.

An important member of this family is the Hamming distance which can be obtaining by plugging the discrete metric into the full sum method.

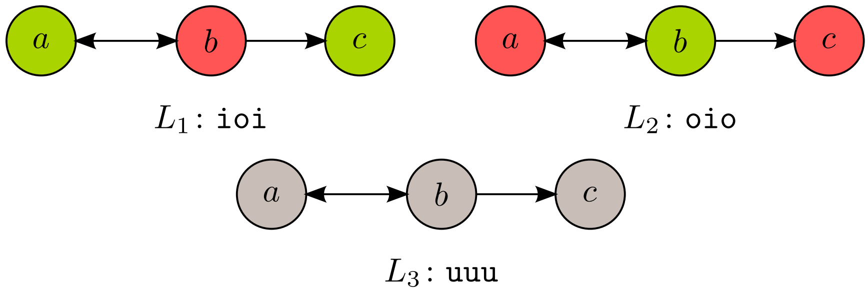



Figure 11: Three possible complete labellings $L_1, L_2$ and $L_3$

**Definition 37.** *A* Hamming distance *is a full sum method with diff* $=$ DM *for all* $\mathcal{A}$. *We will denote it* Hd.

**Example 38.** *Consider for instance the results for Figure 11, where we see that* $\mathrm{Hd}(L_1, L_2) = 3 = \mathrm{Hd}(L_1, L_3)$. *Thus, according to* Hd, *labellings* $L_2$ *and* $L_3$ *are equidistant from* $L_1$.

Hamming distance satisfies all the metric properties. In fact it is due to the fact that we used discrete metric for *diff*. In general metric properties of the full sum measure are inherited from the *diff* function.

**Proposition 39.** *If* d *is a full sum distance method defined via a label measure diff then* d *satisfies* **(REF)**, **(DD)**, **(SYM)**, **(TRI)** *if the corresponding properties* **(ref)**, **(dd)**, **(sym)**, **(tri)** *are satisfied by diff.*

We will obtain proofs of the propositions in this section as a corollaries of more general results in the sections that follow.
Hamming distance satisfies **(BTM$_\triangleright$)** but fails **(BTM$_\blacktriangleright$)**.

**Example 40.** *Consider again AF from Figure 11. We have* $L_1 \blacktriangleright\blacktriangleright L_2 \blacktriangleleft\blacktriangleleft L_3$ *but* $\mathrm{Hd}(L_1, L_2) = 3 = \mathrm{Hd}(L_1, L_3)$.

**Proposition 41.** *If diff satisfies* **(ref)** *and* **(dd)** *then the full sum measure satisfies* **(BTM$_\triangleright$)**. *If diff satisfies also* **(hcs+)** *then the full sum measure satisfies* **(BTM$_\blacktriangleright$)**.

Hamming distance does not satisfy **(hcs+)** but we can easily refine it replacing discrete metric with the refined version:

$$\mathrm{rDM}(\mathtt{in}, \mathtt{out}) = 2, \qquad \mathrm{rDM}(\mathtt{in}, \mathtt{undec}) = \mathrm{rDM}(\mathtt{out}, \mathtt{undec}) = 1.$$

Note $\mathrm{rDM}(x, y)$ may be thought of as the length of the shortest path between $x$ and $y$ in the neighbourhood graph $\mathtt{in} - \mathtt{undec} - \mathtt{out}$ over the labels. It is still a metric over set of labels. The distance obtained by plugging rDM into the full sum measure we call the Refined Hamming distance.

**Definition 42.** *A* Refined Hamming distance *is a full sum method with diff* $=$ rDM *for all* $\mathcal{A}$. *We will denote it* rHd.

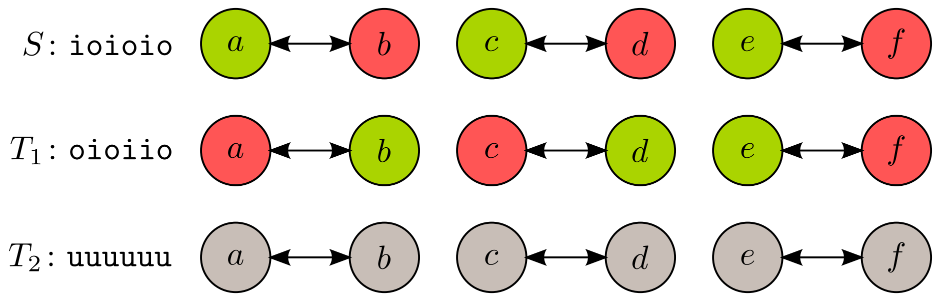



Figure 12: Example showing opposite results of Hd and rHd.

**Example 43.** *Going back to Figure 11, we have*

$$\mathrm{rHd}(L_1, L_2) = 3 \times \mathrm{rDM}(\mathtt{in}, \mathtt{out}) = 6 \textit{ and}$$
$$\mathrm{rHd}(L_1, L_3) = 3 \times \mathrm{rDM}(\mathtt{in}, \mathtt{undec}) = 3,$$

*yielding the expected* $\mathrm{rHd}(L_1, L_3) < \mathrm{rHd}(L_1, L_3)$.

Note that there is an incompatibility between Hd and rHd, in the sense that there exist examples in which Hd and rHd yield opposite conclusions regarding the relative proximity of two labellings $T_1$, $T_2$ to a given S. For example, consider the three complete labellings of the AF containing three pairs of mutually attacking arguments in Figure 12.
Here we have $\mathrm{Hd}(S, T_1) = 4 < 6 = \mathrm{Hd}(S, T_2)$ and $\mathrm{rHd}(S, T_1) = 8 > 6 = \mathrm{rHd}(S, T_2)$.
Until now the Refined Hamming distance satisfies all the mentioned properties because refined discrete metric is still a metric and the previous results still holds. Let us investigate what happens with **(QDA)** and **(SQDA)**. In Chapter 3 we observed that **(SQDA$_{\ominus_{RHP}}$)** is incompatible with **(QDA$_{\ominus_{HS}}$)** (Example 19). In the context of the full sum distances, commitment to qualitative distance given by Hamming set or Refined Hamming pair leads to Hamming or Refined Hamming distance.

**Proposition 44.** *A full sum distance method satisfies* **(QDA$_{\ominus_{HS}}$)** *and* **(SQDA$_{\ominus_{HS}}$)** *if it is defined via diff function satisfying* **(dd)** *and* **(gcs)**.

This sufficient condition happens also to be necessary for most interesting semantics as shown by the following example.

**Example 45.** *Consider* $\mathcal{A} = (\{a, b, c\}, \{(a, b), (b, a), (b, c)\})$ *and any semantics Sem such that* $\{\mathtt{ioi}, \mathtt{oio}, \mathtt{uuu}\} \subseteq \mathit{Sem}(\mathcal{A})$*, e.g.the complete semantics (See again Figure 11). Since* $(\mathtt{ioi} \ominus_{HS} \mathtt{oio}) = (\mathtt{ioi} \ominus_{HS} \mathtt{uuu}) = (\mathtt{uuu} \ominus_{HS} \mathtt{oio}) = \{a, b, c\}$ *for any full sum distance measure* $d^{\mathit{diff}}$ *satisfying* **(QDA$_{\ominus_{HS}}$)** *defined via diff function satisfying* **(sym)** *the corresponding distances* $d^{\mathit{diff}}(\mathtt{ioi}, \mathtt{oio}) = 3 \times \mathit{diff}(\mathtt{in}, \mathtt{out})$*,* $d^{\mathit{diff}}(\mathtt{ioi}, \mathtt{uuu}) = 2 \times \mathit{diff}(\mathtt{in}, \mathtt{undec}) + \mathit{diff}(\mathtt{out}, \mathtt{undec})$ *and* $d^{\mathit{diff}}(\mathtt{uuu}, \mathtt{oio}) = \mathit{diff}(\mathtt{in}, \mathtt{undec}) +$

$2 \times$ *diff*(out, undec) *need to be equal as well. Comparing the last two we obtain diff*(in, undec) $=$ *diff*(out, undec)*. Using this while comparing the first distances we obtain diff*(in, out) $=$ *diff*(out, undec).

The above example shows that Hamming distance (up to constant) is the only full sum distance satisfying **(QDA$_{\ominus_{HS}}$)**. Note that this is incompatible with **(BTM$_{\blacktriangleright}$)** which requires in general that *diff* satisfies **(hcs+)**.

**(QDA$_{\ominus_{RHP}}$)** is weaker than **(QDA$_{\ominus_{HS}}$)** and leaves a bit more space for *diff*.

**Proposition 46.** *A full sum distance method* d *satisfies* **(QDA$_{\ominus_{RHP}}$)** *if it is defined via diff function satisfying* **(dd)**, **(scs)** *and* **(hcs)**. *If diff function satisfies also* **(hcs+)** *then* d *satisfies* **(SQDA$_{\ominus_{RHP}}$)**.

The above result shows that Refined Hamming distance (up to constants attached to hard and soft conflict) is the only full sum distance satisfying **(SQDA$_{\ominus_{RHP}}$)**.

The above results pin down the conditions imposed by the Intuition-based postulates on full sum distance. We will see that they carry over to a more general family of product distance methods. On one hand they may not seem surprising, because restrictions follow the strengthening of the postulates observed in the previous chapter. On the other hand it is good to be aware how strong they are.

Let us now investigate how does full sum distances behave with regards to Compositionality and Equivalence postulates?

**Proposition 47.** *Every full sum distance method satisfies* **(LAB$_{Sem}$)** *(hence* **(ISO)** *and* **(AUTO)***) and* **(COM$_{Sem}$)**.

**(LAB$_{Sem}$)** and **(COM$_{Sem}$)** put no constraints on the choice of *diff* function. However a big problem with full sum methods is that they fail to satisfy **(IPI)**, as the following example (for the case of Hd) shows.

**Example 48.** *Again consider AF in Figure 11. This is an extension of the AF without argument* c*. Any distance measure satisfying* **(IPI)** *should put* $\mathrm{d}(\mathrm{io},\mathrm{oi}) = \mathrm{d}(\mathrm{ioi},\mathrm{oio})$ *but we have* $2 = \mathrm{Hd}(\mathrm{io},\mathrm{oi}) \neq \mathrm{Hd}(\mathrm{ioi},\mathrm{oio}) = 3$.

The above example works not only for Hamming distance. In fact any full sum distance method defined via *diff* function satisfying **(dd-)** will add positive value to the 'tail' and fail **(IPI)**. To find a distance satisfying **(IPI)** we need to consider a wider family of distances.

## 4.3 CRITICAL SET BASED METHODS

The only way to fix (IPI) is to ignore some arguments. This leads us to the study of other selection functios $\mathfrak{S}$.

The first idea comes from a concept introduced by Gabbay (2009) (but here generalised to make it relative to a set of labellings $\mathfrak{X}$ - Gabbay considered only the case $\mathfrak{X} = Comp(\mathcal{A})$). Instead of looking at all arguments, one specifically focuses on the *critical subsets*.

**Definition 49.** *Given $\mathcal{A}$ and $\mathfrak{X} \subseteq Labs(\mathcal{A})$, a set of arguments $A \subseteq Args_{\mathcal{A}}$ is $\mathfrak{X}$-*critical (for $\mathcal{A}$) *iff for any $L_1, L_2 \in \mathfrak{X}$, if $L_1[A] \equiv L_2[A]$ then $L_1 \equiv L_2$. We denote the set of critical subsets for $\mathcal{A}$ and the set of labellings $\mathfrak{X}$ by $\mathfrak{X}$-crit$(\mathcal{A})$.*

A $\mathfrak{X}$-critical set is a set of arguments such that any two labellings of $\mathfrak{X}$ are different iff they label at least one argument in the set differently.

In other words a $\mathfrak{X}$-critical set for $\mathcal{A}$ is a set of arguments whose status is enough to determine the status of *all* the arguments in *Args* (knowing that they are labelled by one of labellings from $\mathfrak{X}$). Clearly at least one critical subset exist, for $Args_{\mathcal{A}}$ is obviously $\mathfrak{X}$-critical for any set of labellings $\mathfrak{X}$. In general the following holds.

**Observation 50.** *For any two sets of $\mathcal{A}$-labellings $\mathfrak{X}_1 \subseteq \mathfrak{X}_2$, if a set of arguments $A$ is $\mathfrak{X}_2$-critical then it is $\mathfrak{X}_1$-critical. In the limiting cases, any set (also empty) is $\emptyset$-critical and $\{L\}$-critical while there is only one $Labs(\mathcal{A})$-critical set which is the set of all arguments $Args_{\mathcal{A}}$.*

We formulated the notion of critical set in terms of arbitrary sets of labellings but of course such a set is given a meaning. Usually it will be a set of labellings accepted by a particular labelling semantics. In such cases we will write *Sem*-critical set instead of *Sem*$(\mathcal{A})$-critical. This means that $Args_{\mathcal{A}}$ is the only $\mathtt{Labs}$-critical set for $\mathcal{A}$, there are usually more *Comp*-critical sets, and for single status semantics, like grounded any set of arguments is critical.

**Definition 51.** *A* critical set method *is a product distance method with $\mathfrak{S}(\mathcal{A})$ is a Sem-critical set for all $\mathcal{A}$.*

Critical sets are important because the Metric and Betweenness constraints discussed in the previous section hold not only for the set of all arguments but for all critical sets.

**Theorem 52.** *Let $\mathcal{A}$ be an AF, $\mathfrak{X} \subseteq Labs(\mathcal{A})$ and $d^{diff}_{\mathfrak{S}(\mathcal{A})}$ be product distance measure defined via diff and $\mathfrak{S}$. If diff is a metric over labels then $d^{diff}_{\mathfrak{S}(\mathcal{A})}$ is a pseudometric over $\mathfrak{X}$ (satisfies* **(REF, SYM, TRI)** *but not* **(DD)***). Moreover, $d^{diff}_{\mathfrak{S}(\mathcal{A})}$ is a metric over $\mathfrak{X}$ iff $\mathfrak{S}(\mathcal{A})$ is $\mathfrak{X}$-critical.*

Before we start to prove this result let us recall a few metric transforms from topology - constructions which define metric over one set in terms of other metrics.

**Lemma 53.** *(Metric Transforms)*

1. ***Induced metric* (*Deza and Deza (2009)* *page 41*)** *Given a metric space* $(X, d)$ *and a subset* $X' \subset X$*, an induced metric is the restriction* $d'$ *of* $d$ *to* $X'$ *. A metric space* $(X', d')$ *is called a metric subspace of* $(X, d)$*, and the metric space* $(X, d)$ *is called a metric extension of* $(X', d')$*;*

2. ***t-scaled metric* (*Deza and Deza (2009)* *page 80*)** *For a metric space* $(X, d)$ *and* $t > 0$*,* $(X, d_t)$ *is a metric space defining* $d_t(x, y) = t * d(x, y)$*;*

3. *For* $f : X \to Y$ *and a metric* $d$ *over* $Y$ *function* $d_f(a, b) = d(f(a), f(b))$ *is a pseudometrics over* $X$*. Moreover, if* $f$ *is an injection then* $d_f$ *is a metric (See also Deza and Deza (2009) page 81* ***Pullback metric****);*

4. ***Product metric* (*Deza and Deza (2009)* *page 83*)** *For metric spaces* $(X_1, d_1), \dots, (X_n, d_n)$ *functions*

$$D_{sum}(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i),$$
$$D(x, y) = \Big(\sum_{i=1}^{n} d_i(x_i, y_i)^p\Big)^{1/p}, 1 < p < \infty,$$
$$D_{max}(x, y) = \max_{i=1\dots n} d_i(x_i, y_i),$$
$$D(x, y) = \sum_{i=1}^{n} \frac{1}{2^i} \frac{d_i(x_i, y_i)}{1 + d_i(x_i, y_i)},$$

*are a metrics on Cartesian product* $X = X_1 \times \ldots \times X_n$*.*

*Proof.* We check the conditions of Definition 12 to show that the functions defined above are indeed metrics.

1. Trivial, we restricted scope of quantification.

2. Trivial, we multiply both sides of each equation by $t$.

3. **(DD)** For injective $f$ and $a \neq b$ we have $f(a) \neq f(b)$ and since $d$ is a metrics $d_f(a, b) = d(f(a), f(b)) > 0$. Now if $f$ is not injective then there exists some $a \neq b$ such that $f(a) = f(b)$ and then $d_f(a, b) = 0$; **(REF)** $d_f(a, a) = d(f(a), f(a)) = 0$; **(SYM)** $d_f(a, b) = d(f(a), f(b)) = d(f(b), f(a)) = d_f(b, a)$; **(TRI)** $d_f(a, b) = d(f(a), f(b)) \leqslant d(f(a), f(c)) + d(f(c), f(b)) = d_f(a, c) + d_f(c, b)$.

4. We just prove the first case $D_{sum}$ which we use later on. **(DD)** For $a \neq b$ we have at least one $j$ such that $a_j \neq b_j$ therefore $D(a, b) = \sum_{i=1}^{n} d_i(a_i, b_i) \geqslant d_j(a_j, b_j) > 0$; **(REF)** $D(a, a) = \sum_{i=1}^{n} d_i(a_i, a_i) = \sum_{i=1}^{n} 0 = 0$; **(SYM)** $D(a, b) = \sum_{i=1}^{n} d_i(a_i, b_i) = \sum_{i=1}^{n} d_i(b_i, a_i) = D(b, a)$; **(TRI)** $D(a, b) = \sum_{i=1}^{n} d_i(a_i, b_i) \leqslant \sum_{i=1}^{n} \big(d_i(a_i, c_i) + d_i(c_i, b_i)\big) = \sum_{i=1}^{n} d_i(a_i, c_i) + \sum_{i=1}^{n} d_i(c_i, b_i) = D(a, c) + D(c, b)$.

□

Now we can proceed with the proof of Theorem 52.

*Proof.* The difficulty is to notice that function given by Equation 1 is a pullback metric of the product metric space $\{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}^{|\mathfrak{S}(\mathcal{A})|}$ with the sum metric.

Let denote by $n$ the number of selected arguments, i.e. $|\mathfrak{S}(\mathcal{A})| = n$. For label metrics $\mathtt{diff}_1, ..., \mathtt{diff}_n$ the function $D_{sum}(x, y) = \sum_{i=1}^{n} \mathtt{diff}_i(x_i, y_i)$ defines sum metric over Cartesian product $\{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}^n$. Let us order elements of $\mathfrak{S}(\mathcal{A})$ by a bijection $o\colon \mathfrak{S}(\mathcal{A}) \to \{1, \ldots, n\}$. We associate with it a function $g_o\colon \mathtt{Labs}(\mathcal{A}) \supseteq \mathcal{X}(\mathcal{A}) \to \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}^n$ assigning to the labelling $L$ a vector of $n$ labels assigned by $L$ to the selected arguments in order given by $o$, i.e. $g_o(L) = (L(o^{-1}(1)), \ldots, L(o^{-1}(n)))$. Now consider the pullback metric $(D_{sum})_{g_o}$. In Equation 1 the same label metrics is used for each argument, i.e. $\mathtt{diff}_i = \mathtt{diff}$ for $i = 1..n$, and ordering of arguments $o$ is not mention since the outcome of the sum does not depend on the bijection $o$. Finally note that $g_o$ is an injection iff $\mathfrak{S}(\mathcal{A})$ is a $\mathcal{X}$-critical set. □

Now we proceed with Betweenness Monotonicity.

**Theorem 54.** *If diff satisfies* **(ref)** *and* **(dd)** *then any critical set method satisfies* **(BTM$_\triangleright$)**. *If diff satisfies also* **(hcs+)** *then any critical set method satisfies* **(BTM$_\blacktriangleright$)**.

*Proof.* Let $L_1, L_2, L_3$ by such that $L_1 \triangleright L_2 \triangleleft\!\triangleleft L_3$ respectively $L_1 \blacktriangleright L_2 \blacktriangleleft\!\blacktriangleleft L_3$. We need to show that for all critical sets $\mathcal{C}$ it holds that:

$$d_{\mathcal{C}}(L_1, L_2) = \sum_{a \in \mathcal{C}} \mathit{diff}(L_1(a), L_2(a)) < \sum_{a \in \mathcal{C}} \mathit{diff}(L_1(a), L_3(a)) = d_{\mathcal{C}}(L_1, L_3).$$

We compare sums argument by argument. Consider the following (not exclusive) cases:

CASE A: $L_1(a) = L_2(a)$
: we have $0 = \mathit{diff}(L_1(a), L_2(a)) \leqslant \mathit{diff}(L_1(a), L_3(a))$ due to **(ref)** and **(dd)**.

CASE B: $L_2(a) = L_3(a)$
: we have $\mathit{diff}(L_1(a), L_2(a)) = \mathit{diff}(L_1(a), L_3(a))$.

CASE C: $L_2(a) = \mathtt{undec} \wedge L_1(a) \neq L_3(a)$
: either one of previous cases hold or we have $\mathit{diff}(L_1(a), L_2(a)) = \mathit{diff}(L_1(a), \mathtt{undec}) < \mathit{diff}(L_1(a), L_3(a)) \in \{\mathit{diff}(\mathtt{in}, \mathtt{out}), \mathit{diff}(\mathtt{out}, \mathtt{in})\}$ by **(hcs+)**.

Note that in all cases the left side is not bigger than the right side. We need to show that there exists $a$ for which right side is strictly bigger. $L_2 \neq L_3$ and $\mathcal{C}$ is critical therefore there exists an argument $a \in \mathcal{C}$ for which $L_2(a) \neq L_3(a)$. This excludes Case B. In Case C the right side is bigger by **(hcs+)**, in Case A the right side is strictly bigger since $L_2(a) \neq L_3(a)$ by **(dd)**. For $L_1 \triangleright L_2 \triangleleft \triangleleft L_3$ Case C is excluded so we do not need to assume **(hcs+)**.

Since for all arguments $a \in \mathcal{C}$, $\mathit{diff}(L_1(a), L_2(a)) \leqslant \mathit{diff}(L_1(a), L_3(a))$ and for some argument $a \in \mathcal{C}$, $\mathit{diff}(L_1(a), L_2(a)) < \mathit{diff}(L_1(a), L_3(a))$ we have $d_{\mathcal{C}}(L_1, L_2) < d_{\mathcal{C}}(L_1, L_3)$. □

Note the above result holds taking $\mathcal{C}$ to be *any* critical subset, therefore by taking $\mathcal{C} = \mathit{Args}$ we obtain Proposition 39, Proposition 41 in the previous section as corollaries of Theorem 52 and Theorem 54 respectively.

Qualitative distance alignment doesn't depend on the selection function $\mathfrak{S}$ at all.

**Theorem 55.** *A product distance measure* d *defined via diff function satisfying* **(dd)**, **(scs)** *and* **(hcs)**, *satisfies* **(QDA$_{\ominus_{RHP}}$)**. *If diff function satisfy* **(dd)** *and* **(gcs)** *then* d *satisfies* **(QDA$_{\ominus_{HS}}$)**.

*Proof.* Let $d_{\mathcal{A}r}$ be a product distance measure defined via *diff*. Based on the assumption about diff we group addends by conflict type.

Let $C_L = (L_1 \ominus_{HS} L_2) \subseteq (N_1 \ominus_{HS} N_2) = C_N$ and assume diff satisfies **(dd)** and **(gcs)** ($\mathit{diff}(x, y) = c$ for $x \neq y$, $c > 0$). We have:

$$
\begin{aligned}
d_{\mathcal{A}r}(N_1, N_2) = \sum_{a \in \mathcal{A}r} & \mathit{diff}(N_1(a), N_2(a)) \\
=& c \times |\mathcal{A}r \cap C_N| \\
=& c \times (|\mathcal{A}r \cap (C_N \setminus C_L)| + |\mathcal{A}r \cap C_L|) \qquad (*) \\
=& c \times |\mathcal{A}r \cap (C_N \setminus C_L)| \\
& + d_{\mathcal{A}r}(L_1, L_2).
\end{aligned}
$$

(*) The fact that $C_L \subseteq C_N$ allows us to partition $C_N$ into $C_N \setminus C_L$ and $C_L$. For assumed diff we have $c > 0$ and $|\mathcal{A}r \cap (C_N \setminus C_L)| \geqslant 0$. It is clear from the reformulation above that $d(N_1, N_2)$ has not decreased.

We proceed similarly with $\ominus_{RHP}$. Let $\langle C_L, H_L \rangle = (L_1 \ominus_{RHP} L_2) \subseteq (N_1 \ominus_{RHP} N_2) = \langle C_N, H_N \rangle$ which translates into two inclusions $C_L \subseteq C_N$ and $H_L \subseteq H_N$ and assume diff satisfies **(dd)**, **(scs)** and **(hcs)**. For this assumptions *diff* is characterised by two constants $c_h \geqslant c_s > 0$ assigned to soft and hard conflicts respectively. We have:

$$
\begin{aligned}
d(N_1, N_2) = \sum_{a \in \mathcal{A}r} & \mathit{diff}(N_1(a), N_2(a)) \\
=& (c_h - c_s) \times |\mathcal{A}r \cap H_N| \qquad (*) \\
& + c_s \times |\mathcal{A}r \cap C_N| \\
=& (c_h - c_s) \times (|\mathcal{A}r \cap (H_N \setminus H_L)| + |\mathcal{A}r \cap H_L|) \\
& + c_s \times (|\mathcal{A}r \cap (C_N \setminus C_L)| + |\mathcal{A}r \cap C_L|) \qquad (**) \\
=& (c_h - c_s) \times |\mathcal{A}r \cap (H_N \setminus H_L)| \\
& + c_s \times |\mathcal{A}r \cap (C_N \setminus C_L)| \\
& + d(L_1, L_2).
\end{aligned}
$$

(*) Since hard conflict are also conflicts ($H_N \subseteq C_N, H_L \subseteq C_L$) we substract $(c_h - c_s)$ to avoid double counting; (**) Again partitioning is possible by assumed inclusions $H_L \subseteq H_N$ and $C_L \subseteq C_N$; For assumed $\mathtt{diff}$ we have $(c_h - c_s) \geqslant 0$ and $c_s > 0$ hence $d(N_1, N_2)$ has not decreased.

Note that here the set $\mathcal{A}r$ plays no role since we only need to show that distance has not decreased. It is different in the case of **(SQDA)** (Theorem 79). □

Note that by taking $\mathcal{A}r = Args$ we obtain **(QDA)** part of Propositions 44 and 46 in the previous section as corollaries of the above theorem.

As we have stated in the beginning, to satisfy **(IPI)** we need to ignore some arguments. The first idea is to concentrate on the *minimal* critical subsets.

**Definition 56.** *Given $\mathcal{A}$ and $\mathcal{X} \subseteq Labs(\mathcal{A})$, we denote the collection of set-theoretically minimal subsets of $\mathcal{X}$-crit$(\mathcal{A})$ by $\mathcal{X}$-mincrit$(\mathcal{A})$, i.e.,*
$\mathcal{X}\text{-}mincrit(\mathcal{A}) \stackrel{\text{def}}{=} \{\mathcal{C} \in \mathcal{X}\text{-}crit(\mathcal{A}) \mid \nexists \mathcal{C}'(\mathcal{C}' \in \mathcal{X}\text{-}crit(\mathcal{A}) \wedge \mathcal{C}' \subset \mathcal{C})\}$.

**Definition 57.** *A* minimal critical set method *is a product distance method with $\mathfrak{S}(\mathcal{A}) \in Sem\text{-}mincrit(\mathcal{A})$ for all $\mathcal{A}$.*

Does any minimal critical set measure satisfy **(IPI)**? The answer depends on the assigned selection function.

**Lemma 58.** *Let $\mathcal{A} = (Args, \rightharpoonup)$ be an arbitrary AF and let $\mathcal{A}^+$ be any framework obtained from $\mathcal{A}$ by adding a single new argument $b \notin Args$ along with a* **single** *attack $a \rightharpoonup b$ from some $a \in Args$. Then $Sem\text{-}crit(\mathcal{A}) \subseteq Sem\text{-}crit(\mathcal{A}^+)$ and $Sem\text{-}mincrit(\mathcal{A}) \subseteq Sem\text{-}mincrit(\mathcal{A}^+)$ for any complete based semantics Sem.*

*Proof.* For all $L \in Sem(\mathcal{A}^+)$ it holds that $L(a) = \neg L(b)$ because *Sem* is complete-based and $a$ is the only attacker of $b$. Therefore the argument $b$ contains 'the same information' as $a$ and a restriction $Sem(\mathcal{A}^+) \ni L \mapsto L[Args] \in Sem(\mathcal{A})$ is a bijection. □

The above lemma shows that for all $\mathcal{A}$ and its single argument extension $\mathcal{A}^+$ (as specified in **(IPI)**) it is possible to put

$$\mathfrak{S}(\mathcal{A}) = \mathfrak{S}(\mathcal{A}^+). \tag{2}$$

**Proposition 59.** *Let $d$ be a critical set method defined via selection function $\mathfrak{S}$. If $\mathfrak{S}$ satisfy 2 then $d$ satisfies* **(IPI)**.

*Proof.* Follows from the definition of product distance measure. □

The similar situation is with **(COM$_{Sem}$)**.

**Lemma 60.** *Let* $A$, $B$ *be semantically independent sets of an argumentation framework* $\mathcal{A}$ *under semantics Sem and let* $\mathcal{A}_1 = \mathcal{A}_{\downarrow A}, \mathcal{A}_2 = \mathcal{A}_{\downarrow B}$ *be the corresponding two subframeworks. If* $\mathcal{C}_1 \in \textit{Sem-crit}(\mathcal{A}_1)$ *and* $\mathcal{C}_2 \in \textit{Sem-crit}(\mathcal{A}_2)$ *then* $\mathcal{C}_1 \cup \mathcal{C}_2 \in \textit{Sem-crit}(\mathcal{A})$*. Moreover if* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *are minimal then* $\mathcal{C}_1 \cup \mathcal{C}_2$ *is also minimal.*

*Proof.* Take $L \in \textit{Sem}(\mathcal{A}) \neq \emptyset$. Since $A$, $B$ are semantically independent sets of an argumentation framework $\mathcal{A}$ then $L = L_A \cup L_B$ for some $L_A \in \textit{Sem}(\mathcal{A}_1)$ and $L_B \in \textit{Sem}(\mathcal{A}_2)$. We can identify $L_A$ and $L_B$ by $L[\mathcal{C}_1]$ and $L[\mathcal{C}_2]$ because $\mathcal{C}_1, \mathcal{C}_2$ are critical sets of $\mathcal{A}_1, \mathcal{A}_2$ respectively, hence $\mathcal{C}_1 \cup \mathcal{C}_2$ is a critical set. Since all combination of labellings $\textit{Sem}(\mathcal{A}_1)$ and $\textit{Sem}(\mathcal{A}_2)$ can be used to form a labelling of $\textit{Sem}(\mathcal{A})$, no information about $L_A$ can be inferred from $L_B$ and hence minimality follows. □

The above lemma shows that for all $\mathcal{A}$ whose arguments can be partitioned into independent (under semantics *Sem*) sets $A$, $B$ (as specified in **(COM$_{Sem}$)**) it is possible to put

$$\mathfrak{S}(\mathcal{A}) = \mathfrak{S}(\mathcal{A}_{\downarrow A}) \cup \mathfrak{S}(\mathcal{A}_{\downarrow B}). \tag{3}$$

Note that we can consider just $\mathcal{A}$ such that $\textit{Sem}(\mathcal{A}) \neq \emptyset$ because for $\mathcal{A}$ with $\textit{Sem}(\mathcal{A}) = \emptyset$ distance measures postulates are empty-fulfilled.

**Proposition 61.** *Let* $d$ *be a critical set measure defined via diff and selection function* $\mathfrak{S}$*. If* $\mathfrak{S}$ *satisfies 3 then* $d$ *satisfies* **(COM$_{Sem}$)**.

*Proof.* Follows from the definition of product distance measure. □

Unfortunately, this approach, although can be made to satisfy **(IPI)** and **(COM$_{Sem}$)**, has three problems which we will discuss now.

*Minimal critical set problems*

The first problem is that more than one minimal critical subset may exist. We would like the distance (or at least the similarity ordering induced by it) to be independent of the particular choice. Unfortunately this does not always hold in general, as the next example shows.

**Example 62.** *Let us return to the AF* $\mathcal{A}_1$ *depicted in Figure 13. It is not the case that by knowing the label of one argument we know the full complete labelling, however, one can check that if we know the label of any* pair *of arguments, we automatically know the label of the third. Thus we have* $\textit{Comp-mincrit}(\mathcal{A}_1) = \{\{a,b\},\{a,c\},\{b,c\}\}$*. We have* $d^{\textit{diff}}_{\{a,b\}}(L_1, L_2) = 2 \times \textit{diff}(\mathtt{in}, \mathtt{out})$ *and* $d^{\textit{diff}}_{\{a,b\}}(L_1, L_3) = \textit{diff}(\mathtt{in}, \mathtt{out})$*. Thus if we focus on the critical subset* $\{a,b\}$ *we obtain that* $L_3$ *is closer to* $L_1$ *than* $L_2$ *is. But if instead we focus on critical subset* $\{a,c\}$ *we obtain the opposite conclusion, for* $d^{\textit{diff}}_{\{a,c\}}(L_1, L_2) = \textit{diff}(\mathtt{in}, \mathtt{out})$ *and* $d^{\textit{diff}}_{\{a,c\}}(L_1, L_3) = 2 \times \textit{diff}(\mathtt{in}, \mathtt{out})$.

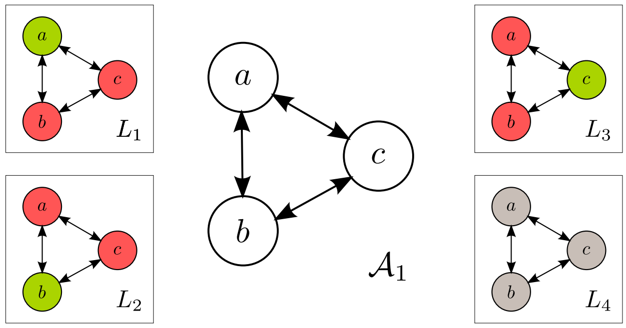



Figure 13: The AF $\mathcal{A}_1$ and its four possible complete labellings $L_1$-$L_4$

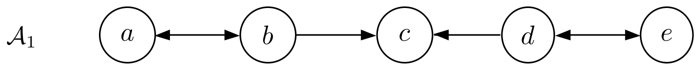



Figure 14: The AF for which all minimal critical set distance measures fail **(SQDA)**

This sensitivity to the choice of critical subset is somewhat undesirable.

The second problem is that the minimal critical set measure fails to satisfy **(AUTO)**, as the following example shows:

**Example 63.** *Consider $\mathcal{A}_1$ in Figure 13 and consider the mapping $g$ such that $g(a) = b$, $g(b) = c$ and $g(c) = a$. It is easy to see that $g$ is an automorphism on $\mathcal{A}_1$. Assume $\mathfrak{S}(\mathcal{A}_1) = \{a, b\}$. Recall $L_1 = \{(a, \mathtt{in}), (b, \mathtt{out}), (c, \mathtt{out})\}$ and $L_3 = \{(a, \mathtt{out}), (b, \mathtt{out}), (c, \mathtt{in})\}$. So $g(L_1) = \{(a, \mathtt{out}), (b, \mathtt{in}), (c, \mathtt{out})\}$ and $g(L_3) = \{(a, \mathtt{in}), (b, \mathtt{out}), (c, \mathtt{out})\}$. Then if* **(AUTO)** *were satisfied we would expect $d^{diff}_{\{a,b\}}(L_1, L_3) = d^{diff}_{\{a,b\}}(g(L_1), g(L_3))$, but $d^{diff}_{\{a,b\}}(L_1, L_3) = diff(\mathtt{in}, \mathtt{out}) \neq 2 \times diff(\mathtt{in}, \mathtt{out}) = d^{diff}_{\{a,b\}}(g(L_1), g(L_3))$. Note that this example assumes $\mathfrak{S}(\mathcal{A}_1) = \{a, b\}$, but it should be clear that counterexamples can also be found if either of the other two elements of Comp-mincrit$(\mathcal{A}_1)$ were selected.*

The above example show that this method doesn't satisfy **($\mathbf{LAB}_{Sem}$)** or **(ISO)** which are stronger than **(AUTO)**. Intuitively it is because there is not enough information in the labellings to select a particular minimal critical set. Such selection function $\mathfrak{S}$ needs to be given from outside. In consequence, first it is not easy to specify, second satisfaction of **($\mathbf{COM}_{Sem}$)** depends on user choice.

The third problem is that minimal critical set measure does not satisfy **(SQDA)** as can be seen in the following example.

**Example 64.** *Consider $\mathcal{A}_1$ from Figure 14. The label of $c$ is determined by the labels of the other arguments, so there are 4 possible minimal* Comp-*critical sets $\{a, b\} \otimes \{d, e\}$. Let's take $\{a, d\}$ as our selected minimal set (though the counterexample will also work for any of the other three possible choices). Consider labellings $l_1 : \mathtt{oioio}$, $l_8 : \mathtt{ioioi}$ and $l_2 : \mathtt{oiooi}$, $l_7 : \mathtt{iooio}$.*

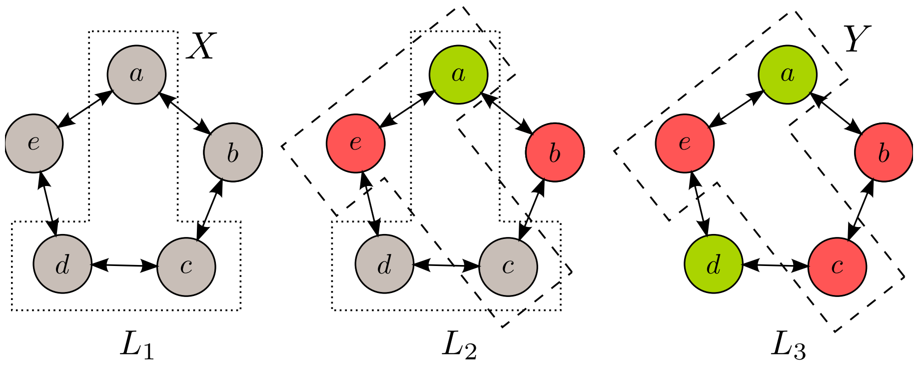



Figure 15: Example showing that *mincd* fails the triangle inequality

*We have* $(l_2 \ominus_{RHP} l_7) = (\{a,b,d,e\},\{a,b,d,e\}) \subsetneq (\{a,b,c,d,e\},\{a,b,c,d,e\}) = (l_1 \ominus_{RHP} l_8)$ *and also* $(l_2 \ominus_{HS} l_7) = \{a,b,d,e\} \subsetneq \{a,b,c,d,e\} = (l_1 \ominus_{HS} l_8)$ *so distance satisfying* **(SQDA)** *in both versions require* $d^{diff}_{\{a,d\}}(l_2,l_7) < d^{diff}_{\{a,d\}}(l_1,l_8)$ *but we have* $d^{diff}_{\{a,d\}}(l_2,l_7) = d^{diff}_{\{a,d\}}(l_1,l_8) = 2 \times diff(\texttt{in},\texttt{out})$.

We now try to fix these problems.

## 4.4 IRRESOLUTE PRODUCT MEASURES

The first idea is to keep the critical sets distance, but factor away the sensitivity to all the selected minimal critical sets by simply combining the values of $d^{diff}_{\mathcal{C}}$ for *all* $\mathcal{C} \in$ *Sem-mincrit*$(\mathcal{A})$. The following three spring to mind:

- $mincd^{diff}(S,T) \stackrel{\text{def}}{=} \min\{d^{diff}_{\mathcal{C}}(S,T) \mid \mathcal{C} \in \textit{Sem-mincrit}(\mathcal{A})\}$
- $sumcd^{diff}(S,T) \stackrel{\text{def}}{=} \sum\{d^{diff}_{\mathcal{C}}(S,T) \mid \mathcal{C} \in \textit{Sem-mincrit}(\mathcal{A})\}$
- $maxcd^{diff}(S,T) \stackrel{\text{def}}{=} \max\{d^{diff}_{\mathcal{C}}(S,T) \mid \mathcal{C} \in \textit{Sem-mincrit}(\mathcal{A})\}$

In the above way we escape the problem of selection of a particular minimal-critical set. It can be shown that all the above methods satisfy **(LAB$_{Sem}$)** Unfortunately, each of them has other drawbacks. *mincd* fails to satisfy the triangle inequality **(TRI)** as can be seen in the following example.

**Example 65.** *In Figure 15 are depicted three different complete labellings of an AF. (All other labellings can be obtained by applying authomorphism to one of them). It can be checked that the set of minimal critical sets contains set* $\{a,d,c\}$ *and four other sets obtained by turns. We have*

$$3 = mincd^{DM}(L_1,L_3) \not\leq mincd^{DM}(L_1,L_2) + mincd^{DM}(L_2,L_3) = d^{DM}_X(L_1,L_2) + d^{DM}_Y(L_2,L_3) = 1+1 = 2$$

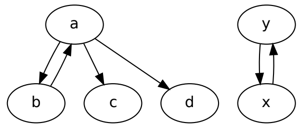



Figure 16: AF with two independent parts.

*sumcd* fails to satisfy **(IPI)** and **(COM$_{Sem}$)**. Informally we can explain it as follows. By adding a new argument or combining two independent parts of the framework the number of minimal critical sets multiplies which causes rapid increase of *sumcd* while those postulates require the distances of the framework to be proportional. The example below illustrates this problem for **(COM$_{Comp}$)**.

**Example 66.** *Reconsider AF $\mathcal{A}$ from Example 23 depicted again in Figure 16 which can be divided into two independent parts $\mathcal{A}_1$ and $\mathcal{A}_2$.*

$$\mathcal{A} = \{\{a, b, c, d, x, y\}, \{(a, b)(b, a)(a, c)(a, d), (x, y), (y, x)\}\}$$
$$\mathcal{A}_1 = \{\{a, b, c, d\}, \{(a, b)(b, a)(a, c)(a, d)\}\}$$
$$\mathcal{A}_2 = \{\{x, y\}, \{(x, y), (y, x)\}\}$$

*We have $Comp(\mathcal{A}_1) = \{\texttt{iooo}, \texttt{oiii}, \texttt{uuuu}\}$ and $Comp(\mathcal{A}_2) = \{\texttt{io}, \texttt{oi}, \texttt{uu}\}$ and $Comp(\mathcal{A}) = Comp(\mathcal{A}_1) \otimes Comp(\mathcal{A}_2)$. Any singleton set is a minimal critical set of the part, i. e.Comp-mincrit$(\mathcal{A}_1) = \{\{a\}, \{b\}, \{c\}, \{d\}\}$ and Comp-mincrit$(\mathcal{A}_2) = \{\{x\}, \{y\}\}$ and they can be combined in any way to form critical set of $\mathcal{A}$, i. e.*

$$\textit{Comp-mincrit}(\mathcal{A}) = \textit{Comp-mincrit}(\mathcal{A}_1) \otimes \textit{Comp-mincrit}(\mathcal{A}_2).$$

*We have*

$$\begin{aligned}
&mincd^{diff}(\texttt{iooooio}, \texttt{oiiioi}) = |\textit{Comp-mincrit}(\mathcal{A})| \times \textit{diff}(\texttt{in}, \texttt{out}) = \\
&8 \times \textit{diff}(\texttt{in}, \texttt{out}) > (4 \times \textit{diff}(\texttt{in}, \texttt{out})) + (2 \times \textit{diff}(\texttt{in}, \texttt{out})) = \\
&(|\textit{Comp-mincrit}(\mathcal{A}_1)| \times \textit{diff}(\texttt{in}, \texttt{out})) + (|\textit{Comp-mincrit}(\mathcal{A}_2)| \times \textit{diff}(\texttt{in}, \texttt{out})) = \\
&mincd^{diff}(\texttt{iooo}, \texttt{oiii}) + mincd^{diff}(\texttt{io}, \texttt{oi}).
\end{aligned}$$

All of them fail **(SQDA)** which can be seen going back to Example 64.

**Example 67.** *Consider situation from Example 64.*

*We have $(l_2 \ominus_{HS} l_7) \subsetneq (l_1 \ominus_{HS} l_8)$ and $(l_2 \ominus_{RHP} l_7) \subsetneq (l_1 \ominus_{RHP} l_8)$. What differentiate these qualitative distances is argument* c *which doesn't belong to any minimal Comp-critical set. There is simply not enough information to make a distinction. For all minimal Comp-critical sets $\mathcal{C}$, $d^{diff}_{\mathcal{C}}(l_2, l_7) = d^{diff}_{\mathcal{C}}(l_1, l_8) = 2 \times \textit{diff}(\texttt{in}, \texttt{out})$ and taking minimum, maximum or sum over them will still keep these distances equal.*

To summarise, *maxcd* can be used to solve the problem of sensitivity to the selection of minimal critical set and **(LAB$_{Sem}$)**. To satisfy **(SQDA)** we need another solution. It leads us to studies of critical sets which are small enough to satisfy **(IPI)** and big enough to satisfy **(SQDA)**.

## 4.5 ISSUE-BASED MEASURES

In this section we will take another approach to remove redundant arguments and satisfy **(IPI)**.

We want to capture the idea that the labels of two arguments are "tied together". For example in a simple 2-argument AF consisting of two arguments $a$ and $b$ mutually attacking each other, there may be two arguments but to all intents and purposes there is really only one "issue" at stake, and that is whether $a$ or $b$ (or none) should be accepted. We want to isolate these different issues which are being argued over.

**Definition 68.** *Given AF $\mathcal{A}$ and $\mathcal{X} \subseteq Labs(\mathcal{A})$ we define an equivalence relation $\equiv_{\mathcal{X}}$ over $Args_{\mathcal{A}}$ by setting $a \equiv_{\mathcal{X}} b$ iff* either *[$L(a) = L(b)$ for all $L \in \mathcal{X}(\mathcal{A})$]* or *[$L(a) = \neg L(b)$ for all $L \in \mathcal{X}(\mathcal{A})$]. Each $\equiv_{\mathcal{X}}$-equivalence class is called an* $\mathcal{X}$-issue, *and we denote the $\equiv_{\mathcal{X}}$-equivalence class to which $a$ belongs by $[a]_{\mathcal{X}}$. The set of all $\mathcal{X}$-issues we will denote by $\mathbb{I}_{\mathcal{X}}(\mathcal{A})$.*

**Proposition 69.** *Given AF $\mathcal{A}$ and $\mathcal{X} \subseteq Labs(\mathcal{A})$, $\equiv_{\mathcal{X}}$ is an equivalence relation over $Args_{\mathcal{A}}$.*

*Proof.* Reflexivity trivially holds. Symmetry follows from the symmetry of equality (=) and the fact that $\neg\neg l = l$ for $l \in \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$. Transitivity is also direct. Assume $a \equiv_{\mathcal{X}} b$ and $b \equiv_{\mathcal{X}} c$. If either $L(a) = L(b)$ or $L(b) = L(c)$ then $a \equiv_{\mathcal{X}} c$ holds trivially. In case when $L(a) = \neg L(b)$ and $L(b) = \neg L(c)$ we have $L(a) = L(c)$. □

For example, it can be checked that the *Comp*-issues for the AF in Figure 17 are $\{a, b, c, d\}$, $\{e, f\}$ and $\{g, h\}$. In the AF of Figure 13 there are 3 *Comp*-issues $\{a\}$, $\{b\}$ and $\{c\}$.

Within each $\equiv_{\mathcal{X}}$-equivalence class, considering just labellings from $\mathcal{X}$ there are at most 3 possible assignments of labels which can occur: either *(i)* all its elements are labelled `undec`, or *(ii)* all its elements are set to `in` or `out`, or *(iii)* the "inverse" labelling to *(ii)* occurs, in which those arguments labelled `in` become `out` and those labelled `out` are now `in`. Essentially each equivalence class acts as a single 3-valued argument.

Now the idea is to take a sum over one representative from each *Sem*-issue. It can be seen as taking a minimal hitting set of all the issues.

**Definition 70.** *Given $\mathcal{A}$ and $\mathcal{X} \subseteq Labs(\mathcal{A})$, a set of arguments $\mathcal{H} \subseteq Args_{\mathcal{A}}$ is* $\mathcal{X}$-issue critical (for $\mathcal{A}$) *iff for any $I \in \mathbb{I}_{\mathcal{X}}(\mathcal{A})$ we have $|I \cap \mathcal{H}| = 1$.*

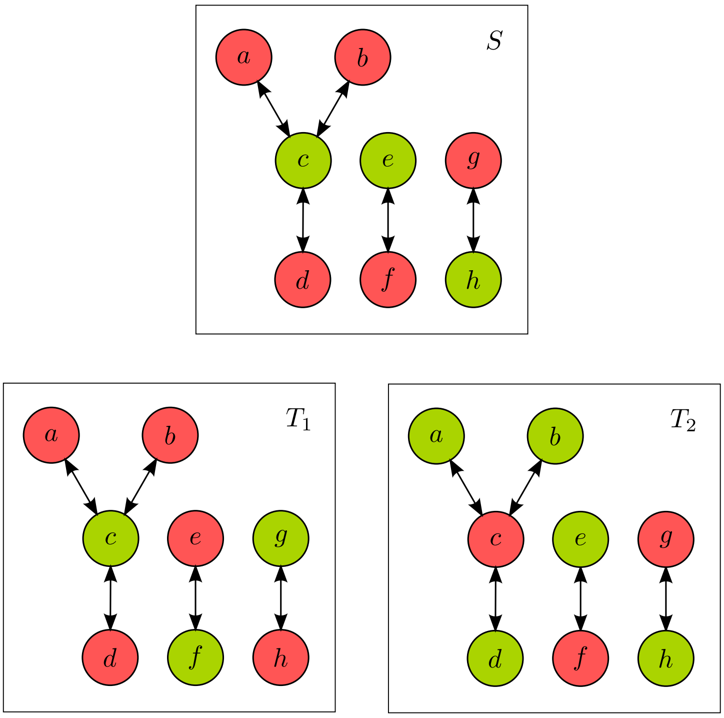



Figure 17: Source labelling S and 2 target labellings $T_1$, $T_2$

**Definition 71.** *An Sem-*issue-based method *is a product distance method with $\mathfrak{S}$ selecting Sem-issue critical set for all $\mathcal{A}$.*

In the example in Figure 17 we can check that $\mathcal{H} = \{a, e, g\}$ is *Comp*-issue critical set. We have $d^{diff}_{\mathcal{H}}(S, T_1) = 2 \times diff(\mathtt{in}, \mathtt{out})$ and $d^{diff}_{\mathcal{H}}(S, T_2) = diff(\mathtt{in}, \mathtt{out})$, as with the critical subsets approach of Section 4.3. For the example in Figure 13 we get just one *Comp*-issue critical set $\mathcal{H} = \{a, b, c\}$. We have $d^{diff}_{\mathcal{H}}(L_1, L_2) = 2 \times diff(\mathtt{in}, \mathtt{out}) = d^{diff}_{\mathcal{H}}(L_1, L_3)$. Thus according to the issue-based distance $d^{diff}_{\mathcal{H}}$, $L_2$ and $L_3$ are equidistant from $L_1$.

What is a relation between issue critical sets and minimal critical sets? We can see that in case of Figure 17 minimal critical sets and issue sets overlap but in case of Figure 13 minimal critical sets are subsets of a issue critical set.

The key to the answer is the following Lemma which states that arguments in a single issue share conflict type.

**Lemma 72.** *Given $\mathcal{X} \subseteq \mathtt{Labs}(\mathcal{A})$ and any two labellings $L_1, L_2 \in \mathcal{X}$ the following holds: (1) $a \in S(L_1, L_2)$ iff $[a]_{\mathcal{X}} \subseteq S(L_1, L_2)$, (2) $a \in H(L_1, L_2)$ iff $[a]_{\mathcal{X}} \subseteq H(L_1, L_2)$, (3) $a \in C(L_1, L_2)$ iff $[a]_{\mathcal{X}} \subseteq C(L_1, L_2)$.*

*Proof.* Since $S(L_1, L_2) = C(L_1, L_2) \setminus H(L_1, L_2)$ it is enough to prove (2) and (3). ($\leftarrow$) Follows trivially from the fact that any element is a member of its equivalence class. ($\rightarrow$) Take any argument $b \in [a]_{\mathcal{X}}$. We have two cases: 1. Positive correlation In this case $L_1(a) = L_1(b)$

and $L_2(a) = L_2(b)$. If $L_1(a) \neq L_2(a)$ then also $L_1(b) \neq L_2(b)$ and both $a, b \in C(L_1, L_2)$. If $L_1(a) = \neg L_2(a)$ then also $L_1(b) = \neg L_2(b)$ and both $a, b \in H(L_1, L_2)$; 2. Negative correlation In this case $L_1(a) = \neg L_1(b)$ and $L_2(a) = \neg L_2(b)$. If $L_1(a) \neq L_2(a)$ then also $\neg L_1(a) \neq \neg L_2(a)$ which is equivalent to $L_1(b) \neq L_2(b)$. Therefore both $a, b \in C(L_1, L_2)$. If $L_1(a) = \neg L_2(a)$ then $\neg L_1(b) = \neg(\neg L_2(b)) = L_2(b)$ and both $a, b \in H(L_1, L_2)$. □

This result has two important consequences. First we obtain the following relation to critical sets:

**Proposition 73.** *For any AF $\mathcal{A}$ and set of labellings $\mathcal{X} \subseteq Labs(\mathcal{A})$ if $\mathcal{H}$ is an $\mathcal{X}$-issue critical set then $\mathcal{H}$ is $\mathcal{X}$-critical set.*

*Proof.* Assume $\mathcal{H}$ is a $\mathcal{X}$-issue critical set of $\mathcal{A}$. We will show the contraposition, for any $L_1, L_2 \in \mathcal{X}$ if $L_1 \neq L_2$ then $lab_1[\mathcal{H}] \neq lab_2[\mathcal{H}]$. From $L_1 \neq L_2$ follows that there exists $a \in Args_{\mathcal{A}}$ such that $L_1(a) \neq L_2(a)$. Since $\mathcal{H}$ is $\mathcal{X}$-issue set of $\mathcal{A}$ there exists $b \in [a]_{\mathcal{X}}$. We have $a \in C(L_1, L_2)$ and by Lemma 72 $b \in [a]_{\mathcal{X}} \subseteq C(L_1, L_2)$. For such $b$, $lab_1[\mathcal{H}](b) \neq lab_2[\mathcal{H}](b)$. □

This result justifies the name $\mathcal{X}$-issue 'critical', since they are indeed a special kind of $\mathcal{X}$-critical sets. Second, for a label distance *diff* satisfying **(scs)** (and **(sym)**, **(ref)** which is our basic assumption) the choice of a particular issue set does not matter.

**Proposition 74.** *Let $d_{\mathcal{H}}, d_{\mathcal{H}'}$ be two Sem-issue-based measures defined via diff. If diff satisfies* **(scs)** *(and* **(sym)**, **(ref)***) then $d_{\mathcal{H}} \equiv d_{\mathcal{H}'}$.*

*Proof.* Both $d_{\mathcal{H}}, d_{\mathcal{H}'}$ take a sum over all issues possibly selecting a different member of an issue. It is enough to show that for all labellings $L_1, L_2 \in Sem$ and any two members of an issue $a, b \in I$ we have $diff(L_1(a), L_2(a)) = diff(L_1(b), L_2(b))$. From Lemma 72 we know that $a, b$ share the same conflict type. We have three cases:

CASE 1: $a, b \notin C(L_1, L_2)$ (no conflict)
$\mathtt{diff(in, in)} = \mathtt{diff(out, out)} = \mathtt{diff(undec, undec)} = 0$ by **(ref)**.

CASE 2: $a, b \in H(L_1, L_2)$
by **(sym)** $\mathtt{diff(in, out)} = \mathtt{diff(out, in)}$.

CASE 3: $a, b \in S(L_1, L_2)$
$\mathtt{diff(in, undec)} = \mathtt{diff(undec, in)} = \mathtt{diff(out, undec)} = \mathtt{diff(undec, out)}$ by **(sym)** and **(scs)**.

□

Thus issue-based measure can be thought of as a critical-set based measure which chooses from among a particular class of critical sets, viz. those which contain one argument from each issue. However the critical set chosen need not be a minimal one, i.e., an element of $Comp\text{-}mincrit(\mathcal{A})$, as can be seen already in the AF of Figure 13. We immediately inherit general critical set results from the previous section - Theorem 52 and Theorem 54.

We have usually more arguments in one issue and hence many issue critical sets. One may suspect to encounter the similar problem to the one with selection of the minimal critical set. Indeed we could use any of the techniques discussed for irresolute product measures but a more promising way is to use a label distance *diff* satisfying **(sym)** and **(scs)** which makes the precise choice of arguments for each issue irrelevant. For such *diff* any issue-based method can be seen as max over all issue-based methods. For that reason issue-based method avoid the problems of minimal critical set method in a similar way to the irresolute product methods.

**Proposition 75.** *Any Issue-based distance method defined via diff satisfying* **(sym)**, **(ref)** *and* **(scs)** *satisfies* **($\mathbf{LAB}_{Sem}$)**.

It is a consequence of the following property.

**Lemma 76.** *Let $\mathcal{A}_1, \mathcal{A}_2$ be two labelling equivalent AFs (under semantics Sem) via bijection $g$. Then $g$ preserves Sem-issues, i.e.* $\{g(I) \mid I \in \mathbb{I}_{Sem}(\mathcal{A}_1)\} = \mathbb{I}_{Sem}(\mathcal{A}_2)$.

*Proof.* We have $g(L)(g(a)) = L(g^{-1}(g(a))) = L(a)$ so $a \equiv_{Sem(\mathcal{A}_1)} b$ iff $g(a) \equiv_{g(Sem(\mathcal{A}_1))} g(b)$ which is equivalent to $g(a) \equiv_{Sem(\mathcal{A}_2)} g(b)$ because $g$ (extended) is a bijection between $Sem(\mathcal{A}_1)$ and $Sem(\mathcal{A}_2)$. □

We prove Proposition 75.

*Proof.* Let $\mathcal{A}_1, \mathcal{A}_2$ be two labelling equivalent AFs (under semantics *Sem*) via bijection $g$ and $d_{\mathcal{A}_1}, d_{\mathcal{A}_2}$ issue-based distance methods defined by *diff* satisfying **(sym)**, **(ref)** and **(scs)**. By Proposition 74 the selection of the particular *Sem*-issue critical set $\mathcal{H}$ does not matter. By Lemma 76 $g(\mathcal{H})$ is an *Sem*-issue critical set of $\mathcal{A}_2$. We calculate

$$\begin{aligned}
d_{\mathcal{A}_1}(L_1, L_2) &= \sum_{a \in \mathcal{H}} diff(L_1(a), L_2(a)) \\
&= \sum_{a \in \mathcal{H}} diff(L_1(g^{-1}(g(a))), L_2(g^{-1}(g(a)))) \\
&= \sum_{b \in g(\mathcal{H})} diff(L_1(g^{-1}(b)), L_2(g^{-1}(b))) \\
&= \sum_{b \in g(\mathcal{H})} diff(g(L_1)(b), g(L_2)(b)) \\
&= d_{g(\mathcal{H})}(g(L_1), g(L_2)) = d_{\mathcal{A}_2}(g(L_1), g(L_2)).
\end{aligned}$$

□

The following proposition holds.

**Proposition 77.** *Let Sem be any complete-based semantics. All issue-based methods defined via diff satisfying* **(sym)**, **(ref)** *and* **(scs)** *satisfy* **(IPI)**.

*Proof.* Let $\mathcal{A} = (Args, \rightharpoonup)$ be an arbitrary AF and let $\mathcal{A}^+$ be any framework obtained from $\mathcal{A}$ by adding a single new argument $b \notin Args$ along with a **single** attack $a \rightharpoonup b$ from some $a \in Args$, $Sem$ be any complete-based semantics and $d_{\mathcal{A}}$, $d_{\mathcal{A}^+}$ a distance measures assigned by issue-based method defined via *diff* satisfying **(sym)**, **(ref)** and **(scs)**. For all $L \in Sem(\mathcal{A}^+)$ it holds that $L(a) = \neg L(b)$ because $Sem$ is complete-based and $a$ is the only attacker of $b$. Therefore $a$ and $b$ belong to the same issue. By Proposition 74 the choice of the labelling critical set $\mathcal{H}$ of $\mathcal{A}^+$ does not matter therefore we can assume $d_{\mathcal{A}^+} = d_{\mathcal{H}}$ for $\mathcal{H} \subseteq Args$. Since restriction $Sem(\mathcal{A}^+) \ni L \mapsto L[Args] \in Sem(\mathcal{A})$ is a bijection $\mathcal{H}$ is also labelling critical set of $\mathcal{A}$ therefore again by Proposition 74 $d_{\mathcal{A}} = d_{\mathcal{H}}$. □

**Proposition 78.** *All Sem-issue-based methods defined via diff satisfying* **(sym)**, **(ref)** *and* **(scs)** *satisfy* **(COM$_{Sem}$)**.

*Proof.* Let A, B be semantically independent partition of $\mathcal{A}$ under semantics $Sem$, i.e. $Sem(\mathcal{A}) = Sem(\mathcal{A}_{\downharpoonright A}) \otimes Sem(\mathcal{A}_{\downharpoonright B})$, and $d_{\mathcal{A}}, d_{\mathcal{A}_{\downharpoonright A}}, d_{\mathcal{A}_{\downharpoonright B}}$ distance measures assigned by $Sem$-issue-based method defined via *diff* satisfying **(sym)**, **(ref)** and **(scs)** to $\mathcal{A}, \mathcal{A}_{\downharpoonright A}, \mathcal{A}_{\downharpoonright B}$ respectively. We need to show that $d_{\mathcal{A}}(L_1, L_2) = d_{\mathcal{A}_{\downharpoonright A}}(L_1[A], L_2[A]) + d_{\mathcal{A}_{\downharpoonright B}}(L_1[B], L_2[B])$.

By Proposition 74 it is enough to show that $\mathbb{I}(\mathcal{A}) = \mathbb{I}(\mathcal{A}_{\downharpoonright A}) \cup \mathbb{I}(\mathcal{A}_{\downharpoonright B})$. It is true with the exception of constant issues which we define for any $\mathcal{A}$ as follows: $\mathtt{uI}(\mathcal{A}) = \{a \in Args_{\mathcal{A}} \mid L(a) = \mathtt{undec}$ for all $L \in Sem(\mathcal{A})\}$ and $\mathtt{ioI}(\mathcal{A}) = \{a \in Args_{\mathcal{A}} \mid L(a) = l \in \{\mathtt{in}, \mathtt{out}\}$ for all $L \in Sem(\mathcal{A})\}$. It is easy to check those set are issues (if non-empty) for any $\mathcal{A}$. The constant issues of semantically independent parts will join but since they are constant they do not contribute to the distance and can be ignored. Non-constant issues stay the same. For all $a, b \in Args \setminus (\mathtt{uI}(\mathcal{A}) \cup \mathtt{ioI}(\mathcal{A}))$ we have two cases:

$a, b \in A$ (OR SIMILARLY $a, b \in B$) Since by semantic independence $Sem(\mathcal{A}_{\downharpoonright A}) = Sem(\mathcal{A})[A]$, $a \equiv_{Sem(\mathcal{A})} b$ iff $a \equiv_{Sem(\mathcal{A}_{\downharpoonright A})} b$.

$a \in A, b \in B$ Since $a, b$ are non-constant there exist $L_1, L_2 \in Sem(\mathcal{A}_{\downharpoonright A})$ such $L_1(a) \neq L_2(a)$, and $L_3 \in Sem(\mathcal{A}_{\downharpoonright B})$. It is not the case that $a \equiv_{Sem}(\mathcal{A}) b$ because by semantic independence both $L_1 \cup L_3, L_2 \cup L_3 \in Sem(\mathcal{A})$ and those labellings neither label $a, b$ the same nor the opposite.

□

We finally verify that this time **(SQDA)** is satisfied.

**Theorem 79.** *The issue-based method* $d$ *defined via* $diff$ *function satisfying* **(dd)**, **(scs)** *and* **(hcs+)**, *satisfies* **(SQDA$_{\ominus_{RHP}}$)**. *The issue-based method* $d$ *defined via* $diff$ *function satisfying* **(gcs)** *and* **(dd)**, *satisfies* **(SQDA$_{\ominus_{HS}}$)**.

*Proof.* Let us remind reformulations of product distance measures $d_{\mathcal{A}r}$ defined via $diff$ in terms of conflict sets from Theorem 55. For $C_L = (L_1 \ominus_{HS} L_2) \subseteq (N_1 \ominus_{HS} N_2) = C_N$ we have

$$d_{\mathcal{A}r}(N_1, N_2) = c \times |\mathcal{A}r \cap (C_N \setminus C_L)| + d_{\mathcal{A}r}(L_1, L_2).$$

For $\langle C_L, H_L \rangle = (L_1 \ominus_{RHP} L_2) \subseteq (N_1 \ominus_{RHP} N_2) = \langle C_N, H_N \rangle$ we have

$$\begin{aligned} d(N_1, N_2) = &(c_h - c_s) \times |\mathcal{A}r \cap (H_N \setminus H_L)| \\ &+ c_s \times |\mathcal{A}r \cap (C_N \setminus C_L)| + d(L_1, L_2). \end{aligned}$$

For the assumption about *diff* all constants $c, c_s$ and $(c_h - c_s)$ (note **(hcs+)** instead of **(hcs)**) are positive. We need to show that the respective set intersections are non-empty for issue critical set $\mathcal{A}r$. For strict inclusion $(L_1 \ominus_{HS} L_2) \subset (N_1 \ominus_{HS} N_2)$ we have $C_N \setminus C_L \neq \emptyset$. Let $a \in C_N \setminus C_L$. From Lemma 72 it follows tha $[a]_{Sem} \subseteq C_N \setminus C_L$ and by the issue criticality of $\mathcal{A}r$ we have $|\mathcal{A}r \cap (C_N \setminus C_L)| \geqslant 1$. Similarly, from $\langle C_L, H_L \rangle = (L_1 \ominus_{RHP} L_2) \subset (N_1 \ominus_{RHP} N_2) = \langle C_N, H_N \rangle$ it follows that $C_N \setminus C_L \neq \emptyset$ or $H_N \setminus H_L \neq \emptyset$ and we reason in the same way. □

Since the set of all arguments contain any issue critical set we obtain Proposition 44 and Proposition 46 as corollaries of the above theorem.

## 4.6 CONCLUSION AND RELATED WORK

### 4.6.1 *Conclusions*

In this chapter we have analysed product distance measures, distance measures of a special form, which can be specified by label distance function *diff* and selection function $\mathfrak{S}$.

We identified the properties of *diff* and $\mathfrak{S}$ on which depend satisfaction of the postulates from the previous chapter.

We have proven sufficient conditions to satisfy distance measure postulates (Metric and Intuition-based) (Table 2) and shown that in cases we are interested in they are also necessary (Example 45

The incompatibility between **(SQDA$_{\ominus_{RHP}}$)** and **(QDA$_{\ominus_{RHP}}$)** in context of product measures boils down to incompatibility between **(gcs)** and **(hcs+)**. We have proposed to use discrete metrics DM or refined discrete metrics rDM for satisfying Hamming set or Refined Hamming Pair related postulates respectively.

The restriction on the selected set of argument vary from none, through critical set to hitting set of issues. The set of all arguments satisfy all those restrictions.

| Property | *diff* | $\mathfrak{S}$ | Results |
|---|---|---|---|
| **(REF)** | **(ref)** | - | Theorem 52 |
| **(DD)** | **(dd)** | **(crit)** | Theorem 52 |
| **(SYM)** | **(sym)** | - | Theorem 52 |
| **(TRI)** | **(tri)** | - | Theorem 52 |
| **(BTW$_\triangleright$)** | **(dd)** | **(crit)** | Theorem 54 |
| **(BTW$_\blacktriangleright$)** | **(dd), (hcs+)** | **(crit)** | Theorem 54 |
| **(SQDA$_{\ominus_{RHP}}$)** | **(dd), (scs),(hcs+)** | **(issue)** | Theorem 79 |
| **(SQDA$_{\ominus_{HS}}$)** | **(dd), (gcs)** | **(issue)** | Theorem 79 |
| **(QDA$_{\ominus_{RHP}}$)** | **(scs), (hcs)** | - | Theorem 55 |
| **(QDA$_{\ominus_{HS}}$)** | **(gcs)** | - | Theorem 55 |

Table 2: Restrictions placed by Metric and Intuition-based postulates

In contrast to distance measure postulates which restricted label distance function *diff* and the properties of the set of arguments selected by selection function $\mathfrak{S}$, the distance method postulates restrict the way selection function assigns sets of important arguments to different AFs (Table 3).

| Property | Restriction on $\mathfrak{S}$ | Results |
|---|---|---|
| **(COM$_{Sem}$)** | $\mathfrak{S}(\mathcal{A}) = \mathfrak{S}(\mathcal{A}_{\downarrow A}) \cup \mathfrak{S}(\mathcal{A}_{\downarrow B})$ (3) | Lemma 60 |
| **(LAB$_{Sem}$)** | $g(\mathfrak{S}(\mathcal{A}_1)) = \mathfrak{S}(\mathcal{A}_2)$ | |
| **(IPI$_{Comp}$)** | $\mathfrak{S}(\mathcal{A}) = \mathfrak{S}(\mathcal{A}^+)$ (2) | Lemma 58 |

Table 3: Restrictions placed by Compositionality and Equivalence

Choosing all arguments of the AF satisfy **(COM$_{Sem}$)** and **(LAB$_{Sem}$)** but fails **(IPI)**. It rises the question how to choose the particular critical set or hitting set of issues in a systematic way which satisfy the restriction posed on distance method? We started with minimal critical sets and continued with issue sets (which are the minimal hitting sets of all issues). In both cases there exists several ways to define selection functions which satisfy **(COM$_{Sem}$)** and **(IPI)**.

When selecting among minimal critical sets we identified two problems. First, there is no choice satisfying **(LAB$_{Sem}$)**. We can interpret it as follows. All minimal critical sets are indistinguishable considering the set of feasible labellings, therefore commitment to one particular set needs to be made upon external information not contained in the set of labellings. At the same time **(LAB$_{Sem}$)** postulates that distance measure should depend solely on information included in the set of feasible labellings. We proposed to address that problem by combining all distance measures of the class by taking their minimum, maximum or sum. We call the family of distance obtained in this way irresolute distance measures. While all of them satisfy **(LAB$_{Sem}$)**, taking minimum fails **(TRI)** and taking sum fails **(IPI)** and **(COM)** therefore taking a maximum is a preferred method.

The second problem of minimal critical set measures that cannot be solved by irresolute measures is a failure of **(SQDA)**. The reason is that minimal critical set ignores too many arguments. We solve it by defining issues which partition all arguments into groups which behaves similarly. We showed that any hitting set covering all the issues, i. e.the set of argument which has non-empty intersection with any issue, is a critical set. Again we considered minimal sets which we call issue sets. Despite the fact that there exists more such sets, for any *diff* function satisfying **(sym)** and **(scs)** choice of the particular issue set doesn't matter since it defines the same function.

Combining all the results above we propose three concrete distance methods, each in two variants depending whether one commits to qualitative distance defined by Hamming set or Refined Hamming pair.

The issue-based measure defined via *diff* function satisfying **(scs)** and **(sym)** is the only proposal which satisfy all the postulates. We define $\mathrm{IBd}$ and $\mathrm{rIBd}$ by plugging discrete $\mathrm{DM}$ and refined discrete metric $\mathrm{rDM}$ respectively over some issue critical set $\mathcal{H}$, i.e. $\mathrm{IBd} = d_{\mathcal{H}}^{\mathrm{DM}}, \mathrm{rIBd} = d_{\mathcal{H}}^{\mathrm{rDM}}$. Note that selection of the issue critical set doesn't matter.

**(SQDA)** happened to be quite strong. If one is willing to give it up the maximum over all minimal critical set measures is a valid way to satisfy **(IPI)**. We define $\mathrm{MMCSd} = mincd^{\mathrm{DM}}, \mathrm{rMMCSd} = mincd^{\mathrm{rDM}}$.

If **(IPI)** is not a vital request taking full sum measure is worth considering because of it simplicity and use Hamming distance $\mathrm{Hd}$ or Refined Hamming distance $\mathrm{rHd}$.

The properties satisfied by each of the above methods is gathered in Table 4.

### 4.6.2 *Related Work*

We have initiated the investigation of the notion of distance between two reasonable evaluations of an argument graph. While this issue has been investigated in non-argument based accounts of both belief revision Delgrande (2004); Lehmann et al. (2001); Peppas et al. (2004), in judgement aggregation Miller and Osherson (2009); Pigozzi (2006), and in abstract preferences Baigent (1987), to our knowledge we are the first to study it in the context of formal argumentation theory.

Coste-Marquis et al. presented an approach for merging multiple Dung-style argumentation graphs presented by multiple agents Coste-Marquis et al. (2007). The authors use a combination of graph expansion, distance calculation and voting in order to arrive at a single

| Property | Hd | rHd | MMCSd | rMMCSd | IBd | rIBd |
|---|---|---|---|---|---|---|
| **(BTW$_{\triangleright}$)** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **(BTW$_{\blacktriangleright}$)** | × | ✓ | × | ✓ | × | ✓ |
| **(SQDA$_{\ominus_{HS}}$)** | ✓ | × | × | × | ✓ | × |
| **(SQDA$_{\ominus_{RHP}}$)** | × | ✓ | × | × | × | ✓ |
| **(QDA$_{\ominus_{HS}}$)** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **(QDA$_{\ominus_{RHP}}$)** | × | ✓ | × | × | × | ✓ |
| **(COM$_{Sem}$)** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **(LAB$_{Sem}$)** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **(IPI)** | × | × | ✓ | ✓ | ✓ | ✓ |

Table 4: Summary of Properties of proposed Distance Measures. All included distances satisfy **(REF)**,**(DD)**,**(SYM)** and **(TRI)**.

argumentation framework. This work addresses a fundamentally different problem, since agents may differ over which arguments exist, or which arguments attack which other arguments. In contrast, in our work, we assume that all arguments are available to all agents (e.g. as in a jury hearing), and that the attack relation is not a subjective matter (i.e. it is objectively determined by the underlying logical system, as is for instance the case in Gorogiannis and Hunter (2011); Prakken (2010) ). In other words, the distance measures introduced by Coste-Marquis et al are between different graphs, and thus address a fundamentally different problem (e.g. the edit distance, which measures the number of insertions/deletions of attacks needed to turn one entire argument graph into another). Our notion of distance, on the other hand, is aimed at quantifying disagreement over the evaluation of the *given* graph; i.e. it is the distance between different evaluations of the given evidence, not between different perspectives on what the evidence is.

# 5 USING DISTANCES FOR AGGREGATION IN ABSTRACT ARGUMENTATION

## 5.1 INTRODUCTION

Individuals presented with the same set of conflicting arguments might take different rational positions. In such a situation one often faces the problem of how to aggregate them into a collective one. This problem has been explored in a number of recent papers Booth et al. (2014); Caminada and Pigozzi (2011); Rahwan and Tohmé (2010) which employ techniques from *judgement aggregation* (JA) List and Puppe (2009) to the problem of aggregating 3-valued argument labellings. These works have shown that, as with classical JA, it is not possible to define general aggregation operators that satisfy a number of seemingly mild constraints while ensuring collective rationality of the outcome.

One way of getting around this problem is to first use an *initial* aggregation operator, which intuitively can be thought of as a *gold standard* operator that satisfies a number of basic postulates, without always yielding collectively rational results and then to *repair* the result of this operator in the cases when it does not give a collectively rational outcome.

In the argumentation setting, Caminada and Pigozzi suggested one way to carry out such a repair, using what they called the *down-admissible* and *up-complete* procedures (2011). In the JA setting, another way to carry out such a repair is to use one of the *distance-based solution methods* that were studied by Miller and Osherson (hereafter MO) (2009) within the framework of binary judgement aggregation. As the name suggests these depend on a provided notion of distance measure between binary judgement sets.

In this chapter we show how this option can also be used in the argumentation setting. We first modify the MO framework for our purposes, using the distances measures defined in Chapter 4. We thus illustrate the usefulness of these distance measures.

Along the way we generalise and extend some of the results. For example some of the MO aggregation methods require a distance to be defined between any two arbitrary labellings of an argumentation framework, whereas the most interesting distance measures of Chapter 4, such as the *issue-based* distance, are defined only between *Sem*-labellings. We thus extend the definition of these distances to apply to arbitrary labellings. Although in Chapter 4 we defined distance generally between a selected set of *Sem*-labellings, in this chapter we will assume the set of feasible labellings is given by complete semantics. We make a link between the **(IPI)** postulate and the problem of agenda manipulation in JA. Finally, we show that the *down-admissible* and *up-complete* procedures Caminada and Pigozzi (2011) can be viewed as a special case of one of the MO methods.

## 5.2 AGGREGATION PRELIMINARIES

We first turn to aggregation. We assume a set of agents $Ag = \{1, \dots, n\}$ (with $n \geqslant 2$) is fixed. In our argumentation setting, the roles of the agenda and the judgement set are filled by the AF $\mathcal{A}$ and $\mathcal{A}$-labelling respectively.

**Definition 80.** *Let $\mathcal{A}$ be an AF. An $\mathcal{A}$-*profile *is an $n$-tuple of $\mathcal{A}$-labellings $\mathbf{L} = (L_1, \dots, L_n)$. If every $L_i$ is a complete $\mathcal{A}$-labelling then we call $\mathbf{L}$ a* complete *$\mathcal{A}$-profile.*

What we seek is a way to construct aggregation operators that, given any $\mathcal{A}$-profile $\mathbf{L}$ as input, return a set of $\mathcal{A}$-labellings $F_{\mathcal{A}}(\mathbf{L})$. Note that an aggregation operator is always defined in a context of some specific AF $\mathcal{A}$. More generally we are interested in an aggregation *method* that given any context AF $\mathcal{A}$ will return in a systematic way an aggregation operator for $\mathcal{A}$.

**Definition 81.** *Let $\mathcal{A}$ be an AF. An* (irresolute) aggregation operator *(for $\mathcal{A}$) is a function $F_{\mathcal{A}}$ that assigns, to each $\mathcal{A}$-profile $\mathbf{L}$ a set $F_{\mathcal{A}}(\mathbf{L}) \subseteq$ Labs$(\mathcal{A})$. An* aggregation method *is a mapping that, given any context AF $\mathcal{A}$, returns an aggregation operator $F_{\mathcal{A}}$ for $\mathcal{A}$.*

From now on we drop the *irresolute* and just say *aggregation operator*. In previous work on aggregation in argumentation the output is usually taken to be a single labelling, but we relax that here. When important, we call an operator which returns always a singleton set a *resolute* operator. Also note for each $\mathcal{A}$, $F_{\mathcal{A}}$ is defined for *all* $\mathcal{A}$-profiles (not necessarily just the complete ones), and that the output of $F_{\mathcal{A}}(\mathbf{L})$ is allowed to be *any* subset of $\mathcal{A}$-labellings. Ideally, of course, we would like the output to consist only of *complete* $\mathcal{A}$-labellings, i.e., we want the following to hold:

> **Collective Rationality** For all $\mathcal{A}$ and $\mathcal{A}$-profiles $\mathbf{L}$, $F_{\mathcal{A}}(\mathbf{L}) \subseteq$ *Comp*$(\mathcal{A})$

## 5.3 MILLER AND OSHERSON AGGREGATION METHODS.

Miller and Osherson (2009) describe a framework for using *distance measures* to define aggregation methods in binary judgement aggregation. In that setting agents evaluate a set of logical propositions called an *agenda* by providing a *judgement set* which is an assignment of either True or False to each proposition in the agenda. Their framework requires specification of two things:

1. An *initial* ***resolute*** *aggregation method* M that is able to give collectively rational answers in simple cases. In their case they only considered the proposition-wise majority method, but in principle any M can be used. The intuition is that if the outcome produced by M happens to be collectively rational then there is no need to choose a different outcome.

2. A measure of *distance* between any two judgement sets of any given agenda. This measure is assumed to be a *metric*.

We can extend a distance measure $d_{\mathcal{A}}$ so that it also returns distance from an $\mathcal{A}$-profile to an $\mathcal{A}$-labelling, as well as distance between two $\mathcal{A}$-profiles. For $L \in Labs(\mathcal{A})$ and profiles $\mathbf{L} = (L_1, \dots, L_n)$, $\mathbf{L}' = (L'_1, \dots, L'_n) \in Labs(\mathcal{A})^n$ we define

$$d_{\mathcal{A}}(\mathbf{L}, L) = \sum_{i=1}^{n} d_{\mathcal{A}}(L_i, L), \qquad d_{\mathcal{A}}(\mathbf{L}, \mathbf{L}') = \sum_{i=1}^{n} d_{\mathcal{A}}(L_i, L'_i).$$

Also required for MO is the notion of an M-*consistent* profile. These are the profiles that, when passed to M, result in a collectively rational outcome.

**Definition 82.** *Let $\mathcal{A}$ be an AF, $\mathbf{L} \in Labs(\mathcal{A})^n$ and M a resolute aggregation method. Then $\mathbf{L}$ is* M-consistent (for $\mathcal{A}$) *iff $M_{\mathcal{A}}(\mathbf{L}) \in Comp(\mathcal{A})$. We denote by $\mathrm{Cons}_{\mathcal{A}}(M)$ the set of M-consistent $\mathcal{A}$-profiles, and by $\mathrm{Cons}_{\mathcal{A}}(M, Comp)$ the set $\mathrm{Cons}_{\mathcal{A}}(M) \cap Comp(\mathcal{A})^n$.*

MO describe four different ways in which all the above ingredients can be combined, resulting in four classes of aggregation methods which we now describe. First some notation: For any function $f : X \to Y$ and sets $D \subseteq X$, $C \subseteq Y$ we denote the image of D by $f(D) = \{f(x) \mid x \in D\}$ and the inverse image of C by $f^{-1}(C) = \{x \in X \mid f(x) \in C\}$. The subset of D for which f obtains its minimal value is returned by the operator $\arg\min_{x \in D} f(x) = \{x \in D \mid f(x) \leqslant f(x') \text{ for all } x' \in D\}$.

**Definition 83** (Miller and Osherson (2009))**.** *Let d be distance method and M a resolute aggregation method. The four aggregation methods $\mathtt{Prototype}^d$, $\mathtt{Endpoint}^{M,d}$, $\mathtt{Full}^{M,d}$ and $\mathtt{Output}^{M,d}$ are defined by setting, for each AF $\mathcal{A}$ and $\mathcal{A}$-profile $\mathbf{L}$:*

$$\mathtt{Prototype}^d_{\mathcal{A}}(\mathbf{L}) = \underset{L \in Comp(\mathcal{A})}{\arg\min}\ d_{\mathcal{A}}(\mathbf{L}, L)$$

$$\mathtt{Endpoint}^{M,d}_{\mathcal{A}}(\mathbf{L}) = \underset{L \in Comp(\mathcal{A})}{\arg\min}\ d_{\mathcal{A}}(M_{\mathcal{A}}(\mathbf{L}), L)$$

$$\mathtt{Full}^{M,d}_{\mathcal{A}}(\mathbf{L}) = M_{\mathcal{A}}\left(\underset{\mathbf{L}' \in \mathrm{Cons}_{\mathcal{A}}(M, Comp)}{\arg\min}\ d_{\mathcal{A}}(\mathbf{L}, \mathbf{L}')\right)$$

$$\mathtt{Output}^{M,d}_{\mathcal{A}}(\mathbf{L}) = M_{\mathcal{A}}\left(\underset{\mathbf{L}' \in \mathrm{Cons}_{\mathcal{A}}(M)}{\arg\min}\ d_{\mathcal{A}}(\mathbf{L}, \mathbf{L}')\right)$$

All four MO aggregation methods minimise distance to ensure the collective outcome is rational. The `Prototype` and `Endpoint` methods minimise the distance over all complete labellings. `Prototype` returns the complete labellings closest to the profile $\mathbf{L}$. `Endpoint` returns the complete labellings closest to the labelling returned by initial aggregator $M_{\mathcal{A}}(\mathbf{L})$, which possibly is not complete. The `Full` and `Output` methods select the $M$-consistent profiles closest to $\mathbf{L}$ and then applies $M_{\mathcal{A}}$ to them. The difference bettween these two is that `Output` performs its selection from among *all* $M$-consistent profiles, while `Full` selects only from those that are, in addition, themselves complete.

Some observations in these definitions:

(i) All four aggregation methods are potentially irresolute.

(ii) `Prototype` does not require an initial aggregator $M$, only a distance method $d$. The other three all rely on $M$.

(iii) A distance method $d$ used in `Endpoint` and `Output` needs to return the distance between *all* labellings. In contrast, for `Prototype` and `Full` it is enough that $d$ is defined only between complete labellings.

If we want to apply MO to our problem of aggregating labellings we need to instantiate the two parameters $M$ and $d$. Let's look at each in turn.

### 5.3.1 *Initial Aggregation Methods*

A family of resolute aggregation methods capturing many operators in a uniform way has been defined in Booth et al. (2014), namely the *interval aggregation methods*.

Formally, let $Int_n$ be the set of *intervals* of non-zero length in $\{0, 1, \dots, n\}$, i.e., $Int_n = \{(k, l) \mid k < l,\ k, l \in \{0, 1, \dots, n\}\}$. Let $Y \subseteq Int_n$ be some subset of distinguished intervals in $Int_n$. Then we define aggregation method $F^Y$ by setting, for each $\mathcal{A}$, $\mathcal{A}$-profile $\mathbf{L}$ and $a \in Args_{\mathcal{A}}$:

$$[F^Y_{\mathcal{A}}(\mathbf{L})](a) = \begin{cases} \mathtt{x} & \text{if } \mathtt{x} \in \{\mathtt{in}, \mathtt{out}\} \text{ and } (|V^{\mathbf{L}}_{a:\neg \mathtt{x}}|, |V^{\mathbf{L}}_{a:\mathtt{x}}|) \in Y \\ \mathtt{undec} & \text{otherwise,} \end{cases}$$

where, for any $x \in \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$, $V^{\mathbf{L}}_{a:x}$ denotes $\{i \in Ag \mid L_i(a) = x\}$. A particular member of this family, which we use in our examples, is the *credulous* aggregation method Caminada and Pigozzi (2011),

$$[\mathtt{cio}_{\mathcal{A}}(\mathbf{L})](a) = \begin{cases} \mathtt{in} & \text{if } \exists L \in \mathbf{L} : L(a) = \mathtt{in} \text{ and} \\ & \nexists L \in \mathbf{L} : L(a) = \mathtt{out}, \\ \mathtt{out} & \text{if } \exists L \in \mathbf{L} : L(a) = \mathtt{out} \text{ and} \\ & \nexists L \in \mathbf{L} : L(a) = \mathtt{in}, \\ \mathtt{undec} & \text{otherwise,} \end{cases}$$

which can be also defined by taking $Y_{cio} = \{(0, l) \in Int_n \mid l > 0\}$.

The credulous method cio returns a collective label of in (resp. out) to an argument if at least one agent votes for in (resp. out) while none vote for the opposite label out (resp. in). Otherwise it returns undec.

Interval methods may be characterised by a number of postulates such as *Anonymity*, *Unanimity* and *AF-Independence* (the collective label of $a$ is calculated independently of which other arguments might be present or absent from $\mathcal{A}$) Booth et al. (2014). However, despite their simplicity, there is *no* interval method that satisfies *Collective Rationality*. For the sake of simplicity, from now on we stick to the cio method, but in principle any interval method could be used as an initial aggregation operator (we refer the reader to Booth et al. (2014) for the details).

### 5.3.2 *Extension to the Set of All Labellings*

We would like to use one of the distance methods defined in Chapter 4 but there is one problem. It was assumed distances were defined only between complete labellings (or more general the set of feasible labellings). In this case issue-based distance $\mathtt{id}^{diff}$ defined via *diff* satisfying **(scs)** seems to be a good candidate to use in MO. But two of the MO aggregation methods, namely Endpoint and Output, require distance to be defined between *all* labellings, and in this case Theorem 52 gives us a problem, for it tells us that the **only** way for any product distance method to yield a metric over the whole set $Labs(\mathcal{A})$ is if $\mathfrak{S}(\mathcal{A}) = Args_{\mathcal{A}}$. The question is, is there any alternative way to define a distance method such that $d_{\mathcal{A}}$ is a metric over $Labs(\mathcal{A})$, but which agrees with $\mathtt{id}^{diff}$ on $Comp(\mathcal{A})$? Here we give one possibility. The idea is to take a sum over all arguments, but to weigh the contribution of an argument $a$ in the sum by the inverse of the size of the *Comp*-issue to which $a$ belongs. This gives rise to the *extended issue-based distance method* $\mathtt{eid}^{diff}$.

$$\mathtt{eid}^{diff}_{\mathcal{A}}(L_1, L_2) = \sum_{a \in Args_{\mathcal{A}}} \frac{diff(L_1(a), L_2(a))}{|[a]_{Comp}|}$$

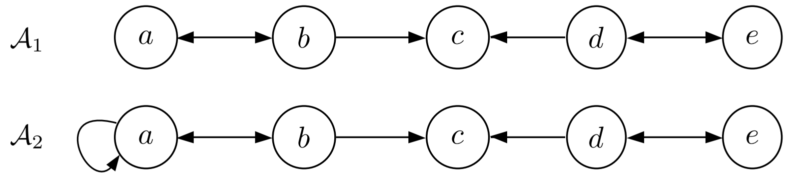



Figure 18: Two aggregation frameworks.

| | $l_1$ | $l_2$ | $l_3$ | $l_4$ | $l_5$ | $l_6$ | $n_1$ | $n_2$ | $n_3$ | $n_4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $l_1$: oioio | 0.0 | 2.0 | 1.0 | 1.0 | 4.0 | 3.0 | 2.0 | 5.0 | 5.0 | 4.5 |
| $l_2$: oiooi | 2.0 | 0.0 | 1.0 | 3.0 | 2.0 | 3.0 | 2.0 | 3.0 | 3.0 | 2.5 |
| $l_3$: oiouu | 1.0 | 1.0 | 0.0 | 2.0 | 3.0 | 2.0 | 1.0 | 4.0 | 4.0 | 3.5 |
| $l_4$: uuoio | 1.0 | 3.0 | 2.0 | 0.0 | 3.0 | 2.0 | 1.0 | 4.0 | 4.0 | 4.5 |
| $l_5$: uuuoi | 4.0 | 2.0 | 3.0 | 3.0 | 0.0 | 1.0 | 2.0 | 1.0 | 1.0 | 1.5 |
| $l_6$: uuuuu | 3.0 | 3.0 | 2.0 | 2.0 | 1.0 | 0.0 | 1.0 | 2.0 | 2.0 | 2.5 |
| $n_1$: uuouu | 2.0 | 2.0 | 1.0 | 1.0 | 2.0 | 1.0 | 0.0 | 3.0 | 3.0 | 3.5 |
| $n_2$: iouoi | 5.0 | 3.0 | 4.0 | 4.0 | 1.0 | 2.0 | 3.0 | 0.0 | 2.0 | 2.5 |
| $n_3$: uuioi | 5.0 | 3.0 | 4.0 | 4.0 | 1.0 | 2.0 | 3.0 | 2.0 | 0.0 | 0.5 |
| $n_4$: uiioi | 4.5 | 2.5 | 3.5 | 4.5 | 1.5 | 2.5 | 3.5 | 2.5 | 0.5 | 0.0 |

Table 5: Distance between labellings by $\mathtt{eid}^{diff}$ with refined Hamming distance $diff_{rh}$ over the labels of $\mathcal{A}_2$ from Figure 18.

**Proposition 84.** *(i)* $\mathtt{eid}^{diff}_{\mathcal{A}}$ *is a metric over* $Labs(\mathcal{A})$

*(ii)* $\mathtt{eid}^{diff}_{\mathcal{A}}(L_1, L_2) = \mathtt{id}^{diff}_{\mathcal{A}}(L_1, L_2)$ *for all* $L_1, L_2 \in Comp(\mathcal{A})$.

*Proof.* (i) The $\mathtt{eid}^{diff}_{\mathcal{A}}$ is constructed the same way as $d^{diff,\mathfrak{S}}_{\mathcal{A}}$ in Theorem 52 but this time instead of using the same label metric $\mathtt{diff}$ we use t-scaled metrics with $t = \frac{1}{|[a]_{Comp}|}$ depending on the size of the issue argument belongs to. The set of all arguments is *Labs*-critical therefore $\mathtt{eid}^{diff}_{\mathcal{A}}$ is a metric (not only pseudometric). (ii) For complete labelling all arguments from the issue have the same conflict. The difference is counted for each member of the issue and then divided by the number of the elements in the issue. □

We define Extended Issue-based distance method in two variant by plugging two variants of discrete metric ($\mathtt{EIBd}, \mathtt{rEIBd}$). Therefore we have two distance methods that can be used freely on $Labs(\mathcal{A})$: $\mathtt{Hd}/\mathtt{rHd}$ and $\mathtt{EIBd}/\mathtt{rEIBd}$.

**Example 85.** *Table 5 presents some distances returned by* $\mathtt{rEIBd}_{\mathcal{A}_2}$*, where* $\mathcal{A}_2$ *is from Figure 18.*

$Args_{\mathcal{A}_2}$ *partitions into three Comp-issues* $\{a, b\}, \{c\}$ *and* $\{d, e\}$*. Let us calculate a few entries as an example:* $\mathtt{rEIBd}(l_1, l_2) = 2$ *because there is no*

| | *abcde* | *abcde* |
|---|---|---|
| $\mathsf{cio}(\mathbf{L})$ | $n_1$ : uuouu | $n_1$ : uuouu |
| Prototype($\mathbf{L}$) | $l_5$ : uuuoi | $l_5$ : uuuoi |
| Endpoint($\mathbf{L}$) | $l_3$ : oiouu<br>$l_4$ : uuoio<br>$l_6$ : uuuuu | $l_6$ : uuuuu |
| Output($\mathbf{L}$) | $l_3$ : oiouu<br>$l_6$ : uuuuu | $l_6$ : uuuuu |
| Full($\mathbf{L}$) | $l_3$ : oiouu | $l_3$ : oiouu |

Table 6: Aggregation of the profile $\mathbf{L} = (l_4\colon \texttt{uuoio}, l_5\colon \texttt{uuuoi}, l_5\colon \texttt{uuuoi})$ - outcomes for different MO aggregation methods used with: extended issue-based labelling distance (rEIBd) (left column), and sum over all arguments (rHd) (right column). In both columns refined hamming distance over the labels and credulous initial operator is used.

*conflict over the first two issues and there is a hard conflict over the last one* $(0+0+2)$*;* $\mathsf{rEIBd}(l_3, l_4) = 2$ *because there are two soft conflicts over the first and the last issue and no conflict on the middle one* $(1+0+1)$*;* $\mathsf{rEIBd}(l_5, n_4) = 1.5$ *because there is half of a soft conflict over the first issue and a soft conflict over the second one* $(0.5+1+0)$*; etc.*

### 5.3.3 *Example of the MO Methods*

We illustrate the MO methods in our setting by continuing with the AF $\mathcal{A}_2$ from Figure 18. The $\mathcal{A}_2$-profile $\mathbf{L} = (l_4\colon \texttt{uuoio}, l_5\colon \texttt{uuuoi}, l_5\colon \texttt{uuuoi})$ aggregated with cio results in the non-complete labelling $n_1\colon \texttt{uuouu}$. The result of repairing it with MO methods is listed in Table 6.

In the left column the MO methods were instantiated with rEIBd. Method Prototype returns the closest complete labelling to the profile $\mathbf{L}$. We calculate the distance between $\mathbf{L}$ and labellings $l_1 - l_6$ by adding distances from the row $l_4$ (Table 5) to the doubled distances from the row $l_5$ and receive $9, 7, 8, 6, 3, 4$ respectively. The minimum is obtained for labelling $l_5\colon \texttt{uuuoi}$.

The Endpoint procedure returns the closest complete labelling to $\mathsf{cio}(\mathbf{L})$. We inspect the row $n_1$ in Table 5 to find that the minimum distance 1 is obtained for $l_3, l_4$ and $l_6$.

The Full and Output procedures search for closest cio-consistent profiles $\mathbf{L}'$ to profile $\mathbf{L}$. The Full procedure is restricted to the complete cio-consistent profiles. Consider $\mathbf{L}' = (l_1\colon \texttt{oioio}, l_5\colon \texttt{uuoio}, l_5\colon \texttt{uuoio})$. It is a cio-consistent profile with with $\mathsf{cio}(\mathbf{L}') = l_3\colon \texttt{oiouu}$. It is also minimal. The profiles $\mathbf{L}$ and $\mathbf{L}'$ differ just on $l_1$ and $l_4$ with distance 1 (soft conflict on issue $\{a, b\}$). There are no other complete cio-consistent profiles with different cio outcome and same distance,

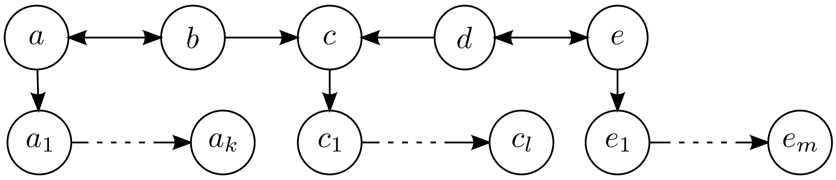



Figure 19: Aggregation frameworks extended with dependant chains.

because any change of labelling $l_4$ to another complete labelling costs more than 1. The only other candidate for change is labelling $l_5$. It can be changed to $l_6$: uuuuu for a cost of 1 but to affect the outcome of cio both occurrences of $l_5$ in **L** would have to be changed with a total cost of 2. The profile $\mathbf{L}'$ works for the `Output` procedure as well but `Output` is not restricted to complete cio-consistent profiles. Changing one of the $l_5$ labellings to a non-complete labelling $n_3$: uuioi with distance 1 creates additional conflict on argument c. As a result the labelling $l_6$: uuuuu is produced.

The results change when we switch to using rHd rather than rEIBd (right column). In this case the size of the issues does matter. The `Endpoint` procedure only selects $l_6$ because it differs with $n_1$ over issue $\{c\}$ with one argument, while the other two labellings $l_3, l_4$ differ over the two-argument issues $\{a, b\}$ and $\{d, e\}$ respectively. Similarly in the case of `Output` changing $l_5$ into $n_3$ over a single-argument issue is closer than the change of $l_4$ into $l_1$ over an issue with two arguments.

## 5.4 AGENDA MANIPULATION

In this section we explain informally why Indifference to peripheral issues **(IPI)** introduced in Section 3.3.4 of Chapter 3 is important in the context of judgement aggregation. Let us remind that **(IPI)** states that the distance between corresponding labellings of an AF and its extension with a single attacked argument should be the same.

**Example 86.** *Consider AF $\mathcal{A}$ from Figure 19 which is AF $\mathcal{A}_1$ from Figure 18 extended wiith the chains of attacking arguments attached to arguments* a, c *and* e. *Consider the chain of arguments* $a_1, ..., a_k$ *attached to the argument* a *with the attacks* $(a, a_1), (a_1, a_2), ..., (a_{k-1}, a_k)$. *The status of such chain under complete semantics is determined by the labelling of argument* a *therefore they belong to the* $[a]_{Comp}$. *The same holds for the chains attached to the arguments* c *and* e. *We have three Comp-issues* $[a]_{Comp} = \{a, b, a_1, \ldots, a_k\}$, $[c]_{Comp} = \{c, c_1, \ldots, c_l\}$ *and* $[e]_{Comp} = \{d, e, e_1, \ldots, e_m\}$. *Let* $\mathcal{A}[k, l, m]$ *for* $k, l, m \geqslant 0$ *denote AF extended with* $k, l, m$ *arguments attached to* a,c *and* e *respectively.*

*Assume, we have two agents with expertise in issue* $[a]_{Comp}$ *and* $[e]_{Comp}$ *respectively with opinions expressed by complete* $\mathcal{A}_1$*-labellings* iouuu *and*

$\texttt{uuuoi}$, *i.e. the first agent accepts argument* $a$, *rejects argument* $b$ *and remains undecided about arguments from other issues while the second agent accepts argument* $e$ *rejects argument* $d$ *and remains undecided about arguments in other issues. Since agents are rational i.e. express opinions using complete labellings we can assume that while presented with any* $\mathcal{A}[k,l,m]$ *agents will extend their labellings in a unique way according to the rules of complete semantics. We denote those extended labellings by* $\overline{\texttt{iouuu}}, \overline{\texttt{uuuoi}}$.

*Assume experts are questioned in a court and their opinions* $\mathbf{L} = (\overline{\texttt{iouuu}}, \overline{\texttt{uuuoi}})$ *are agregated using* $\texttt{Endpoint}$ *and* $\texttt{cio}$. *If distance satisfying* **(IPI)**, *like issue-based* $IBd$, *is used 'the chains are ignored'. We obtain*

$$\texttt{Endpoint}^{\texttt{cio},IBd}_{\mathcal{A}[k,l,m]}(\mathbf{L}) = \{\overline{\texttt{ioioi}}\}$$

*for any* $k, l, m$. *But if distance failing* **(IPI)** *is used, like full sum distance* $Hd$, *then dependent on the lengths of the chains* $k, l, m$ *we have*

$$\texttt{Endpoint}^{\texttt{cio},Hd}_{\mathcal{A}[0,0,0]}(\mathbf{L}) = \{\overline{\texttt{ioioi}}\},$$
$$\texttt{Endpoint}^{\texttt{cio},Hd}_{\mathcal{A}[0,3,0]}(\mathbf{L}) = \{\overline{\texttt{iouuu}}, \overline{\texttt{uuuoi}}\},$$
$$\texttt{Endpoint}^{\texttt{cio},Hd}_{\mathcal{A}[1,3,0]}(\mathbf{L}) = \{\overline{\texttt{uuuoi}}\},$$
$$\texttt{Endpoint}^{\texttt{cio},Hd}_{\mathcal{A}[0,3,1]}(\mathbf{L}) = \{\overline{\texttt{iouuu}}\}.$$

*This kind of behaviour allows the judge to change the outcome of the expertise by formulating problem in different seemingly equivalent ways. Intuitively we would like to avoid such behaviour.*

Similar examples can be constructed for other MO methods. This kind of behaviour is referred to as agenda manipulation. We leave formalisation of the above problem for a future work.

## 5.5 DOWN-ADMISSIBLE AND UP-COMPLETE AS A MO METHOD

Caminada and Pigozzi (2011) explore another way of repairing the results of an initial aggregator $M$ in the argumentation setting, using the *down-admissible* and *up-complete* labellings. To recall them we need some more notation.

**Definition 87** (Caminada and Pigozzi (2011))**.** *Let* $\mathcal{A}$ *be an AF and* $L_1, L_2 \in Labs(\mathcal{A})$. *We say that* $L_1$ *is* less or equally committed as $L_2$, written $L_1 \sqsubseteq L_2$, iff $\texttt{in}(L_1) \subseteq \texttt{in}(L_2)$ and $\texttt{out}(L_1) \subseteq \texttt{out}(L_2)$.

It can be observed Caminada and Pigozzi (2011) that $\sqsubseteq$ defines partial order over all $\mathcal{A}$-labellings.

**Definition 88** (Caminada and Pigozzi (2011))**.** *Given an* $\mathcal{A}$*-labelling* $L$,

- *the* down-admissible *labelling of* $L$, *denoted by* $\downarrow L$, *is the (unique) greatest element (under* $\sqsubseteq$*) of the set of all admissible* $\mathcal{A}$*-labellings* $N$ *such that* $N \sqsubseteq L$,

| | *abcde* | *abcde* |
|---|---|---|
| $\mathbf{L}$ | $l_5$: uuuoi | $l_4$: uuoio |
| | $l_9$: iouuu | $l_5$: uuuoi |
| $\mathtt{cio}(\mathbf{L})$ | $n_2$: iouoi | $n_1$: uuouu |
| $\lfloor \mathtt{cio}(\mathbf{L})$ | $n_2$: iouoi | $l_6$: uuuuu |
| $\rceil\lfloor \mathtt{cio}(\mathbf{L})$ | $l_8$: ioioi | $l_6$: uuuuu |

Table 7: Two profiles $(l_5, l_9)$ and $(l_4, l_5)$ for which cio violates collective rationality and is repaired with down-admissible up-complete procedures.

- *the* up-complete *labelling of* L*, denoted by* $\rceil$L*, is the (unique) smallest element (under* $\sqsubseteq$*) of the set of all complete* $\mathcal{A}$*-labellings* N *such that* $L \sqsubseteq N$.

Two computation procedures were given to calculate $\lfloor L$ and $\rceil L$ Caminada and Pigozzi (2011). Namely to obtain $\lfloor L$ one need to relabel any illegally `in` and illegally `out` argument into `undec` as long as there are illegal `in` (`out`) arguments. The order of relabelling does not matter. To obtain $\rceil L$ one need to relabel any illegally `undec` argument into legally `in` or legally `out`. Again the order in which we relabel does not influence the outcome. We denote by $\rceil\lfloor L$ the composite operation of performing the down-admissible followed by the up-complete procedures on an $\mathcal{A}$-labelling L.

**Example 89.** *In Table 7 two profiles (considered over* $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *respectively from Figure 18) are aggregated with the* cio *method and repaired with down-admissible up-complete procedures (assuming* $n = 2$*).*

*For the left column we consider the* $\mathcal{A}_1$*-profile* $(l_5, l_9)$*. The labelling* $\mathtt{cio}_{\mathcal{A}_1}(l_5, l_9) = n_2$ *is admissible therefore the down-admissible procedure is vacuous. But although* c *was legally-*`undec` *in each of the labellings in the profile it is illegally-*`undec` *because all its attackers are labelled* `out`*. The up-complete procedure relabels it to* `in` *and a complete labelling* $l_8$ *is produced.*

*For the right column we consider the* $\mathcal{A}_2$*-profile* $(l_4, l_5)$*. The labelling* $\mathtt{cio}_{\mathcal{A}_2}(l_4, l_5) = n_1$ *is not complete because argument* c *is labelled* `out` *despite the fact that none of its attackers is* `in`*. Thus it needs to be repaired. The down-admissible procedure relabels* c *to* `undec` *to obtain* $\lfloor n_1 = l_6$: uuuuu*. Since there are no illegally-*`undec` *arguments the up-complete procedure is not needed and so* $\rceil\lfloor n_1 = l_6$.

The above procedure can be viewed more generally as an aggregation method that takes any initial aggregation operator (not just cio) as a parameter.

**Definition 90.** *Let* M *be any initial aggregation method. The aggregation method* $\mathtt{DAUC}^M$ *is defined by setting, for any AF* $\mathcal{A}$ *and* $\mathcal{A}$*-profile* $\mathbf{L}$*,* $\mathtt{DAUC}^M_{\mathcal{A}}(\mathbf{L}) = \{\rceil\lfloor L \mid L \in M_{\mathcal{A}}(\mathbf{L})\}$.

By construction $\mathtt{DAUC}^{M}$ is collectively rational, whatever we take $M$ to be. The above definition is applicable also for cases in which $M$ might not be resolute. However in case $M$ is resolute then clearly so too is $\mathtt{DAUC}^{M}$.

How does this aggregation compare to the MO methods? It turns out that DAUC can be viewed as an instance of the Endpoint method of MO by taking the following distance.

$$d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_2) = \begin{cases} 0 & \text{if } L_1 = L_2 \\ 1 & \text{if } L_1 \neq L_2 \text{ and } \downharpoonright L_1 = \downharpoonright L_2 \\ 2 & \text{if } \downharpoonright L_1 \neq \downharpoonright L_2 \text{ and} \\ & \upharpoonright\downharpoonright L_1 = \upharpoonright\downharpoonright L_2 \\ 3 & \text{otherwise.} \end{cases}$$

**Proposition 91.** (i). $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}$ *is a metric (over Labs($\mathcal{A}$)).*
(ii). *For resolute initial aggregation operator* $M$, $\mathtt{DAUC}_{\mathcal{A}}^{M} = \mathtt{Endpoint}_{\mathcal{A}}^{M, d^{\upharpoonright\downharpoonright}}$

*Proof.* (i) Metrics. The conditions **(REF,DD,SYM)** (Definition 12) are self evident. Only condition **(TRI)**, the triangle inequity - $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_2) \leqslant d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_3) + d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_3, L_2)$ - can rise doubts. We prove by considering all possible distances between $L_1$ and $L_2$. If $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_2) = 3$ then $\upharpoonright\downharpoonright L_1 \neq \upharpoonright\downharpoonright L_2$. In this case either $\upharpoonright\downharpoonright L_3 \neq \upharpoonright\downharpoonright L_1$ or $\upharpoonright\downharpoonright L_3 \neq \upharpoonright\downharpoonright L_2$ from which follows that either $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_3) = 3$ or $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_3, L_2) = 3$. Therefore at least one of the non-negative addends of the right side is equal to the left side and the inequity holds. Similarly, if $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_2) = 2$ then $\downharpoonright L_3 \neq \downharpoonright L_1$ or $\downharpoonright L_3 \neq \downharpoonright L_2$ and so $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_1, L_3) \geqslant 2$ or $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L_3, L_2) \geqslant 2$. And so on, at least one of the addends of the right side is bigger than the left side.

(ii) Take any $\mathcal{A}$-profile $\mathbf{L}$. We have

$$\mathtt{DAUC}_{\mathcal{A}}^{M}(\mathbf{L}) = \{\upharpoonright\downharpoonright M(\mathbf{L})\}$$

and

$$\mathtt{Endpoint}_{\mathcal{A}}^{M, d_{\mathcal{A}}^{\upharpoonright\downharpoonright}} = \underset{L \in Comp(\mathcal{A})}{\arg\min}\ d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L, M(\mathbf{L})).$$

Consider complete labellings over which Endpoint minimalise distance. By definition $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(\upharpoonright\downharpoonright M(\mathbf{L}), M(\mathbf{L})) \leqslant 2$ while for any other complete labelling $L' \neq \upharpoonright\downharpoonright M(\mathbf{L})$, $d_{\mathcal{A}}^{\upharpoonright\downharpoonright}(L', M(\mathbf{L})) = 3$ therefore $\mathtt{Endpoint}_{\mathcal{A}}^{M, d_{\mathcal{A}}^{\upharpoonright\downharpoonright}} = \{\upharpoonright\downharpoonright M(\mathbf{L})\}$. □

## 5.6 CONCLUSIONS

We have continued work on distance methods for argumentation from Chapter 4, illustrating how they can be employed to address problems of aggregation in argumentation. To do this we adapted the framework of Miller and Osherson from binary judgement aggregation to our setting, defining several operators for aggregating argument labellings.

To use the Miller and Osherson framework fully requires distance which is a metric defined for all labellings. From the three distance measures proposed in Chapter 4 only full sum distance Hd satisfies this requirement while issue-based distance IBd and maximum over minimal critical set distance MMCSd are pseudometrics. To address this problem, we extended issue-based distance measure to the set of all labellings. We illustrated informally that methods which fail **(IPI)** may lead to agenda manipulation, which makes IBd superior to the Hd when applied to MO framework. The generalisation of MMCSd, another distance satisfying **(IPI)** is left for future work.

Finally we illustrated the generality of the resulting framework for aggregation by showing how the ⇃⇂ aggregation method of Caminada and Pigozzi (2011) can be viewed as an instance of one of the MO methods.

There are several avenues open for future work. Firstly, a feature of our examples (see Section 5.3.3) is that the different aggregation methods can all yield quite different results. This raises the question of which method to prefer. We plan to classify the different methods in terms of the postulates they satisfy. Previous work on labelling aggregation Booth et al. (2014); Rahwan and Tohmé (2010) have examined postulates for such operators (inspired by postulates from JA), but have done so only for *resolute* aggregation methods. We will generalise these to irresolute methods, perhaps taking a lead from similar generalisations from JA, Grandi and Pigozzi (2012).

Another interesting question is to consider what happens if you aggregate *all* complete $\mathcal{A}$-labellings of an AF. This question was considered in Caminada and Pigozzi (2011), but again only for resolute operators. In this way the operators of Caminada and Pigozzi (2011) were able to characterise certain *single-status* argumentation semantics (i.e., grounded and ideal). Our move to irresolute aggregation opens the possibility that we might be able to capture also some *multiple-status* semantics such as Dung's (1995) *preferred* semantics. That is, does aggregating all complete $\mathcal{A}$-labellings yield precisely the set of preferred $\mathcal{A}$-labellings?

# 6 A PERSUASION DIALOGUE FOR GROUNDED SEMANTICS

## 6.1 INTRODUCTION

The field of formal argumentation consists two main lines of research. One line of research is concerned with the dialectical process of two or more players who are involved in a discussion. This kind of argumentation, referred to as *dialogue theory* in the ASPIC project ASPIC-consortium (2005), can be traced back to the work of Hamblin (1970; 1971) and Mackenzie (1979; 1990). A different line of research is concerned with arguments as a basis for nonmonotonic inference. The idea is that (nonmonotonic) reasoning can be performed by constructing and evaluating arguments, which are composed of a number of reasons that collectively support a particular claim. This line of research can be traced back to the work of Pollock (1992; 1995), Vreeswijk (1993; 1997) and Simari and Loui (1992), and has culminated with the work of Dung (1995), which serves as the basis of much of today's argumentation research.

One particular question one may ask is to what extent it is possible to create links between these two lines of research. One particular way of doing so would be to have an argument accepted (under a particular Dung-style semantics) iff it can be defended in a particular type of formal dialogue. In previous work, Caminada (2010) observed that (credulous) preferred semantics can be reinterpreted as a particular type of Socratic dialogue. That is, an argument is in at least one preferred extension iff the proponent is able to successfully defend the argument in the associated Socratic discussion game, against a maximally sceptical opponent. In the current chapter we take a similar approach, this time not for preferred but for grounded semantics. Our claim is that the acceptance of an argument under grounded semantics coincides with the ability to win a particular type of dialogue, against a maximally sceptical opponent.

One of the aims of our work is to contribute to a conceptual basis for (abstract) argumentation theory. Whereas, for instance, classical logic is based on the notion of truth, it is not immediately obvious where a notion like truth would fit in when it comes to (abstact) argumentation research. Still, one would like to determine what the various argumentation semantics actually constitute to. An answer like "preferred semantics is about the maximal conflict-free fixpoints, whereas grounded semantics is about the minimal conflict-free fixpoint" might be technically correct, but is still conceptually somewhat unsatisfying. We believe that formal dialogue can serve as a conceptual basis for (abstract) argumentation theory. The idea is that one infers not so much what is *true*, as is the case in classical logic, but what can be *defended in rational discussion*. In particular, our aim is to show that different argumentation semantics correspond with different *types* of rational discussion.

The remaining part of this chapter is structured as follows. First, in Section 6.2 we present our formal persuasion dialogue and prove its equivalence with acceptance under grounded semantics. Then we round off in Section 6.3 with a discussion of the obtained results and a treatment of related research.

## 6.2 A DIALOGUE GAME FOR GROUNDED SEMANTICS

Our proposed dialogue game consists of the following moves.

CLAIM This is the first move in the dialogue, at which the proponent claims that a particular argument has to be labelled `in`, creating a commitment at the side of the speaker.

WHY With this move, the opponent asks why a particular argument has to be labelled a particular way.

BECAUSE With this move, similar to the *since* move in Mackenzie's DC, a party explains why the status of a particular argument has to be the way the party stated earlier. The explanation can create new commitments at the side of the speaker.

CONCEDE With this move, a party concedes part of the statements uttered earlier by the other party, creating new commitments at the side of the speaker. Although the act of conceding is left implicit in Mackenzie's DC, we agree with Walton and Krabbe (1995) that it can have advantages to explicitly represent the act of conceding.

Before laying out the precise formal rules of the dialogue game, it can be illustrative to examine some examples. Consider for instance the argumentation framework of Figure 20. Here, the discussion could go as follows.

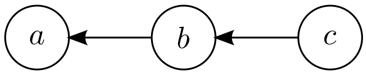



Figure 20: An Argumentation framework with simple reinstatement

| Moves | Commitment | | | |
|---|---|---|---|---|
| | Proponent | | Opponent | |
| | in | out | in | out |
| P: claim in(a) | a | - | - | - |
| O: why in(a) | a | - | - | - |
| P: because out(b) | a | b | - | - |
| O: why out(b) | a | b | - | - |
| P: because in(c) | a,c | b | - | - |
| O: concede in(c), out(b), in(a) | a,c | b | c,a | b |

In essence, our inspiration comes from the argument game for grounded semantics as described in Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009). Here, a game basically consists of a proponent (P) and an opponent (O) taking turns in putting forward arguments (proponent begins). Each moved argument has to be an attacker of the previously moved argument by the other player. In order to ensure that the game terminates, the proponent is disallowed from moving the same argument twice (although the opponent does not have this restriction). A player wins if the other player cannot move any more.

Although the standard grounded game, as described in Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009), can serve fine as a basis for argument-based proof procedures, it does have some properties that deviate from what one would expect for a persuasion dialogue. In particular, where the definition of a complete labelling requires that for an argument to be in *all* of its attackers have to be out, in the standard grounded game, it is only claimed that just *one* of the attackers is out.[1] It is difficult to maintain that an agent can be persuaded that all attackers of an argument are labelled out when this is shown only for one of them. Therefore, the standard grounded game cannot be said to be truly about persuasion, at least not within an individual game.[2,3]

[1] That is, if one interprets the standard grounded game in terms of argument labellings, where the proponent makes in moves and the opponent makes out moves, as is done in Modgil and Caminada (2009).

[2] For establishing correctness and completeness with respect to membership of the grounded extension, the standard grounded game relies not on an individual game, but on the presence of a winning strategy. We refer to Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009) for details.

[3] Similar remarks can be made about other existing dialectical proof procedures for abstract argumentation semantics, such as Dung et al. (2007); Thang et al. (2009).

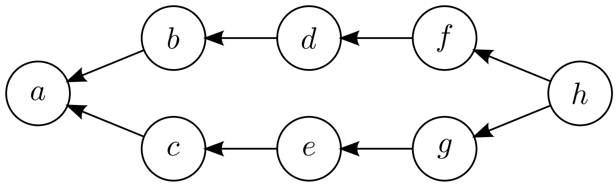



Figure 21: Two paths coming from the same argument

For the grounded game to be categorized as persuasion, it would be highly desirable to be able to evaluate, in the same game, *all* attackers of a particular argument that is claimed to be `in`. This, however, creates a new type of problems. Consider the example in Figure 21.

A discussion in which all attackers of an `in`-labelled argument can be evaluated might look as follows.

| | Commitment | | | |
|---|---|---|---|---|
| Moves | Proponent | | Opponent | |
| | `in` | `out` | `in` | `out` |
| P: `claim in`$(a)$ | a | - | - | - |
| O: `why in`$(a)$ | a | - | - | - |
| P: `because out`$(b)$, `out`$(c)$ | a | b,c | - | - |
| O: `why out`$(b)$ | a | b,c | - | - |
| P: `because in`$(d)$ | a,d | b,c | - | - |
| O: `why in`$(d)$ | a,d | b,c | - | - |
| P: `because out`$(f)$ | a,d | b,c,f | - | - |
| O: `why out`$(f)$ | a,d | b,c,f | - | - |
| P: `because in`$(h)$ | a,d,h | b,c,f | - | - |
| O: concede `in`$(h)$, `out`$(f)$, `in`$(d)$, `out`$(b)$ | a,d,h | b,c,f | h,d | f,b |
| O: `why out`$(c)$ | a,d,h | b,c,f | h,d | f,b |
| P: `because in`$(e)$ | a,d,h,e | b,c,f | h,d | f,b |
| O: `why in`$(e)$ | a,d,h,e | b,c,f | h,d | f,b |
| P: `because out`$(g)$ | a,d,h,e | b,c,f,g | h,d | f,b |
| O: `why out`$(g)$ | a,d,h,e | b,c,f,g | h,d | f,b |
| P: `because in`$(h)$ | a,d,h,e | b,c,f,g | h,d | f,b |
| O: concede `out`$(g)$, `in`$(e)$, `out`$(c)$, `in`$(a)$ | a,d,h,e | b,c,f,g | h,d,e,a | f,b,g,c |

In the above game, the proponent moves the same argument $(h)$ twice. In the standard grounded game, this would not be allowed, since it can lead to non-termination. Take for instance an argumentation framework consisting of two arguments $a$ and $b$ that attack each other as in Figure 22. When one allows for the proponent to repeat arguments, the resulting game can be infinite.

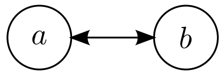



Figure 22: Two arguments attacking each other

| | Commitment | | | |
|---|---|---|---|---|
| Moves | Proponent | | Opponent | |
| | in | out | in | out |
| P: claim in($a$) | a | - | - | - |
| O: why in($a$) | a | - | - | - |
| P: because out($b$) | a | b | - | - |
| O: why out($b$) | a | b | - | - |
| P: because in($a$) | a | b | - | - |
| O: why in($a$) | a | b | - | - |
| P: because out($b$) | a | b | - | - |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

In the standard grounded game, the reason one does not have to repeat argument $h$ in the example of Figure 21 is because the lines of arguments in($a$) — out($b$) — in($d$) — out($f$) — in($h$) and in($a$) — out($c$) — in($e$) — out($g$) — in($h$) are considered to be distinct, constituting different discussions.[4] However, they can only be distinct because one does *not* require all attackers of an in-labelled argument to be evaluated to be labelled out in the same discussion.

Overall, what we are interested in is to define a dialogue game such that (1) all attackers of an argument that is claimed to be in can be evaluated to be out in the same dialogue, (2) each dialogue is guaranteed to terminate, and (3) the ability for the proponent to win the dialogue coincides with membership of the grounded extension.

As labellings can be seen as information about all arguments in the graph, partial labellings are handy to express partial information about a subset of arguments.

**Definition 92.** *Let* $(Args, \rightharpoonup)$ *be an argumentation framework. A* partial argument labelling *is a partial function* $\mathcal{L} : Args \to \{\texttt{in}, \texttt{out}, \texttt{undec}\}$*.*

When we talk about partial labellings we will use caligraphic $\mathcal{L}$ in contrast to usual L. By dom($\mathcal{L}$) we denote a domain of $\mathcal{L}$.

For any arguments $a_1, ..., a_n \in Args$ by in($a_1, ..., a_n$) we mean partial labelling that labels precisely arguments $a_1, ..., a_n$ in. Similarly we use out(...) and undec(...)

To keep notation short and skip writing separate cases for arguments labelled in and out we use the following two definitions.

[4]They are essentially branches in the tree of the winning strategy Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009).

**Definition 93.** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be partial labellings. We say $\mathcal{L}_1$* is a reason for $\mathcal{L}_2$ *iff:*

1. $\mathcal{L}_2 = \mathtt{in}(a)$ *or* $\mathcal{L}_2 = \mathtt{out}(a)$.
2. *if* $\mathcal{L}_2 = \mathtt{in}(a)$ *then* $\mathcal{L}_1 = \mathtt{out}(b_1, ..., b_n)$ *where* $b_i, i \geqslant 0$ *are all attackers of* $a$ *in the framework.*
3. *if* $\mathcal{L}_2 = \mathtt{out}(a)$ *then* $\mathcal{L}_1 = \mathtt{in}(b)$ *for some* $b$ *attacker of* $a$.

We extend this notion to partial labellings of any size.

**Definition 94.** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be partial labellings. We say $\mathcal{L}_1$* contains a reason for $\mathcal{L}_2$ *iff:*

1. $\mathcal{L}_2(a) \in \{\mathtt{in}, \mathtt{out}\}$ *for all arguments* $a$ *in the domain of* $\mathcal{L}_2$.
2. *for any singleton argument labelling, that is a labelling that labels only a single argument,* $\mathcal{L} \subseteq \mathcal{L}_2$*, there exists a labelling* $\mathcal{L}'$ *such that* $\mathcal{L}'$ *is a reason for* $\mathcal{L}$ *and* $\mathcal{L}' \subseteq \mathcal{L}_1$.

Note that if $\mathcal{L}_1$ is a reason for $\mathcal{L}_2$ then $\mathcal{L}_1$ contains a reason for $\mathcal{L}_2$ but usually not the reverse. Directly from Definition 9 follows the observation below.

**Observation 95.** *Let* $L$ *be a complete labelling. If* $\mathcal{L}_1$ *contains a reason for* $\mathcal{L}_2$ *and* $\mathcal{L}_1 \subseteq L$ *then* $\mathcal{L}_2 \subseteq L$.

So if $L$ represents the position of a rational agent that satisfies the condition of a complete labelling and contains reasons for singleton argument labelling $\mathcal{L}_a$, then it must also contain labelling $\mathcal{L}_a$ itself.

The dialogue game is guided by the definition of complete labelling. The opponent is assumed to be maximally sceptical, conceding only if this cannot be avoided. That is, he only concedes that argument is `in` if he is already committed that all attackers are `out` and he only concedes that an argument is `out` if he is already committed that at least one attacker is `in`. Informally, the rules of the dialogue can be described as follows.

- The proponent (P) and opponent (O) takes turns; Each turn of P contains single move `claim` or `because`; In each turn O plays one or more moves. O's turn starts with an optional sequence of `concede` moves and finishes (when possible) with single `why` move.
- P gets committed to arguments used in `claim` and `because` moves; O gets committed to arguments used in `concede` moves.
- P starts with `claim` $\mathtt{in}(a)$ where $a$ is the main argument of the discussion; `claim` cannot be repeated later in the game.

- In consecutive turns P provides reasons for the directly preceding `why` L move of O by moving `because` L′ where L′ is a reason of L.
- P can play `because` only if the reason given does not contain any arguments already mentioned (in P's commitment store) but not yet accepted (not in O's commitment store). We call such arguments *open issues*.
- O addresses the most recent open issue L ($\mathtt{in}(a)$ or $\mathtt{out}(a)$) in the discussion. If O is committed to reasons for L it must `concede` L otherwise O questions all reasons that O is not committed to with `why`.
- O can question with `why` just one argument.
- The moves `claim`, `because` and `concede` can be played only if new commitments do not contradict a previous one.
- The discussion terminates when no more moves are possible. If O conceded the main argument then P wins, otherwise O wins.

We now formally define the above sketched dialogue game.

**Definition 96** (Discussion move). *A discussion move is a triple* $M = (\mathcal{P}, \mathcal{T}, \mathcal{L})$ *where* $\mathcal{P} \in \{\mathtt{proponent}, \mathtt{opponent}\}$ *is a player,* $\mathcal{T} \in \{\mathtt{claim}, \mathtt{why}, \mathtt{because}, \mathtt{concede}\}$ *is a move type and* $\mathcal{L}$ *is a partial labelling.*

**Definition 97** (Discussion). *A discussion is a tuple* $(\mathcal{M}, \mathcal{CS})$ *where* $\mathcal{M}$ *is a finite sequence of moves* $[M_1, ..., M_n]$ *and* $\mathcal{CS}$ *is a function assigning to each player a partial labelling representing his commitment.*

*Additionally we define* open issues *as* $OI(\mathcal{D}) = \mathcal{CS}(\mathtt{proponent}) \setminus \mathcal{CS}(\mathtt{opponent})$ *and the* last open issue $LOI(\mathcal{D}) = OI(\mathcal{D}) \cap \mathcal{L}_k$ *where* $M_k = (\mathcal{P}_k, \mathcal{T}_k, \mathcal{L}_k)$ *is the last move such that* $OI(\mathcal{D}) \cap \mathcal{L}_k$ *is not empty (maximal k).*

**Definition 98** (Grounded discussion). *A discussion* $\mathcal{D} = (\mathcal{M}, \mathcal{CS})$, $\mathcal{M} = [M_1, ..., M_{n+1}]$ *is a* grounded discussion *iff the following recursive conditions (basis) or (construction) hold.*

(BASIS) $n = 0$ *and the following holds*

$I_1$ $M_1 = (\mathtt{proponent}, \mathtt{claim}, \mathtt{in}(A))$

$I_2$ $\mathcal{CS}(\mathtt{proponent}) = \mathtt{in}(A)$ *and* $\mathcal{CS}(\mathtt{opponent}) = \emptyset$

(CONSTRUCTION) $n > 0$ *and* $\mathcal{D}' = (\mathcal{M}', \mathcal{CS}')$ *(where* $\mathcal{M}'$ *is* $[M_1, ..., M_n]$*) is a grounded discussion and one of the following holds:*

$W_1$ $M_{n+1} = (\mathtt{opponent}, \mathtt{why}, \mathcal{L})$ *and*

$W_2$ $M_n = (\mathcal{P}_n, \mathcal{T}_n, \mathcal{L}_n)$, $\mathcal{T}_n \neq \mathtt{why}$ *and*

$W_3$ $\mathcal{L} \subseteq LOI(\mathcal{D}')$, $\#\mathcal{L} = 1$ *and*

$W_4$ *there is no* $\mathcal{L}' \subseteq \mathrm{LOI}(\mathcal{D}'), \#\mathcal{L}' = 1$ *such that* $\mathcal{CS}(\mathtt{opponent})$ *contains the reason for* $\mathcal{L}'$ *and*

$W_5$ $\mathcal{CS} = \mathcal{CS}'$

*or*

$B_1$ $M_{n+1} = (\mathtt{proponent}, \mathtt{because}, \mathcal{L})$ *and*

$B_2$ $M_n = (\mathtt{opponent}, \mathtt{why}, \mathcal{L}')$ *and*

$B_3$ $\mathcal{L}$ *is a reason for* $\mathcal{L}'$ *and*

$B_4$ $\mathcal{L} \cap \mathrm{OI}(\mathcal{D}') = \emptyset$ *and*

$B_5$ $\mathcal{CS}(\mathtt{proponent}) = \mathcal{CS}'(\mathtt{proponent}) \cup \mathcal{L}$ *and*

$B_6$ $\mathcal{CS}(\mathtt{opponent}) = \mathcal{CS}'(\mathtt{opponent})$

*or*

$C_1$ $M_{n+1} = (\mathtt{opponent}, \mathtt{concede}, \mathcal{L})$ *and*

$C_2$ $\mathcal{L} \subseteq \mathrm{LOI}(\mathcal{D}'), \#\mathcal{L} = 1$ *and*

$C_3$ $\mathcal{CS}'(\mathtt{opponent})$ *contains a reason for* $\mathcal{L}$ *and*

$C_4$ $\mathcal{CS}(\mathtt{proponent}) = \mathcal{CS}'(\mathtt{proponent})$ *and*

$C_5$ $\mathcal{CS}(\mathtt{opponent}) = \mathcal{CS}'(\mathtt{opponent}) \cup \mathcal{L}$

*We say that a discussion is* terminated *if it cannot be extended any more. For a terminated discussion* `proponent` *wins if* `opponent` *conceded the main claim of the discussion, otherwise* `opponent` *wins.*

**Observation 99.** *(Grounded discussion properties)*

1. *The* `claim` *move is the first move in every discussion* ($I_1$) *and it is never repeated as it is not listed in* construction *part of Definition 98.*

2. *The* `concede` *is never played after* `why` *as* $W_4$ *excludes* $C_3$ *so* $C_3$ *does not hold when* `why` *is played and also in the next move as* `why` *does not change the commitment store.*

3. *After each* `concede` *move* $\mathcal{CS}(\mathtt{opponent})$ *is enlarged with one argument. Consider a partial labelling* $\mathcal{L}$ *that is added to* $\mathcal{CS}(\mathtt{opponent})$ *during a* `concede` *move* ($C_5$)*. By condition* $C_2$ *it labels exactly one argument and because* $\mathcal{L} \subseteq \mathrm{LOI}(\mathcal{D}') \subseteq \mathrm{OI}(\mathcal{D}')$ *and* $\mathrm{OI}(\mathcal{D}') \cap \mathcal{CS}(\mathtt{opponent}) = \emptyset$ *it was not in* $\mathcal{CS}(\mathtt{opponent})$ *before.*

4. *The* `why` *move cannot be repeated directly after other* `why` *move* ($W_2$)*.*

5. *Whenever a* `concede` *move can be played, by condition* $W_4$ `why` *move cannot.*

6. *The* `because` *move follows directly after a* `why` *move* ($B_2$)*.*

7. *After each* because *move* $\mathcal{CS}$(proponent) *is enlarged. Consider labelling* $\mathcal{L}$ *that is added to* $\mathcal{CS}$(proponent) *during a* because *move* ($B_5$)*.* $\mathcal{L}$ *is a reason for* $\mathcal{L}'$ *from the previous* why *move* ($B_3$)*. By* $W_4$ $\mathcal{L}$ *is not contained in* $\mathcal{CS}$(opponent) *and in particular* $\mathcal{L} \setminus \mathcal{CS}$(opponent) *is not empty. By* $B_4$ $\mathcal{L} \setminus \mathcal{CS}$(opponent) *is also not an open issue. Therefore* $\mathcal{L}$ *needs to contain at least one new element.*

8. *It is always the case that* $\mathcal{CS}(\texttt{opponent}) \subseteq \mathcal{CS}(\texttt{proponent})$*. It holds after first* claim *move* ($I_2$)*, then* opponent*'s commitment store is only modified during* concede *move when it is extended by* $\mathcal{L} \subseteq \mathrm{LOI}(\mathcal{D}) \subseteq \mathrm{OI}(\mathcal{D}) \subseteq \mathcal{CS}(\texttt{proponent})$.

9. *The* because *move cannot be played if the new commitment store defined in* $B_5$ *is not a partial labelling. Grounded discussion is a discussion and so the commitment stores need to be partial labellings. This rules out the possibility that the same argument can be labelled both* in *and* out*. This is not a concern in case of* concede *as opponent's commitment store is always subset of proponent's store (see previous point).*

We simplify the winning criteria as follows. The opponent wins (conversely the proponent loses) if and only if the proponent cannot respond to the why move of the opponent. Thus we can determine the winner of the game just by examining the type of the last move.

**Lemma 100.** *Let* $\mathcal{D} = (\mathcal{M}, \mathcal{CS})$ *with* $\mathcal{M}_i = [M_1, ..., M_k]$ *be a terminated grounded discussion based on argumentation framework* $\mathcal{A} = (Args, \rightharpoonup)$*. The opponent wins iff the last move* $M_k$ *is of type* why*.*

*Proof.* If the opponent wins then the main claim is not in its commitment store but it is in the proponent's commitment store, therefore the set of open issues is not empty. Then the set of last open issues is not empty as well. For all artial labellings $\mathcal{L}$ labelling one argument from open issues the opponent is not committed to any reason for $\mathcal{L}$ otherwise $\mathcal{D}$ which is terminated could be extended with a concede$\mathcal{L}$ move. But then condition $W_4$ (Definition 98) is fulfilled. The only reason for which $\mathcal{D}$ cannot be extended with the why$\mathcal{L}$ move is condition $W_2$ i.e. another why move was already played. By contraposition, players commitments stores in a terminated discussion which doesn't finish with why need to be equal. Since proponent is committed to the main claim so is the opponent. In this case the proponent wins. □

**Theorem 101.** *Any grounded discussion over a finite argumentation framework has to terminate.*

*Proof.* Assume there is an infinite sequence of grounded discussions $\mathcal{D}_i = (\mathcal{M}_i, \mathcal{CS}_i)$ and $\mathcal{M}_i = [M_1, ..., M_i]$ for $i = 1, ..., \infty$. As noticed in Observation 99 the claim move is played only once. Furthermore after both the because move and the concede move, the commitment stores of `proponent` and `opponent` respectively are extended with at least one new argument, so this moves can be played at most N times where N is a number of arguments in argumentation framework. Therefore after $2N+1$ moves only `why` moves can be played. But `why` moves cannot be played uninterrupted (without being interleaved with other types of moves). Therefore, every grounded game must terminate (cannot be extended infinitely many times). □

**Lemma 102.** *For any grounded discussion* $\mathcal{D} = (\mathcal{M}, \mathcal{CS})$ *based on argumentation framework* $\mathcal{A} = (Args, \rightharpoonup)$ *the commitment store of* `opponent` *is in the grounded labelling of* $\mathcal{A}$*. This is* $\mathcal{CS}(\texttt{opponent}) \subseteq L_{gr}$ *where* $L_{gr}$ *is the grounded labelling.*

*Proof.* We use induction over the number of moves in the discussion.

(Basis) For a discussion containing just one move (main `claim`) opponent's store is empty and condition is trivially fulfilled ($\mathcal{CS}(\texttt{opponent}) = \emptyset \subseteq L_{gr}$).

(Step) Assume that for grounded discussion $\mathcal{D}' = (\mathcal{M}', \mathcal{CS}')$, $\mathcal{M}' = [M_1, ..., M_n]$ our claim hold i.e. $\mathcal{CS}'(\texttt{opponent}) \subseteq L_{gr}$.[5] We need to show that it holds for a grounded discussion $\mathcal{D} = (\mathcal{M}, \mathcal{CS})$, $\mathcal{M} = [M_1, ..., M_n, M_{n+1}]$, $M_{n+1} = (\mathcal{P}_{n+1}, \mathcal{T}_{n+1}, \mathcal{L}_{n+1})$ as well.

We need to consider only the case where $\mathcal{T}_{n+1} = \texttt{concede}$. In other cases $\mathcal{CS}(\texttt{opponent}) = \mathcal{CS}'(\texttt{opponent}) \subseteq L_{gr}$ by assumption.

After the `concede` move $\mathcal{CS}(\texttt{opponent}) = \mathcal{CS}'(\texttt{opponent}) \cup \mathcal{L}$ ($C_5$). We only need to show that $\mathcal{L} \subseteq L_{gr}$ . According to $C_3$ $\mathcal{CS}'(\texttt{opponent})$ contain the reasons for $\mathcal{L}$. Then by assumption $L_{gr}$ contains the reasons and by Observation 95 $L_{gr}$ needs to include $\mathcal{L}$ as well. □

**Corollary 103.** *If proponent wins a grounded discussion then the main claim is in the grounded extension that is, it is labelled* `in` *by the grounded labelling.*

To show that our dialogue is not only sound but also complete, we construct the grounded persuasion game won by the proponent for all `in`-labelled arguments of the grounded labelling. For that, we recall the concept of strong admissibility first introduced by Baroni and Giacomin (2007) as an extension-based semantics and recently introduced by Caminada (2014) in terms of labellings. We use later version defined by min-max numbering, i.e. function assigning to each `in`/`out` labelled argument a natural number or infinity.

**Definition 104** (Caminada (2014))**.** *Let* L *be an admissible labelling of argumentation framework* $(Args, \rightharpoonup)$*. A* min-max numbering *is a total function* $\mathcal{MM}_L : \texttt{in}(L) \cup \texttt{out}(L) \to \mathbb{N} \cup \{\infty\}$ *such that for each* $a \in \texttt{in}(L) \cup \texttt{out}(L)$ *it holds that:*

- *if* $L(a) = \texttt{in}$ *then* $\mathcal{MM}_L(a) = \max(\{\mathcal{MM}_L(b) \mid b$ *attacks* $a$ *and* $L(b) = \texttt{out}\}) + 1$ *(with* $\max(\emptyset)$ *defined as* 0*)*

[5] Notice that $\mathcal{CS}'$ is well defined by Definition 98

- *if* $\mathtt{L}(a) = \mathtt{out}$ *then* $\mathcal{MM}_L(a) = \min(\{\mathcal{MM}_L(b) \mid b$ *attacks* $a$ *and* $\mathtt{L}(b) = \mathtt{in}\}) + 1$ *(with* $\min(\emptyset)$ *defined as* $\infty$*)*

Caminada proved that every admissible labelling has a unique min-max numbering. The strongly admissible labellings are those admissible labelling whose min-max numbering is finite, (2014).

**Definition 105** (Caminada (2014))**.** *A* strongly admissible labelling *is an admissible labelling whose min-max numbering yields natural numbers only (so no argument is numbered* $\infty$*).*

We use the fact that the grounded labelling is strongly admissible labelling, (2014).

**Theorem 106.** *Let* $\mathcal{A} = (Args, \rightharpoonup)$ *be argumentation framework and* $L_{gr}$ *its unique grounded labelling. For any argument* $a \in Args$ *if* $L_{gr}(a) = \mathtt{in}$ *then there exists a grounded discussion for* $a$ *that is won by a proponent.*

*Proof.* Let $a$ be an in-labelled argument of $L_{gr}$. We construct the grounded persuasion game $\mathcal{D}$ won by proponent as follows:

1. Start by $(\mathtt{proponent}, \mathtt{claim}, \mathtt{in}(a))$.
2. Extend the game with a valid move $M$, whose labelling agrees with $L_{gr}$ and labels an argument with the lowest possible min-max number.
3. Repeat the previous step as long as possible.

Note that the above rules define a valid game because we start and extend with valid moves, and by Theorem 101 this process stops at some point. The only ambiguous steps are the following:

1. the opponent asks why $\mathtt{out}(b)$ where $b$ is one of several attackers of in-labelled arguments,
2. the proponent answers to the why $\mathtt{out}(b)$ by because $\mathtt{in}(c)$ where $c$ is one of several in-labelled attackers of $b$.

In those cases moves contain partial labelling which labels exactly one argument. It may happen that several arguments in question has the same min-max number but then choice between them is not important.

We show that at each step of the above construction the last open issues are the open issues which have the lowest min-max numbering. We state this property formally. For two partial labellings $\mathcal{L}_1$, $\mathcal{L}_2$ we will write $\mathcal{MM}(\mathcal{L}_1) \geqslant \mathcal{MM}(\mathcal{L}_2)$ iff $\forall a \in \mathrm{dom}(\mathcal{L}_1), b \in \mathrm{dom}(\mathcal{L}_2) :$ $\mathcal{MM}(a) \geqslant \mathcal{MM}(b)$. We prove by induction (with respect to the length of the game) that the following property is preserved:

$$\mathcal{MM}(OI_{\mathcal{D}} \setminus LOI_{\mathcal{D}}) \geqslant \mathcal{MM}(LOI_{\mathcal{D}}) \tag{IP}$$

(BASIS) After the first move $OI_{\mathcal{D}} = LOI_{\mathcal{D}}$ so (IP) trivially holds.

(CONSTRUCTION) Assume (IP) holds for $\mathcal{D}$ with the last move $M = (\mathcal{P}, \mathcal{T}, \mathcal{L})$. Let $\mathcal{D}'$ be an extension of $\mathcal{D}$ with move $M' = (\mathcal{P}', \mathcal{T}', \mathcal{L}')$. We have the following cases based on the type of the extending move:

($\mathcal{T}' = \mathtt{why}$) $OI_{\mathcal{D}} = OI'_{\mathcal{D}}$ and $\mathcal{L}' = LOI_{\mathcal{D}} \subseteq LOI'_{\mathcal{D}}$. By the inductive assumption $\mathcal{MM}(OI_{\mathcal{D}} \setminus LOI_{\mathcal{D}}) \geqslant \mathcal{MM}(LOI_{\mathcal{D}})$ and $\mathcal{MM}(LOI_{\mathcal{D}}) \geqslant \mathcal{MM}(\mathcal{L}')$ because by construction the minimal argument was questioned. Therefore $\mathcal{MM}(OI'_{\mathcal{D}} \setminus LOI'_{\mathcal{D}}) \geqslant \mathcal{MM}(LOI'_{\mathcal{D}})$. In other words, by construction $\mathcal{L}'$ labels a single argument with minimal min-max number among the last open issues of the previous step, which by the inductive assumption labels arguments with minimal min-max numbers among all open issues which are the same before and after extension. Therefore the last open issues labels arguments with minimal min-max numbers among arguments labelled by open issues, i.e. (IP) is preserved.

($\mathcal{T}' = \mathtt{concede}$) After the $\mathtt{concede}$ move the opponent commits to $\mathcal{L}'$ therefore it is no longer an open issue. Either $LOI_{\mathcal{D}'} = \mathcal{L}'' \cap OI_{\mathcal{D}'}$ for some $\mathcal{L}''$ used in earlier move $M''$ or there are no more open issues and game stops. In the later case (IP) is trivially fulfilled. In the case $OI_{\mathcal{D}'}$ is not empty we have $OI_{\mathcal{D}'} \subseteq OI_{\mathcal{D}''}$ because both players commitment stores increase with the length of the game ($\mathcal{CS}_{\mathcal{D}''}(\mathtt{opponent}) \subseteq \mathcal{CS}_{\mathcal{D}'}(\mathtt{opponent}), \mathcal{CS}_{\mathcal{D}''}(\mathtt{proponent}) \subseteq \mathcal{CS}_{\mathcal{D}'}(\mathtt{proponent})$) and $\mathcal{CS}_{\mathcal{D}'}(\mathtt{proponent}) \setminus \mathcal{CS}_{\mathcal{D}'}(\mathtt{opponent}) \subseteq \mathcal{CS}_{\mathcal{D}''}(\mathtt{proponent})$ because of maximality of $M''$ is the latest move with labelling not fully committed by proponent. Then we have also $LOI_{\mathcal{D}'} \subseteq LOI_{\mathcal{D}''}$. By the inductive assumption $LOI_{\mathcal{D}''}$ label arguments with minimal min-max numbering among the arguments labelled by $OI_{\mathcal{D}''}$ and it is preserved to it subsets.

($\mathcal{T}' = \mathtt{because}$) Then $\mathcal{L}' = LOI_{\mathcal{D}'}$ is a reason for $\mathcal{L} = LOI_{\mathcal{D}}$. We have two cases:

$\mathcal{L} = \mathtt{in}(v)$ Then $\mathcal{L}' = \mathtt{out}(w_1, ..., w_2)$ where $w_1, ..., w_2$ are all the attackers of $v$. By definition of min-max numbering $\mathcal{MM}(\mathcal{L}) > \mathcal{MM}(\mathcal{L}')$.

$\mathcal{L} = \mathtt{out}(v)$ Then $\mathcal{L}' = \mathtt{in}(w)$ for some $w$ attacking $v$. By the strategy $w$ has a minimal min-max number and by definition of min-max numbering $\mathcal{MM}(v) > \mathcal{MM}(w)$.

In both cases we have strict inequity $\mathcal{MM}(\mathcal{L}) > \mathcal{MM}(\mathcal{L}')$. Since by the inductive assumption $\mathcal{L}$ labels arguments with minimal min-max numberings among $OI_{\mathcal{D}}$, $OI_{\mathcal{D}'} = OI_{\mathcal{D}} \cup (\mathcal{L}' \setminus \mathcal{CS}(\mathtt{opponent}))$ and $\mathcal{L}'$ labels argument with strictly smaller min-max number than ones labelled by $\mathcal{L}$, $\mathcal{L}'$ labels argument with strictly smaller min-max number than any other argument labelled by $OI_{\mathcal{D}'}$.

We have shown that at each step of the above construction the last open issues are the open issues which have the lowest min-max numbering. Moreover, in the case when the reason of the labelling questioned in the `why` move is given it labels arguments with strictly lower min-max numbers. It means that the order of moves in the construction guarantees that a `because` move never fails because of Condition B4 - $\mathcal{L} \cap \mathrm{OI}(\mathcal{D}) = \emptyset$). Also, there always exists a reason for a labelling questioned in the `why` move because by construction the proponent plays moves compatible with $L_{gr}$ (Condition B3). Therefore the proponent can always address the `why` moves of the opponent. This means that `why` move is not the last move of the game therefore by Lemma 100 the proponent wins. □

Note that the decision points in the construction of Theorem 106 are taken by both parts. Therefore discussion testifying the inclusion of an argument in the grounded extension, although it exists, to be actually played requires the proponent and the opponent to cooperate. It can be shown that the proponent not always has a wining strategy to convince the opponent.

It can be illustrative to examine how the earlier mentioned examples of discussion games (related to Figure 20, Figure 21 and Figure 22) are handled by the formal definition of the grounded discussion. The discussions related to Figure 20 and Figure 21 are in essence instances of the formal grounded game (one difference is that in the formal game, conceding several arguments is done using a sequence of concede moves instead of in a single move) although in the example related to Figure 21 one would have to omit the $15^{th}$ and $16^{th}$ moves, since the fact that $G$ is `out` already follows from the opponent's existing commitments, where $H$ is `in`. The third discussion, that is related to the case of two arguments attacking each other (Figure 22), however, is *not* a legal grounded game. The reason is that in the fifth step ("P: `because in`$(a)$") the proponent gives a reason (`in`$(a)$) that is actually an open issue, which is explicitly forbidden by rule B4 of Definition 98.

Overall, one can observe that our approach to the grounded discussion no longer relies on an implicit tree-like structure (as was still the case in the standard grounded game, in which this tree is essentially a winning strategy of lines of arguments Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009), or in the approach of Prakken (2005)) to be able to allow certain forms of desirable repetition (in different lines of arguments, as is the case in the example related to Figure 21) while at the same time ruling out certain forms of undesirable repetition (in the same line of arguments, as is the example related to Figure 22). By cleverly using the commitment store we made desirable repetition unnecessary, which means that all other forms of repetition are undesirable and can simply be forbidden. We simply do not need the concept of lines of arguments anymore in order to distinguish between desirable repetition and undesirable repetition. In this way, the discussion related to grounded semantics has become in line with standard dialogue theory, where one relies only on the notion of a commitment store, and not on all kinds of implicit mathematical structures (like trees or lines of arguments) to keep track of the status of the dialogue.

## 6.3 DISCUSSION

In the current chapter, we have examined how the notion of grounded semantics can be specified in terms of persuasion dialogue. Unlike for instance the standard grounded game Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009) or the approach of Prakken (2005), our dialogue game does not depend on an implicit tree-like structure, in which the moves have to fit. Also, unlike the approach in for instance Parsons et al. (2002, 2003a,b), we do not merely *apply* the concept of grounded semantics (for instance for determining what moves an agent is allowed to make, depending on its acceptance attitude) but we *characterize* it.

The aim of our work, as well as that of Caminada (2010) is to build a connection between two lines of argumentation research: argumentation as a basis for specifying nonmonotonic inference Dung (1995); Caminada and Amgoud (2007); Prakken (2010); Gorogiannis and Hunter (2011) and argumentation as a dialectical process of structured discussion Hamblin (1970, 1971); Mackenzie (1979, 1990); Yuan et al. (2003). The idea is that argumentation as nonmonotonic inference can be specified *by means of* structured discussion.

It should be observed that although the dialogue game described in the current chapter should not be seen as an algorithm or proof procedure. Although we prove that an argument is in the grounded extension iff the proponent can win the dialogue game for it, we did not provide any procedure for actually *finding* such a dialogue game. Moreover, it might very well be that several such dialogue games exist, of different length. If the aim is for an information system to use dialogue to convince the user to accept a particular argument, then it can have advantages to try to aim for the shortest dialogue that has this effect. The issue of how to find such a shortest dialogue has been left open as a topic for further research.

The prospect of interpreting inference-based argumentation theory in terms of dialogue also opens various other research issues. One of the issues that could for instance be explored is how the dialogue games work out in the context of instantiated argumentation Governatori et al. (2004); Caminada and Amgoud (2007); Wu et al. (2009); Prakken (2010) where an argument is essentially composed of a number of reasons (usually represented as rules) that collectively support a particular conclusion. An argument, in this sense, can be regarded as a defeasible proof. Since approaches like Governatori et al. (2004); Caminada and Amgoud (2007); Wu et al. (2009); Prakken (2010) all apply standard Dung-style semantics at the abstract level, it would be possible to directly apply a dialogue game for abstract argumentation, as is for instance specified in the current chapter. This, however, would imply that a move would consist of an entire argument, for instance an entire tree of rules in case of Caminada and Amgoud (2007); Wu et al. (2009); Prakken (2010). A more natural approach might be to aim for a smaller grain size. Instead of a move consisting of an entire argument (an entire aggregate of rules) one might opt to have a move consisting of a single rule (or, alternatively, of a single premise or assumption), which would allow for an argument to be "rolled off" in a gradual way, starting from its top-conclusion. Ideally, what one would want is that the ability to win such a dialogue coincides with what would be entailed by an instantiated argumentation approach like Governatori et al. (2004); Caminada and Amgoud (2007); Wu et al. (2009); Prakken (2010). We think that the approach in the current chapter as well as in Caminada (2010), where the ability to win a dialogue coincides with standard Dung-style semantics on the abstract level, could serve as a basis for examining more fine-grained approaches.

The approach in the current chapter assumes that both players have access to the same argumentation framework. Although this is a necessary assumption if one's aim is to describe existing argumentation semantics, it could be interesting to examine what happens if one would weaken this assumption and allow for private information to be released during the course of the dialogue. Unlike approaches like Cayrol et al. (2010); Liao et al. (2011) argumentation updates would become available in an interactive way, allowing parties to react and dynamically search for counterarguments to react on the newly received information, which could give rise to a whole range of strategic aspects Rahwan et al. (2009); S. Pan and Rahwan (2010). Overall, the approach would be similar to what humans tend to do when they disagree: to discuss and see whether one has information that can convince the other.

# 7

# IMPLEMENTING CRASH-RESISTANCE AND NON-INTERFERENCE IN LOGIC-BASED ARGUMENTATION

## 7.1 INTRODUCTION

The field of formal argumentation can be traced back to the work of Pollock (1992, 1995), Vreeswijk (1993, 1997), and Simari and Loui (1992). The idea is that (non-monotonic) reasoning can be performed by constructing and evaluating arguments, which are composed of a number of reasons for the validity of a claim. Arguments distinguish themselves from proofs by the fact that they are defeasible, that is, the validity of their conclusions can be disputed by other arguments. Whether a claim can be accepted therefore depends not only on the existence of an argument that supports this claim, but also on the existence of possible counter arguments, that can then themselves be defeated by counter arguments, etc.

Nowadays, much research on the topic of argumentation is based on the abstract argumentation theory of Dung (1995). The central concept in this work is that of an argumentation framework, which is essentially a directed graph in which the arguments are represented as nodes and the defeat relation is represented by the arrows. Given such a graph, one can then examine the question which set(s) of arguments can be accepted: answering this question corresponds to defining an *argumentation semantics*. Various proposals have been formulated in this respect, and in the current thesis we will describe some of the mainstream approaches. It is, however, important to keep in mind that the issue of argumentation semantics is only one specific aspect (although an important one) in the overall theory of formal argumentation. For instance, if one wants to use argumentation theory for the purpose of (non-monotonic) entailment, one can distinguish three steps (see Figure 23). First of all, one would use an underlying knowledge base to generate a set of arguments and determine in which ways these arguments defeat each other (step 1). The result is an argumentation framework, represented as a directed graph in which the internal structure of the arguments, as well as the nature of the defeat relation has been abstracted away. Based on such an argumentation framework, the next step is to determine the sets of arguments that can be accepted, using a pre-defined criterion called an argumentation semantics (step 2). After the set(s) of accepted arguments have been identified, one then has to identify the set(s) of accepted conclusions (step 3), for which there exist various approaches.

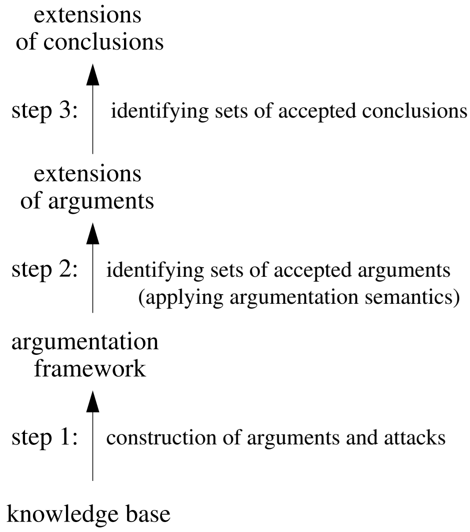



Figure 23: Argumentation for inference

As illustrated in Figure 23, the argumentation approach provides a graph based way of performing non-monotonic reasoning. An interesting phenomenon is that the non-monotonicity is isolated purely in step 2 of the process. Step 1 is monotonic (having additional information in the knowledge base yields an argumentation framework with zero or more additional vertices and edges), just like step 3 is monotonic (having additional arguments in an argument-based extension yields an associated conclusion-based extension with zero or more additional conclusions). Step 2, however, is non-monotonic because adding new arguments and extending the defeat relation can change the status of arguments that were already present in the argumentation framework when it comes to determining the argument-based extensions. That is, when adding new arguments and extending the defeat relation it is by no means guaranteed that the resulting argument-based extensions will be supersets of the previous argument-based extensions. Apart from isolating non-monotonicity in step 2, the argumentation approach to NMR also offers the advantage of different levels of abstraction. The field of abstract argumentation, for instance, only studies step 2 of the overall argumentation process and has now become one of the most popular topics in argumentation research.

Despite its advantages, the argumentation approach to non-monotonic reasoning also has important difficulties that are often overlooked by those studying purely abstract argumentation. The point is that in step 1 of the overall argumentation process, one constructs arguments that have a logical content. Yet, in step 2, one selects the sets of accepted arguments (argument-based extensions) based purely on some topological principle of the resulting graph, without looking what is actually inside of the arguments. The abstract level (step 2) is essentially about how to apply a semantics "blindly", without looking at the logical content of the arguments. But if one cannot see what is inside of the arguments, then how can one make sure that the selected set of arguments makes sense from a logical perspective? For instance, how can one be sure that the conclusions yielded by these sets of arguments (step 3) will be consistent?[1] Or, alternatively, how does one know that these conclusions will actually be closed under logical entailment?

Issues like that of consistency and closure of argumentation-based entailment cannot be handled purely at the level of any of the individual three steps in the overall argumentation process. Instead, they require a carefully selected *combination* of how to carry out *each* of these individual steps. For instance, Caminada and Amgoud (2005; 2007) point out that when applying the argumentation process to a knowledge base consisting of strict and defeasible rules, one can obtain closure and consistency of the resulting conclusions by applying transposition and restricted rebut when constructing the argumentation framework (step 1), in combination with any complete-based argumentation semantics (step 2). Under these conditions, the conclusions associated with the argument-based extensions (step 3) will be consistent and closed under the strict rules Caminada and Amgoud (2007).

Caminada and Amgoud introduce three postulates that they aim to satisfy for argument-based entailment: *Direct Consistency*, *Indirect Consistency* and *Closure*. The current chapter extends this line of research by examining two additional postulates: *Crash-resistance* and *Non-interference* Caminada et al. (2012). It is explained why these postulates matter, and how they are in fact violated by several well-known formalisms for argument-based entailment (including Pollock's OSCAR system Pollock (1995)). We focus on a formalism called ASPICLite, but our findings are also relevant for other formalisms that aim to combine classical logic with defeasible rules, such as Pollock (1995); Prakken (2010); Reiter (1980). Furthermore, we provide a general way of satisfying the postulates of Caminada and Amgoud (2005; 2007) as well as the additional postulates discussed in this chapter, in the context of argumentation formalisms that apply defeasible argument schemes in combination with classical logic.

The remaining part of this chapter is structured as follows. In Section 7.3, the postulates of Non-interference and Crash-resistance are described, and it is examined how these are violated by formalisms

[1] To make an analogy, consider the (fictitious) case of uncle Bob who lives in a retirement home. Every day, he has to take a number of medicines, which come in small bottles that a nurse puts on the table for him. However, some combinations of medicines are poisonous when taken at the same time. Having lost his reading glasses, uncle Bob is unable to read the labels, to determine the actual contents of the medicines. Instead, he chooses which medicines to take purely on how the bottles have been arranged on his table, hoping that the nurse somehow knows his selection criterion and has arranged the bottles accordingly.

like ASPICLite, OSCAR Pollock (1995) and ASPIC Prakken (2010). In Section 7.4, we provide a general solution to satisfy both the postulates introduced in Caminada and Amgoud (2005, 2007) (Direct Consistency, Indirect Consistency and Closure) and the additional postulates examined in the current chapter (Non-interference and Crash-resistance Caminada et al. (2012)). The discussion is then rounded off with some concluding remarks in Section 7.6.

## 7.2 PRELIMINARIES

In this section, we briefly introduce some preliminaries. Following the three-step argumentation process in Figure 23, we first show how to construct an argumentation framework based on the argumentation formalism introduced by Caminada and Amgoud (2007) (step 1). Then we discuss Dung's abstract argumentation semantics (step 2). Finally (step 3), we review how to determine sets (extensions) of accepted conclusions based on the sets (extensions) of accepted arguments. Based on this three-step process, then we restate postulates of Direct Consistency, Indirect Consistency and Closure. Finally, we introduce the ASPICLite system.

### 7.2.1 *The Three-Step Argumentation Process*

#### 7.2.1.1 *Step 1: constructing the argumentation framework*

Given a knowledge base, the question becomes how to construct the associated argumentation framework. In order to illustrate how to construct a Dung-style abstract argumentation framework from a knowledge base, we discuss an argumentation formalism introduced by Caminada and Amgoud (2007). The knowledge base in this case corresponds to the defeasible theory $\mathcal{T}$.

In the following, let $\mathcal{L}$ be a set of literals, closed under negation. Arguments consist of strict or defeasible rules Lin and Shoham (1989); Pollock (1987); Vreeswijk (1993).

**Definition 107** (Strict and defeasible rules (Caminada and Amgoud (2007))). *Let $\varphi_1, \ldots, \varphi_n, \varphi \in \mathcal{L}$ $(n \geqslant 0)$.*

- *A* strict rule *is of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$.*
- *A* defeasible rule *is of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$.*

$\varphi_1, \ldots, \varphi_n$ are called the *antecedents* of the rule and $\varphi$ its *consequent*. A strict rule of the form $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ indicates that if $\varphi_1, \ldots, \varphi_n$ hold, then without exception it holds that $\varphi$. A defeasible rule of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ indicates that if $\varphi_1, \ldots, \varphi_n$ hold, then *usually* it holds that $\varphi$.

**Definition 108** (Closure of a set of formulas (Definition 5 in Caminada and Amgoud (2007))). *Let $\mathcal{P} \subseteq \mathcal{L}$. The closure of $\mathcal{P}$ under the set $\mathcal{S}$ of strict rules, denoted $\mathcal{C}l_{\mathcal{S}}(\mathcal{P})$, is the smallest set such that:*

- $\mathcal{P} \subseteq \mathcal{C}l_{\mathcal{S}}(\mathcal{P})$.
- *if* $\phi_1, \ldots, \phi_n \to \psi \in \mathcal{S}$ *and* $\phi_1, \ldots, \phi_n \in \mathcal{C}l_{\mathcal{S}}(\mathcal{P})$ *then* $\psi \in \mathcal{C}l_{\mathcal{S}}(\mathcal{P})$.

The consistency of a set in $\mathcal{L}$ is defined according to classical negation.

**Definition 109** (Consistent set (Definition 6 in Caminada and Amgoud (2007)) ). *Let $\mathcal{P} \subseteq \mathcal{L}$. $\mathcal{P}$ is consistent iff $\nexists\ \psi, \varphi \in \mathcal{P}$ such that $\psi = \neg\varphi$.*

**Definition 110.** *A defeasible theory $\mathcal{T}$ is a pair $(\mathcal{S}, \mathcal{D})$ where $\mathcal{S}$ is a set of strict rules and $\mathcal{D}$ is a set of defeasible rules.*

Arguments are built from a defeasible theory according to the recursive definition below. In the base case arguments are constructed from rules with empty antecedent. In other cases arguments are constructed by applying strict or defeasible rule to previously constructed arguments. Along the definition we define several helper functions. The conclusion of an argument is returned by a function `Conc` which is the consequent of the root rule of the argument. `Sub` returns all subarguments of the argument and functions `StrictRules` and `DefRules` return all the strict rules and the defeasible rules respectively.

**Definition 111** (Argument (Definition 7 in Caminada and Amgoud (2007))). *Let $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ be a defeasible theory. An argument $A$ is:*

- $A_1, \ldots, A_n \to \psi$ $(n \geqslant 0)$ *if* $A_1, \ldots, A_n$ *are arguments such that there exists a strict rule* $r \in \mathcal{S}$ *and* $r = \mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \to \psi$,

  $\mathtt{Conc}(A) = \psi$,

  $\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n) \cup \{A\}$,

  $\mathtt{TopRule}(A) = r$,

  $\mathtt{DefRules}(A) = \mathtt{DefRules}(A_1) \cup \ldots \cup \mathtt{DefRules}(A_n)$,

  $\mathtt{StrictRules}(A) = \mathtt{StrictRules}(A_1) \cup \ldots \cup \mathtt{StrictRules}(A_n) \cup \{r\}$.

- $A_1, \ldots, A_n \Rightarrow \psi$ $(n \geqslant 0)$ *if* $A_1, \ldots, A_n$ *are arguments such that there exists a defeasible rule* $r \in \mathcal{D}$ *and* $r = \mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \Rightarrow \psi$,

  $\mathtt{Conc}(A) = \psi$,

  $\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \ldots \cup \mathtt{Sub}(A_n) \cup \{A\}$,

  $\mathtt{TopRule}(A) = r$,

  $\mathtt{DefRules}(A) = \mathtt{DefRules}(A_1) \cup \ldots \cup \mathtt{DefRules}(A_n) \cup \{r\}$,

  $\mathtt{StrictRules}(A) = \mathtt{StrictRules}(A_1) \cup \ldots \cup \mathtt{StrictRules}(A_n)$.

*Let Args be the set of all arguments that can be built from* $\mathcal{T}$ *and let* $A, A' \in$ *Args.*

- $A'$ *is a* subargument *of* $A$ *iff* $A' \in \texttt{Sub}(A)$.
- $A'$ *is a* direct subargument *of* $A$ *iff* $A' \in \texttt{Sub}(A)$, $A' \neq A$, $\nexists A'' \in$ *Args such that* $A'' \in \texttt{Sub}(A)$ *and* $A' \in \texttt{Sub}(A'')$, $A \neq A''$ *and* $A' \neq A''$.
- $A$ *is an* atomic argument *iff* $\nexists A' \in$ *Args*, $A' \neq A$ *and* $A' \in \texttt{Sub}(A)$.
- *The* depth *of* $A$ *(*$\text{depth}(A)$*) is* $1$ *if* $A$ *is an atomic argument, or else* $1 + \text{depth}(A')$ *where* $A'$ *is a direct subargument of* $A$ *such that* $\text{depth}(A')$ *is maximal.*

We extend Sub to a set of arguments, i.e. $\texttt{Sub}(\{A_1, \ldots, A_n\}) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n)$ where $A_1, \ldots, A_n \in$ *Args*.

An argument is strict if it is constructed only by strict rules, otherwise it is defeasible.

**Definition 112** (Strict argument and defeasible argument (Definition 8 in Caminada and Amgoud (2007))). *An argument* $A$ *is*

- *strict if* $\texttt{DefRules}(A) = \emptyset$*;*
- *defeasible if* $\texttt{DefRules}(A) \neq \emptyset$*;*

An argument can defeat arguments in two different ways: undercutting and rebutting.

The definition of undercutting, taken from Caminada and Amgoud (2007), applies the objectification operator ($\lceil \ldots \rceil$) introduced by Pollock. The idea is to translate a meta-level expression (in our case: a rule) to an object-level expression (in our case: an element of $\mathcal{L}$) Pollock (1992, 1995). Undercutting an argument means that there is a defeasible rule in the argument that is claimed not to be applicable.

**Definition 113** (Undercutting (Definition 10 in Caminada and Amgoud (2007))). *Argument* $A$ undercuts *argument* $B$ *(on* $B'$*) iff* $\texttt{Conc}(A) = \neg\lceil \texttt{Conc}(B''_1), \ldots, \texttt{Conc}(B''_n) \Rightarrow \psi \rceil$ *for some* $B' \in \texttt{Sub}(B)$ *of the form* $B''_1, \ldots, B''_n \Rightarrow \psi$.

Rebutting an argument means that there is a contrary conclusion of the conclusion of a defeasible rule in the argument so that the conclusion of the defeasible rule is argued against.

**Definition 114** (Rebutting (Definition 15 in Caminada and Amgoud (2007))). *Argument* $A$ rebuts *argument* $B$ *(on* $B'$*) iff* $\texttt{Conc}(A) = \neg\varphi$ *for some* $B' \in \texttt{Sub}(B)$ *of the form* $B''_1, \ldots, B''_n \Rightarrow \varphi$.

Using the notions of rebut and undercut, one can then subsequently define the notion of defeat.

**Definition 115** (Defeat (Definition 16 in Caminada and Amgoud (2007)))**.** *Argument* A defeats *argument* B *iff* A *rebuts* B *or* A *undercuts* B*.*

Note that a defeat is determined by a conclusion of a defeating argument and is always aimed at some defeasible rule of a defeated argument. Among two arguments with contrary conclusions, one of which has a strict top rule and the other a defeasible top rule, the argument with the strict top rule rebuts asymmetrically the one with the defeasible top rule.

Overall, given a defeasible theory $(\mathcal{S}, \mathcal{D})$, one can construct the associated argumentation framework by applying Definition 111 and Definition 115.

**Definition 116** (Argumentation framework)**.** *An* abstract argumentation framework $\mathcal{A}$ *built from a defeasible theory* $\mathcal{T}$ *is a pair* $(Args, \rightharpoonup)$ *such that:*

- *Args is the set of arguments on the basis of* $\mathcal{T}$ *as defined by Definition 111,*
- $\rightharpoonup$ *is the relation on Args given by Definition 115.*

Definition 116 completes the first step in the overall argumentation process: constructing an argumentation framework given a knowledge base.

#### 7.2.1.2 *Step 2: applying abstract argumentation semantics*

Given the argumentation framework as provided at the end of step 1 (Definition 116) the next question then becomes how to determine the associated sets of arguments that can collectively be accepted. In Dung's approach, this question is answered on abstract level without looking at the logical content of the arguments. In the previous chapters we used labellings. In this chapter we use extension approach as described in Subsection 2.2.1 of Chapter 2.

#### 7.2.1.3 *Step 3: determining the sets of justified conclusions*

Depending on the particular abstract argumentation semantics, step 2 provides zero or more extensions of arguments. However, what one is often interested in for practical purposes are not so much the arguments themselves, but the *conclusions* supported by these arguments. That is, for each set (extension) of arguments, one needs to identify the associated set (extension) of conclusions.

**Definition 117.** *Let* $\mathcal{A}r$ *be a set of arguments whose structure complies with Definition 111. We define* $\texttt{Concs}(\mathcal{A}r)$ *as* $\{\texttt{Conc}(A) \mid A \in \mathcal{A}r\}$*.*

Definition 117 makes it possible to refer to the extensions of conclusions under various argumentation semantics. For instance, the extensions of conclusions under preferred semantics are simply the associated conclusions (Definition 117) of each preferred extension of arguments.

The justified conclusions Caminada and Amgoud (2007) are conclusions that are supported by at least one argument in each extension.

**Definition 118** (Justified conclusions (Definition 12 in Caminada and Amgoud (2007))). *Let* $(Args, \rightharpoonup)$ *be an argumentation framework, and* $\{E_1, \ldots, E_n\}(n \geqslant 1)$ *be its set of extensions under a given semantics subsumed by complete semantics.*

$$\text{Output} = \bigcap_{i=1,\ldots,n} \mathtt{Concs}(E_i)$$

Output *is the set of justified conclusions under the given semantics.*

### 7.2.2 *Rationality Postulates*

Caminada and Amgoud (2007) specify the rationality postulates of *Direct Consistency*, *Indirect Consistency* and *Closure*. In this section we restate the definitions of these three postulates.

The idea of Closure is that the conclusions of an argumentation framework should be complete. If there exists a strict rule $a \rightarrow b$ and $a$ is justified then $b$ should be justified too.

An argumentation framework satisfies Closure if its set of justified conclusions, as well as the set of conclusions supported by each extension are closed.

**Definition 119** (Closure (Postulate 1 in Caminada and Amgoud (2007))). *Let* $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ *be a defeasible theory and* $\mathcal{A}$ *be an argumentation framework (Definition 116) built from* $\mathcal{T}$. *Let* $E_1, \ldots, E_n$ *be its extensions and* Output *be its set of justified conclusions under a given argumentation semantics.* $\mathcal{A}$ *satisfies Closure iff:*

*(1)* $\mathtt{Concs}(E_i) = \mathcal{C}l_{\mathcal{S}}(\mathtt{Concs}(E_i))$ *for each* $1 \leqslant i \leqslant n$.

*(2)* $\text{Output} = \mathcal{C}l_{\mathcal{S}}(\text{Output})$.

The following Proposition shows that if the different sets of conclusions of the extensions are closed, then the set Output is also closed.

**Proposition 120** (Proposition 4 in Caminada and Amgoud (2007)). *Let* $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ *be a defeasible theory and* $\mathcal{A}$ *be an argumentation framework built from* $\mathcal{T}$. *Let* $E_1, \ldots, E_n$ *be extensions of* $\mathcal{A}$ *under a given semantics. Let* Output *be its set of justified conclusions. If* $\mathtt{Concs}(E_i) = \mathcal{C}l_{\mathcal{S}}(\mathtt{Concs}(E_i))$ *for each* $1 \leqslant i \leqslant n$ *then* $\text{Output} = \mathcal{C}l_{\mathcal{S}}(\text{Output})$.

An argumentation framework satisfies Direct Consistency if its set of justified conclusions is consistent and the set of conclusions of each individual extension is consistent.

**Definition 121** (Direct Consistency (Postulate 2 in Caminada and Amgoud (2007))). *Let $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ be a defeasible theory and $\mathcal{A}$ be an argumentation framework built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be extensions of $\mathcal{A}$ under a given semantics. Let* Output *be its set of justified conclusions. $\mathcal{A}$ satisfies Direct Consistency iff:*

*(1)* $\texttt{Concs}(E_i)$ *is consistent (according to Definition 109) for each* $1 \leqslant i \leqslant n$.

*(2)* Output *is consistent (according to Definition 109).*

If the closure of the set of justified conclusions is consistent and the closure of conclusions of each extension is consistent, then the argumentation framework satisfies Indirect Consistency.

**Definition 122** (Indirect consistency (Postulate 3 in Caminada and Amgoud (2007))). *Let $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ be a defeasible theory and $\mathcal{A}$ be an argumentation framework built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be extensions of $\mathcal{A}$ under a given semantics. Let* Output *be its set of justified conclusions. $\mathcal{A}$ satisfies Indirect Consistency iff:*

*(1)* $\mathcal{C}l_{\mathcal{S}}(\texttt{Concs}(E_i))$ *is consistent (according to Definition 109) for each* $1 \leqslant i \leqslant n$.

*(2)* $\mathcal{C}l_{\mathcal{S}}(\text{Output})$ *is consistent (according to Definition 109).*

It follows that if Indirect Consistency is satisfied by an argumentation framework, then the argumentation framework also satisfies Direct Consistency.

The following proposition shows that if the closure of conclusions of each extension is consistent, then the closure of the justified conclusions is consistent.

**Proposition 123** (Proposition 5 in Caminada and Amgoud (2007)). *Let $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ be a defeasible theory and $\mathcal{A}$ be an argumentation framework built from $\mathcal{T}$. Let $E_1, \ldots, E_n$ be extensions of $\mathcal{A}$ under a given semantics. Let* Output *be its set of justified conclusions. If* $\mathcal{C}l_{\mathcal{S}}(\texttt{Concs}(E_i))$ *is consistent for each* $1 \leqslant i \leqslant n$ *then* $\mathcal{C}l_{\mathcal{S}}(\text{Output})$ *is consistent.*

**Proposition 124** (Proposition 7 in Caminada and Amgoud (2007)). *Let $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ be a defeasible theory and $\mathcal{A}$ be an argumentation framework built from $\mathcal{T}$. If $\mathcal{A}$ satisfies Closure and Direct Consistency, then it satisfies Indirect Consistency.*

Caminada and Amgoud (2007) prove that if we construct an argumentation framework from the defeasible theory whose set of strict rules $\mathcal{S}$ is closed under transposition (i.e. for any strict rule $\alpha_1, \ldots, \alpha_n \rightarrow \alpha \in \mathcal{S}$ also $\alpha_1, \ldots, \alpha_{j-1}, \neg\alpha, \alpha_{j+1}, \ldots, \alpha_n \rightarrow \neg\alpha_j \in \mathcal{S}$ for $j = 1 \ldots n$) and apply any complete-based semantics (semantics such that each extension of arguments is a complete extension) according to Step 1, 2, 3, the set of conclusions yielded satisfies the postulates of Direct Consistency, Indirect Consistency and Closure.

In the next section we introduce the ASPICLite system.

### 7.2.3 *The* ASPIC Lite *System*

Rule-based argumentation formalisms, i.e. instances that support strict and defeasible rules, assume some underlying language. For example, the formalism of Caminada and Amgoud (2007) described in the previous section assumes the language $\mathcal{L}$ to be a set of literals, closed under negation. In this case the underlying language is simple but it can be more expressive e.g. propositional logic with classical inference relation $\vdash$. If the language has logical structure then there is usually an inference relation that models deductive reasoning. This is similar to the role of strict rules in rule-based argumentation formalisms. Moreover, the rule-based formalisms known so far to satisfy postulates of Consistency and Closure Caminada and Amgoud (2005); Modgil and Prakken (2013); Prakken (2010) require strict rules to be closed under transposition or contraposition. Also, it can be observed that transposition is satisfied by the classical logic entailment operator $\vdash$. This then leads to the question of why not to define strict rules in terms of $\vdash$. At least there should be some synchronisation between $\vdash$ and $\mathcal{S}$. We will explore this approach and show the possible problems related to it.

In the current section we define the ASPICLite formalism that integrates classical logic with the formalism of Caminada and Amgoud (2005, 2007). Our formalism can be specified in the ASPIC+ framework Modgil and Prakken (2013); Prakken (2010) therefore problems described in the next section also apply to at least some of formalisms specified in ASPIC+. We do not model preferences over arguments, which corresponds to using equally preferred defeasible rules in ASPIC+, because, as will be demonstrated in the last section, our solution cannot be generalised immediately to the class of preferences given by the last-link principle. We leave such generalisation for future work.

Let $\mathcal{L}$ be a propositional language. Based on Definition 107, we extend defeasible rules by explicitly showing all their undercutters.

**Definition 125.** *A* defeasible rule *is of the form* $\varphi_1, \dots, \varphi_n \underset{U}{\Rightarrow} \psi$ *where* $U \subseteq \mathcal{L}$.

This form of defeasible rules indicates that if $\neg u$ holds (for any $u \in U$), then the defeasible rule $\varphi_1, \dots, \varphi_n \underset{U}{\Rightarrow} \psi$ is inapplicable.

In this work, we use defeasible rules of the form in Definition 125. The definition of undercutting becomes the following.

**Definition 126.** *An argument* A undercuts *argument* B *on* B′ *iff* $\mathtt{Conc}(A) = \neg u$ *where* $u \in U$ *for some* $B' \in \mathtt{Sub}(B)$ *of the form* $B''_1, \dots, B''_n \underset{U}{\Rightarrow} \psi$.

In the remaining part of this chapter, we use the notion of undercutting specified by Definition 126 instead of the one specified by Definition 113.

The first step of the three-step argumentation process is constructing an argumentation framework from a given knowledge base. In the previous section the defeasible theory was playing that role. In ASPICLite we have a defeasible theory of strict and defeasible rules, as before, but the set of strict rules will be (partly) generated by standard propositional entailment. We define a knowledge base as a pair consisting of a set of premises and a set of defeasible rules.

**Definition 127** (Knowledge base)**.** *A* knowledge base *in the* ASPICLite *system is a pair* $(\mathcal{P}, \mathcal{D})$ *where* $\mathcal{P} \subseteq \mathcal{L}$ *is a propositionally consistent (i.e. there is no* $\alpha \in \mathcal{L}$ *such that* $\mathcal{P} \vdash \alpha$ *and* $\mathcal{P} \vdash \neg\alpha$ *where* $\vdash$ *is classical consequence relation)*[2] *set of premises and* $\mathcal{D}$ *is a set of defeasible rules.*

In the rest of the chapter we use $\texttt{Atoms}(\mathcal{F})$ for the atoms that occur in a set of formulas $\mathcal{F}$. For instance: $\texttt{Atoms}(\{a \wedge b, \neg b \vee c\}) = \{a, b, c\}$. Furthermore, if $At$ is a set of atoms and $\mathcal{F}$ is a set of formulas, then we write $\mathcal{F}_{|At}$ for formulas in $\mathcal{F}$ that contain only atoms from $At$. For instance: $\{a \wedge b, \neg b \vee c\}_{|\{a,b\}} = \{a \wedge b\}$. We say that two sets of formulas $\mathcal{F}_1$ and $\mathcal{F}_2$ are *syntactically disjoint* iff $\texttt{Atoms}(\mathcal{F}_1) \cap \texttt{Atoms}(\mathcal{F}_2) = \emptyset$. For a strict rule $s = \varphi_1, \ldots, \varphi_n \rightarrow \psi$, $\texttt{Atoms}(s) = \texttt{Atoms}(\{\varphi_1, \ldots, \varphi_n, \psi\})$. Similarly, for a defeasible rule $d = \varphi_1, \ldots, \varphi_n \underset{U}{\Rightarrow} \psi$, $\texttt{Atoms}(d) = \texttt{Atoms}(\{\varphi_1, \ldots, \varphi_n, \psi\}) \cup \texttt{Atoms}(U)$. $\texttt{Atoms}(\mathcal{S}) = \texttt{Atoms}(s_1) \cup \ldots \cup \texttt{Atoms}(s_n)$ where $\mathcal{S} = \{s_1, \ldots, s_n\}$ is a set of strict rules. Similarly for a set of defeasible rules $\mathcal{D} = \{d_1, \ldots, d_n\}$, $\texttt{Atoms}(\mathcal{D}) = \texttt{Atoms}(d_1) \cup \ldots \cup \texttt{Atoms}(d_n)$. Then for a knowledge base $\mathcal{B} = (\mathcal{P}, \mathcal{D})$, $\texttt{Atoms}(\mathcal{B}) = \texttt{Atoms}(\mathcal{P}) \cup \texttt{Atoms}(\mathcal{D})$. For an argument $A$, $\texttt{Atoms}(A) = \texttt{Atoms}(\texttt{StrictRules}(A)) \cup \texttt{Atoms}(\texttt{DefRules}(A))$ and for a set of arguments $\mathcal{A}r = \{A_1, \ldots, A_n\}$, $\texttt{Atoms}(\mathcal{A}r) = \texttt{Atoms}(A_1) \cup \ldots \cup \texttt{Atoms}(A_n)$.

**Definition 128** (Defeasible theory)**.** *A* defeasible theory *associated with a knowledge base* $(\mathcal{P}, \mathcal{D})$ *is a pair* $(\mathcal{S}(\mathcal{P}), \mathcal{D})$ *such that* $\mathcal{S}(\mathcal{P}) = \{\rightarrow \varphi \mid \varphi \in \mathcal{P}\} \cup \{\varphi_1, \ldots, \varphi_n \rightarrow \psi \mid \varphi_1, \ldots, \varphi_n, \psi \in \mathcal{L}$ *and* $\varphi_1, \ldots, \varphi_n \vdash \psi$ *and* $\texttt{Atoms}(\{\varphi_1, \ldots, \varphi_n, \psi\}) \subseteq \texttt{Atoms}(\mathcal{P}) \cup \texttt{Atoms}(\mathcal{D})\}$.

Note that a set of strict rules $\mathcal{S}(\mathcal{P})$ generated from premises $\mathcal{P}$ is closed under transposition, therefore from Caminada and Amgoud (2007) it follows that the ASPICLite system satisfies Closure, Direct Consistency and Indirect Consistency when used with complete-based semantics.

**Definition 129** (Argumentation framework)**.** *An* abstract argumentation framework $\mathcal{A}$ *built from a knowledge base* $\mathcal{B} = (\mathcal{P}, \mathcal{D})$ *is a pair* $(Args, \rightharpoonup)$ *such that:*

[2] The notion of *propositional consistency* used here is different from the notion of consistency defined in Definition 109. The notion from Definition 109 is used by Caminada and Amgoud in a context of language containing literals and closed under negation. We use it in preliminaries and in the formulation of postulates of Closure and Consistency, which we kept in the original form. In the context of the ASPICLite system we rely on the propositional consistency defined here, which sometimes we call simply consistency.

- *Args is the set of arguments on the basis of* $\mathcal{T} = (\mathcal{S}(\mathcal{P}), \mathcal{D})$ *as defined by Definition 111,*
- $\rightharpoonup$ *is the relation on Args given by Definition 115.*[3]

In the ASPICLite system, the definitions except the definitions of defeasible rules, undercutting, defeasible theory and argumentation framework are the same as the definitions in Section 7.2.

## 7.3 PROBLEMS AND POSTULATES

Although Closure, Direct Consistency and Indirect Consistency are three important properties of the ASPICLite system, they are not the only properties that matter. There are other postulates that an argumentation framework should satisfy to make the system stable and prevent the system from crashing because of some problematic piece of information. So Non-interference and Crash-resistance are two more postulates we will discuss in this section.

In order to illustrate the kinds of problems that we are interested in, let us first take a look at Example 130 which is essentially equivalent to an example in Pollock (1994).

**Example 130** (Caminada (2005))**.** *Consider a knowledge base* $\mathcal{B} = (\mathcal{P}, \mathcal{D})$ *where*

$$\mathcal{P} = \{\mathsf{says\_J\_s}, \quad \text{(“John says the cup of coffee contains sugar.”)}$$

$$\mathsf{says\_M\_ns}, \quad \text{(“Mary says the cup of coffee does not contain sugar.”)}$$

$$\mathsf{says\_WF\_r}\}, \quad \text{(“The weather forecaster predicts rain today.”)}$$

$$\mathcal{D} = \{\mathsf{says\_J\_s} \underset{\{\mathsf{rel\_J}\}}{\Rightarrow} \mathsf{s}, \quad \text{(“If John says the cup of coffee contains sugar then the cup of coffee probably contains sugar.”)}$$

$$\mathsf{says\_M\_ns} \underset{\{\mathsf{rel\_M}\}}{\Rightarrow} \neg\mathsf{s}, \quad \text{(“If Mary says the cup of coffee does not contain sugar then the cup of coffee probably does not contain sugar.”)}$$

$$\mathsf{says\_WF\_r} \underset{\{\mathsf{rel\_WF}\}}{\Rightarrow} \mathsf{r}\}. \quad \text{(“If the weather forecaster predicts rain today then it probably rains today.”)}$$

*Note that we use a propositional language. For example* says_WF_r *is a single atom representing the fact that the weather forecaster announced rain and* rel_WF *is a single atom representing the fact that the weather forecaster is reliable*[4]*. Consider the following arguments:*

$$J_1 : \mathsf{says\_J\_s} \underset{\{\mathsf{rel\_J}\}}{\Rightarrow} \mathsf{s} \qquad W_1 : \mathsf{says\_WF\_r} \underset{\{\mathsf{rel\_WF}\}}{\Rightarrow} \mathsf{r}$$

$$M_1 : \mathsf{says\_M\_ns} \underset{\{\mathsf{rel\_M}\}}{\Rightarrow} \neg\mathsf{s} \qquad JM : J_1, M_1 \rightarrow \neg\mathsf{r}$$

[3] Definition 115 referred here is meant to use the notion of undercut defined by Definition 126.

[4] Our notation differs from Caminada (2005), where an atom says_WF_r is represented by an instantiated propositional schema Says(WF, r).

*where* rel_J *means that John is reliable and* rel_M *means that Mary is reliable.*
*The resulting argumentation framework is partly illustrated in Figure 24.*

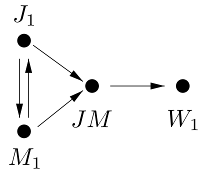



Figure 24: Arguments $J_1$ and $M_1$ "contaminate" argument $W_1$.

In the argumentation framework of Figure 24, arguments $J_1$ and $M_1$ defeat each other. As one may expect, a particularly troublesome argument is JM. It is composed of $J_1$ and $M_1$, together with the strict rule $s, \neg s \rightarrow \neg r$ (which exists because of the fact that $s, \neg s \vdash \neg r$). JM defeats (rebuts) $W_1$ and is defeated (rebutted) by $J_1$ and $M_1$. Figure 24 illustrates a general problem when trying to embed classical logic into rule-based argumentation formalisms: whenever there are two arguments that rebut each other (like $J_1$ and $M_1$) it is possible to combine them into an argument with any arbitrary conclusion (like JM) that can then be used to defeat any defeasible argument (like $W_1$).

Simply forbidding rules with inconsistent antecedents does not provide a solution. For instance, in the case of Example 130, classical logic also generates the strict rules $s \rightarrow s \vee \neg r$ and $s \vee \neg r, \neg s \rightarrow \neg r$ which can then be used (together with $J_1$ and $M_1$) to construct an argument for $\neg r$, even in the absence of $s, \neg s \rightarrow \neg r$ or any other rule with inconsistent antecedent.

The argumentation framework of Figure 24 is particularly troublesome under grounded semantics. Since every defeasible argument is defeated, the grounded extension contains strict arguments only. It means that effectively all defeasible inferences are blocked. Hence, under grounded semantics, the weather forecast is not justified because John and Mary are having a disagreement about a cup of coffee. This is of course absurd.

However, as has been observed by Prakken (2010), the problem seems to go away when using preferred semantics. In the example of Figure 24, there exist two preferred extensions: $\{J_1, W_1\}$ and $\{M_1, W_1\}$. Each of these extension contains $W_1$. A similar observation can be made for stable Dung (1995), semi-stable Caminada (2006b); Verheij (1996) semantics. So if the problem of Figure 24 only becomes serious under grounded semantics, then why not for instance use preferred semantics instead? The problem, however, is that although preferred semantics "solves" the problem of Figure 24, there exist other more complex situations where preferred semantics alone does *not* provide a solution. One of these situations is described in Example 131 (taken from Caminada (2005)).

**Example 131** (Caminada (2005))**.** *Given a knowledge base* $\mathcal{B} = (\mathcal{P}, \mathcal{D})$ *where*

$$\mathcal{P} = \{\mathtt{says\_J\_s}, \quad \text{(“John says the cup of coffee contains sugar.”)}$$
$$\mathtt{says\_M\_ns}, \quad \text{(“Mary says the cup of coffee does not contain sugar.”)}$$
$$\mathtt{says\_J\_urJ}, \quad \text{(“John says John is unreliable.”)}$$
$$\mathtt{says\_M\_urM}, \quad \text{(“Mary says Mary is unreliable.”)}$$
$$\mathtt{says\_WF\_r}\}, \quad \text{(“The weather forecaster predicts rain today.”)}$$
$$\mathcal{D} = \{\mathtt{says\_J\_s} \underset{\{\mathtt{rel\_J}\}}{\Rightarrow} \mathtt{s}, \quad \text{(“If John says the cup of coffee contains sugar then the cup of coffee probably contains sugar.”)}$$
$$\mathtt{says\_M\_ns} \underset{\{\mathtt{rel\_M}\}}{\Rightarrow} \neg\mathtt{s}, \quad \text{(“If Mary says the cup of coffee does not contain sugar then the cup of coffee probably does not contain sugar.”)}$$
$$\mathtt{says\_J\_urJ} \underset{\{\mathtt{rel\_J}\}}{\Rightarrow} \neg\mathtt{rel\_J}, \quad \text{(“If John says John is unreliable then John probably is unreliable.”)}$$
$$\mathtt{says\_M\_urM} \underset{\{\mathtt{rel\_M}\}}{\Rightarrow} \neg\mathtt{rel\_M}, \quad \text{(“If Mary says Mary is unreliable then Mary probably is unreliable.”)}$$
$$\mathtt{says\_WF\_r} \underset{\{\mathtt{rel\_WF}\}}{\Rightarrow} \mathtt{r}\}. \quad \text{(“If the weather forecaster predicts rain today then it probably rains today.”)}$$

*Consider the following arguments:*

| | | |
|---|---|---|
| $J_0 : \mathtt{says\_J\_s}$ | $M_0 : \mathtt{says\_M\_s}$ | $W_0 : \mathtt{says\_WF\_r}$ |
| $J_1 : \mathtt{says\_J\_urJ}$ | $M_1 : \mathtt{says\_M\_urM}$ | $W_1 : W_0 \underset{\{\mathtt{rel\_WF}\}}{\Rightarrow} \mathtt{r}$ |
| $J_2 : J_1 \underset{\{\mathtt{rel\_J}\}}{\Rightarrow} \neg\mathtt{rel\_J}$ | $M_2 : M_1 \underset{\{\mathtt{rel\_M}\}}{\Rightarrow} \neg\mathtt{rel\_M}$ | |
| $J_3 : J_0 \underset{\{\mathtt{rel\_J}\}}{\Rightarrow} \mathtt{s}$ | $M_3 : M_0 \underset{\{\mathtt{rel\_M}\}}{\Rightarrow} \neg\mathtt{s}$ | $JM : J_3, M_3 \rightarrow \neg\mathtt{r}$ |

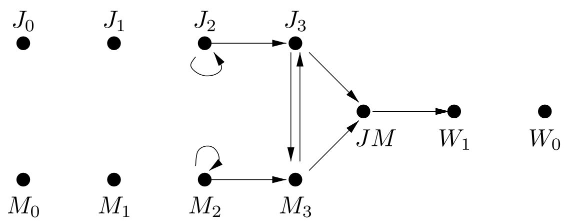



Figure 25: Switching from grounded to preferred semantics still does not always provide a solution.

*When applying the defeat relation specified by Definition 115, the argumentation framework* $\mathcal{A}$ *in Figure 25 can be built (In Figure 25, we give only the arguments that can be captured by intuition. We can use a part of the argumentation framework to illustrate the contamination of the* ASPICLite *system since all defeasible arguments can be defeated by arguments that are obtained by substituting the conclusion of the argument*

$JM$). *In this argumentation framework, only one complete extension exists* $\{J_0, J_1, M_0, M_1, W_0\}$, *which is also a preferred extension and a stable extension. So we have that the weather forecast is not justified because unreliable John and unreliable Mary are having a disagreement about a cup of coffee.*

*Here the argument* $JM$ *is the contaminating information. Without argument* $JM$, *the argumentation framework in Figure 25 can be split into two syntactically disjoint argumentation frameworks. The argumentation framework* $\mathcal{A}_r$ *on the right side part consists of two arguments* $W_0$ *and* $W_1$. *The set* $\{W_0, W_1\}$ *is the only complete extension of the argumentation framework* $\mathcal{A}_r$. *The only complete extension of the whole argumentation framework* $\{J_0, J_1, M_0, M_1, W_0\}$ *does not contain* $W_1$. *The argument* $W_1$ *which is supposed to be accepted in the small argumentation framework, is rejected in the big argumentation framework. This means that the combination of knowledge bases interferes with the reasoning results of the system. In this case, we say that this system does not satisfy the postulate of Non-interference under complete semantics. Besides, there are arguments that can be constructed by replacing the conclusion of argument* $JM$ *with an arbitrary formula. Therefore all defeasible arguments could be defeated by those arguments. The consequence is that only strict arguments can be accepted. Then the argumentation system crashes because it can only supply the set of strict arguments as the result under complete semantics.*

The aim of the current chapter is to come up with a general solution to problems like those illustrated in Figure 24 and Figure 25. However, in order to claim any general solution, we first need to precisely define what it actually is that is violated in the outcome of Example 130 (Figure 24) and Example 131 (Figure 25). To do so, we now proceed to describe *contamination*, *non-triviality* and the postulates of *Non-interference* and *Crash-resistance* Caminada et al. (2012).

The original formulation as stated by Caminada et al. (2012) was done in a very general way for an arbitrary *logical formalism*.

**Definition 132** (Logical formalism (Definition 1 in Caminada et al. (2012)))**.** *A* logical formalism *is a triple* $(\mathcal{A}\mathsf{toms}, \mathcal{F}\mathsf{ormulas}, \mathsf{Cn})$ *where* $\mathcal{A}\mathsf{toms}$ *is a countable (finite or infinite) set of atoms,* $\mathcal{F}\mathsf{ormulas}$ *is the set of all well-formed formulas that can be constructed using* $\mathcal{A}\mathsf{toms}$, *and* $\mathsf{Cn} : 2^{\mathcal{F}\mathsf{ormulas}} \to 2^{2^{\mathcal{F}\mathsf{ormulas}}}$ *is a consequence function.*

We can regard a particular argumentation system as a logical formalism with consequence function defined in the three-step process described in Section 7.2.1. In the following, we first make precise how to treat the ASPICLite formalism as a logical formalism. Then we define Non-interference, Crash-resistance, contamination and non-triviality in the specific case of the ASPICLite formalism. Finally we give a theorem that states that non-triviality and Non-interference together imply Crash-resistance.

In the case of the ASPICLite formalism, the set of atoms is the set of propositional atoms of the propositional language that is a parameter of the formalism. The relation between the knowledge base and the set of atoms is given below Definition 127. We remind that in this chapter, $\texttt{Atoms}(\mathcal{B})$ represents the atoms that occur in a knowledge base $\mathcal{B}$ and $\texttt{Atoms}(Args)$ represents the atoms that occur in a set of arguments *Args*. We say also that two knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$ are *syntactically disjoint* if $\texttt{Atoms}(\mathcal{B}_1) \cap \texttt{Atoms}(\mathcal{B}_2) = \emptyset$.

An ASPICLite knowledge base consists of a set of premises and a set of defeasible rules. To accommodate to the definition of a logical formalism, whose input is a simple set of formulas, we treat a knowledge base as a set of two types of formulas of a single language. To be precise we need to define the union of knowledge bases.

**Definition 133** (Union of knowledge bases). *Let $\mathcal{B}_1 = \langle \mathcal{P}_1, \mathcal{D}_1 \rangle$ and $\mathcal{B}_2 = \langle \mathcal{P}_2, \mathcal{D}_2 \rangle$ be two knowledge bases. The union of $\mathcal{B}_1$ and $\mathcal{B}_2$ (denoted $\mathcal{B}_1 \cup \mathcal{B}_2$) is a knowledge base $\mathcal{B} = \langle \mathcal{P}, \mathcal{D} \rangle$ such that $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ and $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$.*

Finally, we define the consequence function $Cn_{Sem}$, such that $Cn_{Sem}(\mathcal{B})$ is a set of sets of conclusions under certain argumentation semantics.

**Definition 134.** *Let $\mathfrak{B}$ be a set of all possible knowledge bases over some language $\mathcal{L}$. Let $\mathcal{B} \in \mathfrak{B}$ be a knowledge base and $\mathcal{A} = (Args, \rightharpoonup)$ be the argumentation framework built from $\mathcal{B}$. $Cn_{Sem} : \mathfrak{B} \rightarrow 2^{2^{\texttt{Concs}(Args)}}$ is a function such that $Cn_{Sem}(\mathcal{B}) = \{\texttt{Concs}(\mathcal{E}_1), \ldots, \texttt{Concs}(\mathcal{E}_n)\}$ where $\mathcal{E}_1, \ldots, \mathcal{E}_n$ ($n \geqslant 0$) are the extensions of arguments of $\mathcal{A}$ under certain semantics $Sem \in \{\mathtt{complete}, \mathtt{grounded}, \mathtt{preferred}, \mathtt{stable}, \mathtt{semi\text{-}stable}\}$.*

Now we can proceed with the description of the properties.

The idea of Non-interference Caminada et al. (2012) is that for two completely independent knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$, $\mathcal{B}_1$ should not influence the outcome with respect to the language of $\mathcal{B}_2$ and vice versa. In the specific case of the ASPICLite formalism, Non-interference can be described as follows.

**Definition 135** (Non-interference). *The ASPICLite system satisfies Non-interference under a given semantics Sem iff for every pair of syntactically disjoint knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$ it holds that $Cn_{Sem}(\mathcal{B}_1)_{|\texttt{Atoms}(\mathcal{B}_1)} = Cn_{Sem}(\mathcal{B}_1 \cup \mathcal{B}_2)_{|\texttt{Atoms}(\mathcal{B}_1)}$.*

**Example 136.** *Let $\mathcal{B}$ be the knowledge base in Example 131. $\mathcal{B}$ is the union of the following two syntactically disjoint knowledge bases $\mathcal{B}_1 = (\mathcal{P}_1, \mathcal{D}_1)$ where*

$$\mathcal{P}_1 = \{\mathsf{says\_J\_s},$$ (“John says the cup of coffee contains sugar.”)

$$\mathsf{says\_M\_ns},$$ (“Mary says the cup of coffee does not contain sugar.”)

$$\mathsf{says\_J\_urJ},$$ (“John says John is unreliable.”)

$$\mathsf{says\_M\_urM}\},$$ (“Mary says Mary is unreliable.”)

$$\mathcal{D}_1 = \{\mathsf{says\_J\_s} \underset{\{\mathsf{rel\_J}\}}{\Rightarrow} \mathsf{s},$$ (“If John says the cup of coffee contains sugar then the cup of coffee probably contains sugar.”)

$$\mathsf{says\_M\_ns} \underset{\{\mathsf{rel\_M}\}}{\Rightarrow} \neg\mathsf{s},$$ (“If Mary says the cup of coffee does not contain sugar then the cup of coffee probably does not contain sugar.”)

$$\mathsf{says\_J\_urJ} \underset{\{\mathsf{rel\_J}\}}{\Rightarrow} \neg\mathsf{rel\_J},$$ (“If John says John is unreliable then John probably is unreliable.”)

$$\mathsf{says\_M\_urM} \underset{\{\mathsf{rel\_M}\}}{\Rightarrow} \neg\mathsf{rel\_M}\},$$ (“If Mary says Mary is unreliable then Mary probably is unreliable.”)

*and* $\mathcal{B}_2 = (\mathcal{P}_2, \mathcal{D}_2)$ *where*

$$\mathcal{P}_2 = \{\mathsf{says\_WF\_r}\},$$ (“The weather forecaster predicts rain today.”)

$$\mathcal{D}_2 = \{\mathsf{says\_WF\_r} \underset{\{\mathsf{rel\_WF}\}}{\Rightarrow} \mathsf{r}\}.$$ (“If the weather forecaster predicts rain today then it probably rains today.”)

*Consider the following arguments:*

$$W_0 : \mathsf{says\_WF\_r}$$
$$W_1 : W_0 \underset{\{\mathsf{rel\_WF}\}}{\Rightarrow} \mathsf{r}$$

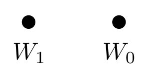



Figure 26: The argumentation framework $\mathcal{A}_2$ built from $\mathcal{B}_2$.

*There is one complete extension* $\{W_0, W_1\}$ *of the argumentation framework* $\mathcal{A}_2$ *which is built from* $\mathcal{B}_2$. *There is only one complete extension of* $\mathcal{A}$ *built from* $\mathcal{B}$ *(Example 131) which is* $\{J_0, J_1, M_0, M_1, W_0\}$. *Atoms of the knowledge base* $\mathcal{B}_2$ *are* $\{\mathsf{says\_WF\_r}, \mathsf{rel\_WF}, \mathsf{r}\}$. *Then* $\mathsf{Cn}_{Sem}(\mathcal{B})_{|\mathtt{Atoms}(\mathcal{B}_2)} = \{\{\mathsf{says\_WF\_r}\}\}$ *and* $\mathsf{Cn}_{Sem}(\mathcal{B}_2)_{|\mathtt{Atoms}(\mathcal{B}_2)} = \{\{\mathsf{says\_WF\_r}, \mathsf{r}\}\}$. *In this case* $\mathsf{Cn}_{Sem}(\mathcal{B})_{|\mathtt{Atoms}(\mathcal{B}_2)} \neq \mathsf{Cn}_{Sem}(\mathcal{B}_2)_{|\mathtt{Atoms}(\mathcal{B}_2)}$. *Therefore, the* ASPICLite *system does not satisfy the postulate of Non-interference.*

There might be a set $\mathcal{F}_1$ of formulas which determines the consequences of every set in which it is contained even if the larger set contains information syntactically disjoint from $\mathcal{F}_1$. We call this undesired phenomenon contamination and a formalism that is free from it will be said to satisfy Crash-resistance. The definition of contamination and the postulate of Crash-resistance in the case of the ASPICLite are as follows.

**Definition 137** (Contamination). *Let $\mathcal{A}toms$ be a set of atoms of the propositional language $\mathcal{L}$. The knowledge base $\mathcal{B}_1$, with $\texttt{Atoms}(\mathcal{B}_1) \subsetneq \mathcal{A}toms$, is called* contaminating *(under a given semantics Sem) iff for every knowledge base $\mathcal{B}_2$ such that $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint it holds that $Cn_{Sem}(\mathcal{B}_1) = Cn_{Sem}(\mathcal{B}_1 \cup \mathcal{B}_2)$.*

**Definition 138** (Crash-resistance). *The* ASPICLite *system satisfies* Crash-resistance *iff there does not exist a knowledge base $\mathcal{B}$ that is contaminating.*

We say that a formalism is non-trivial if for any set of atoms different knowledge bases can be constructed with different consequences. Caminada et al. (2012) proved that for any non-trivial formalism, Non-interference implies Crash-resistance. Since demonstrating non-triviality is much easier than proving Crash-resistance directly, non-triviality is a useful tool that we also use in this chapter. In the specific case of the ASPICLite formalism, non-triviality can be described as follows.

**Definition 139** (Non-trivial). *The* ASPICLite *system is called* non-trivial *iff for each nonempty set $\mathcal{A}t$ of atoms there exist knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\texttt{Atoms}(\mathcal{B}_1) = \texttt{Atoms}(\mathcal{B}_2) = \mathcal{A}t$ and $Cn_{Sem}(\mathcal{B}_1)_{|\mathcal{A}t} \neq Cn_{Sem}(\mathcal{B}_2)_{|\mathcal{A}t}$.*

Below we provide the mentioned proof in the specific case of the ASPICLite formalism.

**Theorem 140.** *If the* ASPICLite *system is non-trivial and satisfies Non-interference then it satisfies Crash-resistance.*

*Proof.* We prove by contradiction: (1) Assume the ASPICLite system does not satisfy Crash-resistance. Hence there exists a knowledge base $\mathcal{B}$ built from a strict subset of all atoms of the propositional language $\mathcal{L}$, i.e. $\texttt{Atoms}(\mathcal{B}) \subsetneq \texttt{Atoms}(\mathcal{L})$, which is contaminating; (2) By assumption the ASPICLite system is non-trivial so for the set of atoms $\mathcal{A}t$ unused in $\mathcal{B}$ ($\mathcal{A}t = \texttt{Atoms}(\mathcal{L}) \setminus \texttt{Atoms}(\mathcal{B})$), there exist two knowledge bases $\mathcal{B}_1, \mathcal{B}_2$ such that $Cn_{Sem}(\mathcal{B}_1)_{|\mathcal{A}t} \neq Cn_{Sem}(\mathcal{B}_2)_{|\mathcal{A}t}$. By construction they both are syntactically disjoint with $\mathcal{B}$; (3) By assumption the ASPICLite system satisfies Non-interference therefore $Cn_{Sem}(\mathcal{B}_1)_{|\mathcal{A}t} = Cn_{Sem}(\mathcal{B}_1 \cup \mathcal{B})_{|\mathcal{A}t}$ and $Cn_{Sem}(\mathcal{B}_2 \cup \mathcal{B})_{|\mathcal{A}t} = Cn_{Sem}(\mathcal{B}_2)_{|\mathcal{A}t}$; (4) For contaminating $\mathcal{B}$ we have $Cn_{Sem}(\mathcal{B}_1 \cup \mathcal{B})_{|\mathcal{A}t} = Cn_{Sem}(\mathcal{B})_{|\mathcal{A}t} = Cn_{Sem}(\mathcal{B}_2 \cup \mathcal{B})_{|\mathcal{A}t}$; (5) From 3 and 4 we have $Cn_{Sem}(\mathcal{B}_1)_{|\mathcal{A}t} = Cn_{Sem}(\mathcal{B}_2)_{|\mathcal{A}t}$. Contradiction with 2. □

## 7.4 SOLUTION

In this section we provide a solution to the contamination of the ASPICLite system. In Figure 25, the argument JM connects the two unconnected graphs. It leads to argument $W_1$ being affected by completely irrelevant information which is the reason of the contamination. The idea is to avoid this by deleting the inconsistent arguments.

We build an argumentation framework $\mathcal{A}$ from a knowledge base, as before, but then we prune from it all inconsistent arguments. We start by formalising the above idea and demonstrating how it deals with our running example.

An argument is inconsistent iff the set of conclusions of all its sub-arguments is propositionally inconsistent.

**Definition 141** (Consistent argument)**.** *An argument* $A \in Args$ *is* consistent *iff* $\{\mathtt{Conc}(A') \mid A' \in \mathtt{Sub}(A)\}$ *is (propositionally) consistent. Otherwise, the argument is* inconsistent*.*

A set of arguments is consistent if the set of conclusions of all sub-arguments of arguments in the set is consistent.

**Definition 142** (Consistent set of arguments)**.** *A set of arguments* $\mathcal{A}r = \{A_1, \dots, A_n\}$ *is consistent if* $\mathtt{Concs}(\mathtt{Sub}(A_1)) \cup \dots \cup \mathtt{Concs}(\mathtt{Sub}(A_n))$ *is (propositionally) consistent, otherwise* $\mathcal{A}r$ *is inconsistent.*

Note that a consistent set of arguments is different from a set of consistent arguments. For example each argument in the set $\{\rightarrow a, \Rightarrow \neg a \wedge \neg b\}$ is consistent, but it is not a consistent set of arguments since the set of conclusions $\{a, \neg a \wedge \neg b\}$ is propositionally inconsistent. Now we can define inconsistency-cleaned argumentation framework.

**Definition 143** (Inconsistency-cleaned argumentation framework)**.** *Let* $(Args, \rightharpoonup)$ *be an argumentation framework built from a knowledge base* $\mathcal{B}$*. We define* $Args_c$ *as* $\{A \mid A \in Args \text{ and } A \text{ is consistent}\}$*, and* $\rightharpoonup_c = \rightharpoonup \cap (Args_c \times Args_c)$*. We refer to* $(Args_c, \rightharpoonup_c)$ *as the inconsistency-cleaned argumentation framework built from* $\mathcal{B}$*.*

In the remaining part of the chapter, we write *the inconsistency-cleaned version* of the ASPICLite system to refer to the ASPICLite system in which the inconsistency-cleaned argumentation framework is constructed.

In the ASPICLite system we distinguish three kinds of arguments depending on their origin:

1. atomic arguments constructed from the strict rules which originate from the set of premises in the knowledge base,
2. arguments with a defeasible top rule, and
3. arguments constructed by application of a strict rule corresponding to the inference relation of the underlying logic to the previously constructed arguments.

We define a special class of arguments, which we call *flat arguments*, and then we demonstrate that for each argument of the last kind there exists a flat argument which is in essence equivalent. Later on, this will allow us to concentrate in the proofs on arguments of this special form.

**Definition 144** (Flat argument)**.** *Let $\mathcal{A} = (Args, \rightharpoonup)$ be an argumentation framework built from some knowledge base and let $A \in Args$. We say that $A$ is* flat *iff* $\texttt{TopRule}(A)$ *is strict,* $A = A_1, \ldots, A_n \rightarrow \alpha$ *and for* $A_i$ $(1 \leqslant i \leqslant n)$ *one of the following conditions holds:*

1. $\texttt{TopRule}(A_i)$ *is defeasible or,*
2. $\texttt{TopRule}(A_i)$ *has an empty antecedent and is not trivial, i.e.* $\texttt{Conc}(A_i)$ *is not a tautology.*

In other words, a flat argument has strict top rule and if that rule is applied to any subarguments with strict top rule then the strict top rule of subargument need to originate from the set of premises in the knowledge base, i.e. it has empty antecedent and a consequent which is a premise. Formally arguments with a strict top rule which have an empty antecedent are also flat, but they are not important to us.

**Lemma 145** (Flattening)**.** *Let $\mathcal{A} = (Args, \rightharpoonup)$ be an argumentation framework built from some knowledge base and let $A$ be any of its arguments with a strict top rule. There exists a flat argument $A'$ such that:*

1. $\texttt{Conc}(A') = \texttt{Conc}(A)$*,*
2. $A'^{+} = A^{+}$*,*
3. $A'^{-} = A^{-}$*,*
4. $A'$ *is propositionally consistent iff* $A$ *is propositionally consistent.*

*Proof.* Let $\mathcal{A} = (Args, \rightharpoonup)$ be an argumentation framework built from some knowledge base. Consider argument $A = A_1, \ldots, A_n \rightarrow \alpha$ with a strict top rule, and without loss of generality, let $A_1 = A^1_1, \ldots, A^1_k \rightarrow \alpha_1$ be an argument with a top rule that is strict and has non-empty antecedent (or has an empty antecedent, i.e. $k = 0$, and $\alpha_1$ is a tautology). There needs to be such an argument, otherwise $A$ is already flat. Consider an argument constructed by absorbing $A_1$ into the main argument $A' = A^1_1, \ldots, A^1_k, A_2, \ldots, A_n \rightarrow \alpha$ (for $k = 0$ absorption means removal of $A_1$). First, the strict rule used in $A'$ exists and therefore it is a proper argument. It is because strict rules with antecedent (or tautological consequences) correspond to the inference relation $\vdash$ and propositional logic satisfy *monotony* (if $\Gamma \vdash \phi$ and $\Gamma \subset \Delta$ then $\Delta \vdash \phi$) and *cut* (if $\Gamma \vdash \phi$ and $\Gamma, \phi \vdash \psi$ then $\Gamma \vdash \psi$). We can derive the needed rule as follows.

| | | |
|---|---|---|
| 1. | $a_1, ..., a_n \vdash a$ | assumption |
| 2. | $b_1, ..., b_k \vdash a_n$ | assumption |
| 3. | $b_1, ..., b_k, a_1, ..., a_{n-1} \vdash a_n$ | 2 + monotony |
| 4. | $b_1, ..., b_k, a_1, ..., a_{n-1}, a_n \vdash a$ | 1 + monotony |
| 5. | $b_1, ..., b_k, a_1, ..., a_{n-1} \vdash a$ | 3 + 4 + cut |

Second, notice that by construction, $A'$ and $A$ share the conclusion and the set of defeasible rules, i.e. $\texttt{DefRules}(A) = \texttt{DefRules}(A')$ (we have removed just one strict rule). The defeat relation in the ASPICLite system depends only on the conclusion of a defeater and the defeasible rules of the defeated argument. Therefore it follows that $A'$ and $A$ defeat and are defeated by exactly the same arguments. Third, it has exactly one less subargument, i.e. $\texttt{Sub}(A') = \texttt{Sub}(A) \setminus \{A_1\}$, and therefore if $A$ was propositionally consistent $A'$ is also propositionally consistent. Conversely, if $A'$ is propositionally consistent, i.e. $\texttt{Concs}(\texttt{Sub}(A'))$ is propositionally consistent, adding the conclusion already derivable from the set doesn't change its consistency.

If $A'$ is flat we are done. If not, we can repeat exactly the same reasoning, and as we remove one subargument each time and the number of subarguments in $A$ is finite, we will finally obtain $A'$ that is requested. □

The first three conditions assures that under a complete-based semantics, $A$ belongs to an extension if and only if $A'$ does and as they share the conclusion they contribute in the same way to the outcome. The fourth condition assures that either both $A$ and $A'$ are in the consistency-cleaned argumentation framework or neither of them is. In the rest of this chapter we will assume that we deal with flat arguments, since the non-flat arguments are essentially superfluous as they do not influence the overall outcome regarding conclusions.

Now let us see whether the problem in Example 131 can be fixed in the inconsistency-cleaned argumentation framework.

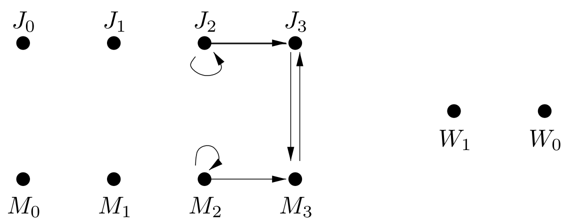



Figure 27: Repaired argumentation framework

**Example 146.** *Consider the argumentation framework generated by Example 131. We delete the argument* JM *from the argumentation framework because it is an inconsistent argument. We then obtain the argumentation framework of Figure 27. In this new argumentation framework, there is only one complete extension, and* $W_1$ *is justified under any mainstream argumentation semantics (including grounded and preferred). The weather forecast is no longer affected by a quarrel between the other two persons over a cup of coffee.*

Since our approach is to delete the class of inconsistent arguments, we have to make sure that this does not cause any problems. Hence, we need to prove not only that our approach satisfies the postulates of Non-interference and Crash-resistance, but that it also continues to satisfy the previously satisfied postulates of Closure, Direct Consistency and Indirect Consistency.

*The prior postulates*

We first demonstrate that without inconsistent arguments the ASPICLite system still satisfies all the prior postulates.

The following lemma is our main tool to demonstrate an inclusion of a particular argument $A$ of an argumentation framework (before cleaning) in a complete extension $E$ of an inconsistency-cleaned argumentation framework. The demonstration of inclusion boils down to verifying two conditions: (1) argument $A$ needs to be consistent and therefore hasn't been cleaned away; and (2) it is less vulnerable then a subset of arguments $S$ in the extension, i.e. every defeater of $A$ is also a defeater of some argument in $S$ for some $S \subseteq E$. Note that set $S$ may range from empty set (a trivial case when $A$ has no defeaters) through singleton set to the whole extension $S = E$, and of course if the property holds for some subset $S$ of the extension it also holds for the extension $E$ itself. This may appear superfluous, but we keep that formulation, because in practice the set $S$ is usually clear from the context.

**Lemma 147.** *Let $\mathcal{A} = (Args, \rightharpoonup)$ be an argumentation framework built from some knowledge base and $(Args_c, \rightharpoonup_c)$ be an inconsistency-cleaned argumentation framework built from the same knowledge base. Let $E$ be a complete extension of an inconsistency-cleaned argumentation framework, $S \subseteq E$ and $A \in Args$. If $A$ is propositionally consistent and $A^- \subseteq S^-$, i.e. every defeater of $A$ is a defeater of $S$, then $A \in E$.*

*Proof.* The argument $A$ is propositionally consistent therefore $A \in Args_c$. The extension $E$ is admissible therefore it defends its subset $S$ and also the argument $A$ because it has less defeaters then $S$. The extension $E$ is complete and therefore includes all defended arguments. Hence $A \in E$. □

The simple consequence of the lemma is that closure under subarguments is satisfied by the inconsistency cleaned version of the ASPICLite system under complete semantics.

**Lemma 148** (Subargument closure)**.** *The inconsistency-cleaned version of the ASPICLite system satisfies closure under subarguments.*

*Proof.* A subset of a consistent set is consistent, therefore a subargument of a consistent argument is consistent. Every defeater of an argument's subargument is also a defeater of the argument itself. Therefore by Lemma 147 (taking $S = \{A\}$) for every complete extension $E$ and every argument $A \in E$, $\mathtt{Sub}(A) \subseteq E$. □

The main challenge in proving satisfaction of the prior postulates is to show that each complete extension is a consistent set of arguments. We will prove it by induction on the heights of arguments. Considering an argument as a tree, the height of an argument is the number of defeasible rules on the path that uses the most defeasible rules.

**Definition 149** (Height of argument). *The height of an argument $A$ (denoted $h_d(A)$) is*

- $0$ *if $A$ consists of a single strict rule with empty antecedent.*
- $1$ *if $A$ consists of a single defeasible rule with empty antecedent.*
- $1 + \max\{h_d(B_1), \ldots, h_d(B_n)\}$ *if* $A = B_1, \ldots, B_n \Rightarrow c$ $(n \geqslant 1)$.
- $\max\{h_d(B_1), \ldots, h_d(B_n)\}$ *if* $A = B_1, \ldots, B_n \to c$ $(n \geqslant 1)$.

The following theorem shows that every complete extension of an inconsistency-cleaned argumentation framework is a propositionally consistent set of argument.

**Theorem 150.** *Let $\mathcal{B} = (\mathcal{P}, \mathcal{D})$ be a knowledge base and $\mathcal{A} = (Args, \rightharpoonup)$ be an inconsistency-cleaned argumentation framework built from $\mathcal{B}$. Let $E$ be a complete extension of $\mathcal{A}$. $E$ is propositionally consistent.*

*Proof.* Let $E$ be an arbitrary complete extension of $\mathcal{A}$. We can partition $E$ into sets of arguments that have different heights as follows:

$$E = \bigcup_{i=0..\infty} E_i \text{ where } E_i = \{A \in E \mid h_d(A) = i\}.$$

Additionally let us denote partial sums:

$$S_k = \bigcup_{0 \leqslant i \leqslant k} E_i.$$

Observations:

- Any argument $A \in E$ is finite so $h_d(A)$ is defined and so $A$ falls into exactly one $E_i$.
- If for some $i$, $E_i = \emptyset$ then $E_j = \emptyset$ for all $j > i$. This follows from Lemma 148.
- $S_i$ is closed under subarguments because $E$ is closed under subarguments and for each $A \in S_i$, it holds that if $A' \in \mathtt{Sub}(A)$ then $h_d(A') \leqslant h_d(A)$. So $A' \in S_i$. It follows that $S_i$ is consistent, i.e. $\mathtt{Concs}(\mathtt{Sub}(S_i))$ is propositionally consistent, if and only if $\mathtt{Concs}(S_i)$ is propositionally consistent.
- Let us denote arguments whose top rule doesn't correspond to the inference relation $\vdash$, i.e. is defeasible or originate in the set of premises $\mathcal{P}$, by $Elem = \{A \in Args \mid \mathtt{TopRule}(A) \in \mathcal{D}\} \cup \{\to p \mid p \in \mathcal{P}\}$. The set $S_i$ is consistent iff $S_i \cap Elem$ is consistent. It is because $\mathtt{Concs}(S_i \cap Elem) \vdash \mathtt{Concs}(S_i \setminus Elem)$.

In order to prove $E$ is consistent, it is enough to show that for any $i$, $\mathtt{Concs}(S_i \cap \mathrm{Elem})$ is propositional consistent. We prove by induction on $i$.

**Basis step:** The set $\mathtt{Concs}(S_0 \cap \mathrm{Elem}) = \mathcal{P}$ which is propositional consistent by assumption.
**Induction step:** Assume $\mathtt{Concs}(S_k \cap \mathrm{Elem})$ is consistent. We need to verify that the set $\mathtt{Concs}(S_{k+1} \cap \mathrm{Elem})$ is consistent. We prove by contradiction. Assume $\mathtt{Concs}(S_{k+1} \cap \mathrm{Elem})$ is inconsistent and let $\{c_1, \dots, c_n\} \subseteq \mathtt{Concs}(S_{k+1} \cap \mathrm{Elem})$ be an inconsistent set of conclusions supported by a finite set of arguments $\mathcal{C} \subseteq S_{k+1} \cap \mathrm{Elem}$, i.e. such that $\mathtt{Concs}(\mathcal{C}) = \{c_1, \dots, c_n\}$. There exists such set because propositional logic is compact. In case there are more such sets we choose $\mathcal{C}$ with minimal number of arguments belonging to $E_{k+1}$, i.e. minimal $\mathcal{C}_2$ where

$$\mathcal{C}_1 = \{A_1, ..., A_l\} = \mathcal{C} \cap S_k,$$
$$\mathcal{C}_2 = \{B_1, ..., B_m\} = \mathcal{C} \setminus S_k$$

is the partition of $\mathcal{C}$. Because $\{c_1, \dots, c_n\}$ is inconsistent, the inference

$$\mathtt{Conc}(A_1), \dots, \mathtt{Conc}(A_l), \mathtt{Conc}(B_1), \dots, \mathtt{Conc}(B_m) \vdash \bot$$

holds and also by transposition

$$\mathtt{Conc}(A_1), \dots, \mathtt{Conc}(A_l), \mathtt{Conc}(B_2), \dots, \mathtt{Conc}(B_m) \vdash \neg\mathtt{Conc}(B_1)$$

holds, so the strict rule

$$\mathtt{Conc}(A_1), \dots, \mathtt{Conc}(A_l), \mathtt{Conc}(B_2), \dots, \mathtt{Conc}(B_m) \rightarrow \neg\mathtt{Conc}(B_1)$$

is in $\mathcal{S}(\mathcal{P})$. Note that the set $\mathcal{C}_2$ is non-empty since $S_k$ is consistent and $\mathcal{C}$ is inconsistent. Then consider an argument

$$K = A_1, \dots, A_l, B_2, \dots, B_m \rightarrow \neg\mathtt{Conc}(B_1).$$

$K$ is consistent, otherwise the set $\mathtt{Sub}(A_1, \dots, A_l, B_2, \dots, B_m) \cap \mathrm{Elem}$ is inconsistent and it has strictly less elements belonging to $E_{k+1}$. Namely, $\big(\mathtt{Sub}(A_1, \dots, A_l, B_2, \dots, B_m) \cap \mathrm{Elem}\big) \setminus S_k = \{B_2, \dots, B_m\} \subsetneq \mathcal{C}_2$. This contradicts with the way the set $\mathcal{C}$ was chosen.

All defeaters of $K$ are also defeaters of its proper arguments which are in the extension, i.e. $K^- \subseteq (\mathtt{Sub}(K) \setminus \{K\})^-, \mathtt{Sub}(K) \setminus \{K\} \subseteq E$, because it has a strict top rule. Therefore by Lemma 147 $K \in E$.

$K$ rebuts $B_i$ because the top rule of $B_i$ is defeasible. Then $K$ defeats $B_i$ and $B_i \in E_{k+1} \subseteq E$. Then $E$ is not conflict-free. Contradiction.

So by the induction, we have proven that $\mathtt{Concs}(S_k \cap \mathrm{Elem})$ is consistent for all $k \in \mathbb{N}$. So $E$ is consistent. □

From the fact that every complete extension is propositionally consistent, it follows that the ASPICLite system without inconsistent arguments satisfies Direct Consistency.

**Theorem 151.** *The inconsistency-cleaned version of the* ASPICLite *system satisfies Direct Consistency under complete semantics.*

*Proof.* Let $\mathcal{B}$ be a knowledge base and $\mathcal{A}_c = (Args_c, \rightharpoonup_c)$ be the inconsistency-cleaned argumentation framework built from $\mathcal{B}$. Let $E$ be a complete extension of $\mathcal{A}_c$. By Theorem 150 $E$ is a consistent set of arguments, then $\mathtt{Concs}(E)$ is propositionally consistent and in particular $\mathtt{Concs}(E)$ does not contain a formula $\phi$ and a formula $\neg\phi$. □

The following theorem shows that inconsistency-cleaned argumentation frameworks satisfy Closure under complete semantics.

**Theorem 152.** *The inconsistency-cleaned version of the* ASPICLite *system satisfies Closure under complete semantics.*

*Proof.* Let $\mathcal{B} = (\mathcal{P}, \mathcal{D})$ be a knowledge base, $\mathcal{T} = (\mathcal{S}, \mathcal{D})$ be a defeasible theory associated with $\mathcal{B}$ and $\mathcal{A}_c = (Args_c, \rightharpoonup_c)$ be the inconsistency-cleaned argumentation framework built from $\mathcal{B}$. Let $E$ be a complete extension of $\mathcal{A}_c$. We need to prove that $\mathtt{Concs}(E) = Cl_{\mathcal{S}}(\mathtt{Concs}(E))$.

"$\subseteq$": This follows from the fact that $Cl_{\mathcal{S}}$ is a closure operator (Definition 108).

"$\supseteq$": Let $c_1, \ldots, c_n \to c \in \mathcal{S}$ and $\mathcal{A}r = \{C_1, \ldots, C_n\} \subseteq E$ such that for each $i \in \{1..n\}$ $\mathtt{Conc}(C_i) = c_i$. It suffices to show that argument $C = C_1, \ldots, C_n \to c$ is (1) consistent, i.e. $\mathtt{Concs}(\mathtt{Sub}(C)) = \mathtt{Concs}(\mathtt{Sub}(\mathcal{A}r)) \cup \{c\}$ is propositionally consistent, and (2) not more vulnerable than $\mathcal{A}r$, then by Lemma 147 $C \in E$. There are two cases.

Case 1: The strict rule originates from the set of premises. In this case $\mathtt{Concs}(\mathtt{Sub}(C)) = \{c\} \subseteq \mathcal{P}$ which is consistent by assumption.

Case 2: The strict rule corresponds to a propositional inference. The set $\mathtt{Concs}(\mathtt{Sub}(\mathcal{A}r))$ is consistent by Theorem 150 and the formula $c$ is a logical consequence of $\mathtt{Concs}(\mathtt{Sub}(\mathcal{A}r))$, i.e. $\mathtt{Concs}(\mathtt{Sub}(\mathcal{A}r)) \vdash c$. In this case the set $\mathtt{Concs}(\mathtt{Sub}(\mathcal{A}r)) \cup \{c\}$ is consistent because extending a consistent set with its logical consequences preserves consistency.

In both cases condition (1) holds. The condition (2) follows from the fact that every defeater of $C$ is a defeater of $\mathcal{A}r$ since $C$ cannot be defeated on its top rule which is strict. □

Since inconsistency-cleaned argumentation frameworks satisfy Closure and Direct Consistency under complete semantics, from Proposition 124, Indirect Consistency is satisfied under complete semantics.

**Theorem 153.** *The inconsistency-cleaned version of the* ASPICLite *system satisfies Indirect Consistency under complete semantics.*

*Proof.* It follows from Theorem 152 and Theorem 151. □

Now that we have proved that the inconsistency-cleaned ASPICLite system satisfies Direct Consistency and Indirect Consistency, as well as Closure, the next step is to prove that the inconsistency-cleaned ASPICLite system satisfies Non-interference and Crash-resistance.

*The new postulates*

The new postulates of Non-interference and Crash-resistance constrain the output an argumentation system produces from some knowledge base $\mathcal{B}_1$ and a knowledge base $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$. In this section we call $\mathcal{B}_1$ the *initial* knowledge base and $\mathcal{B}$ the *extended* knowledge base. Non-interference treats the situation where a knowledge base $\mathcal{B}_1$ is extended with a knowledge base $\mathcal{B}_2$ that is syntactically disjoint, i.e. $\mathtt{Atoms}(\mathcal{B}_1) \cap \mathtt{Atoms}(\mathcal{B}_2) = \emptyset$, and postulates that the output of the system produced from $\mathcal{B}_1$ should be recoverable from the output produced from the extended knowledge base $\mathcal{B}$ in a systematic way, i.e. restricting the outcome to the atoms of the initial knowledge base $\mathcal{B}_1$. Crash-resistance states that it should be possible to extend every knowledge base which does not use all atoms of the underlying language in a meaningful way, i.e. there is no input knowledge base $\mathcal{B}_1$ which determines the output of all its extensions.

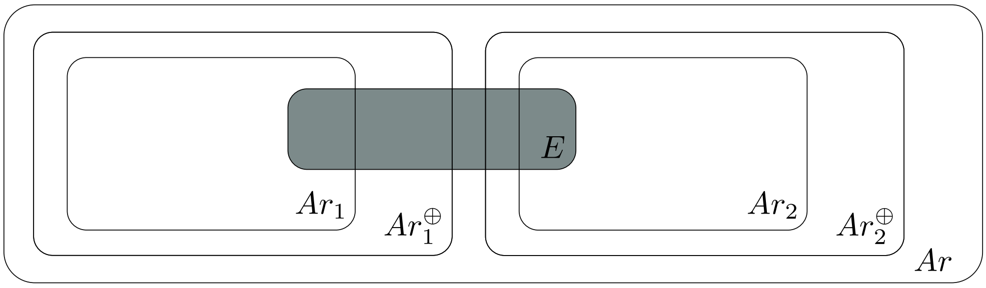



Figure 28: Different types of arguments produced from a union of two syntactically disjoint knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$

We start with a few observations about the ASPICLite system in this context. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two syntactically disjoint knowledge bases and $\mathcal{B}$ be a knowledge base such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$. Let $\mathcal{A} = (\mathit{Args}, \rightharpoonup)$, $\mathcal{A}_1 = (\mathit{Args}_1, \rightharpoonup_1)$ and $\mathcal{A}_2 = (\mathit{Args}_2, \rightharpoonup_2)$ be the argumentation frameworks built from $\mathcal{B}$, $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively. This situation is depicted in figure 28. Finally, let $E$ be an extension of $\mathcal{A}$. For $i \in \{1, 2\}$:

1. The arguments constructed from the knowledge base $\mathcal{B}_i$ are proper arguments of $\mathcal{A}$, i.e. $Ar_i \subseteq Ar$, because the construction of arguments and defeats is monotonic. We will call those arguments *pure*. Moreover, for syntactically disjoint knowledge bases $\mathcal{B}_1, \mathcal{B}_2$, pure arguments are exactly the ones built from atoms present in the knowledge base they come from, i.e. $Ar_i = \{A \in Ar \mid \mathtt{Atoms}(A) \subseteq \mathtt{Atoms}(\mathcal{B}_i)\}$.

2. There are no new defeats between $\mathit{Args}_1$ and $\mathit{Args}_2$ considered as a part of the argumentation framework $\mathcal{A}$, i.e. $\rightharpoonup_i = \rightharpoonup \cap \mathit{Args}_i \times \mathit{Args}_i$.

3. Let the set $\mathsf{Ar}_i^{\oplus}$ contain arguments whose conclusions are built from atoms of the knowledge base $\mathcal{B}_i$, i.e. $\mathsf{Ar}_i^{\oplus} = \{A \in \mathsf{Ar} \mid \mathtt{Atoms}(\mathtt{Conc}(A)) \subseteq \mathtt{Atoms}(\mathcal{B}_i)\}$. We will call those *the arguments with pure conclusions*. They are important for two reasons. First, they are the only arguments outside the argumentation framework $\mathcal{A}_i$ that can defeat arguments in the framework. This is because, to rebut or undercut an argument from $\mathcal{A}_i$ a defeater needs to have a conclusion that is a negation of the conclusion or the undercutting formula ($u \in \mathcal{U}$) which both are constructed from a subset of atoms $\mathtt{Atoms}(\mathcal{B}_i)$. Second, they are the only arguments of the $\mathcal{A}$ that influence the outcome of the system produced from the knowledge base $\mathcal{B}$ restricted to the atoms of $\mathcal{B}_i$.

4. Pure arguments are also arguments with pure conclusions.

5. For syntactically disjoint knowledge bases the sets of arguments with pure conclusions, i.e. $\mathsf{Ar}_1^{\oplus}$ and $\mathsf{Ar}_2^{\oplus}$, are disjoint.

6. There are also *other arguments* constructed by connecting previously discussed arguments with strict rules corresponding to the inference of the propositional logic.

The satisfaction of the postulates in this section is based on the following property of the construction of arguments: *for every argument* $A$ *with pure conclusion there exists a pure argument* $A'$ *with the same conclusion that is not more vulnerable, i.e.* $A'^{-} \subseteq A^{-}$. In the following lemma we formalise this property and prove that it holds for the inconsistency-cleaned version of the ASPICLite system.

**Lemma 154.** *Let* $\mathcal{B}$ *be a knowledge base such that* $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ *where* $\mathcal{B}_1 = (\mathcal{P}_1, \mathcal{D}_1)$ *and* $\mathcal{B}_2 = (\mathcal{P}_2, \mathcal{D}_2)$ *are syntactically disjoint. Let* $\mathcal{A} = (Args, \rightharpoonup)$ *and* $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ *be the inconsistency-cleaned argumentation frameworks built from* $\mathcal{B}$ *and* $\mathcal{B}_1$ *respectively. For each argument* $C \in Args$ *with pure conclusion, i.e.* $\mathtt{Atoms}(\mathtt{Conc}(C)) \subseteq \mathtt{Atoms}(\mathcal{B}_1)$*, there exists a pure argument* $C' \in Args_1$ *such that* $\mathtt{Conc}(C') = \mathtt{Conc}(C)$*,* $C'^{+} = C^{+}$ *and* $C'^{-} \subseteq C^{-}$*.*

*Proof.* We prove by induction on depths of arguments. We will construct argument $C'$ such that $\mathtt{Conc}(C') = \mathtt{Conc}(C)$ and check two properties: (1) $\mathtt{Concs}(\mathtt{Sub}(C')) \subseteq \mathtt{Concs}(\mathtt{Sub}(C))$ which guarantees that $C'$ is consistent and was not cleaned away; (2) $\mathtt{DefRules}(C') \subseteq \mathtt{DefRules}(C)$ which actually proves that $C'^{-} \subseteq C^{-}$ since defeats in the ASPICLite system are always aimed at a particular defeasible rule. Additionally we use the fact that in the ASPICLite system $\mathtt{Atoms}(C) \subseteq \mathtt{Atoms}(\mathcal{B}_1)$ is equivalent to $C \in \mathsf{Ar}_1$ and $C'^{+} = C^{+}$ follows directly from $\mathtt{Conc}(C') = \mathtt{Conc}(C)$.

**Basis step:** $\mathtt{depth}(C) = 1$. Then $C$ is an atomic argument, i.e. $C$ has empty antecedent, therefore $\mathtt{Atoms}(C) = \mathtt{Atoms}(\mathtt{Conc}(C)) \subseteq \mathtt{Atoms}(\mathcal{B}_1)$

which implies $C \in Args_1$. The conditions trivially hold for $C' = C$.

**Induction step:** Assume the condition (1) and (2) hold for each argument C such that $\texttt{depth}(C) \leqslant k$. We show that for each argument C such that $\texttt{depth}(C) = k + 1$ they also hold. There are two possible cases:

1. $\texttt{TopRule}(C)$ is defeasible. The defeasible rules belongs either to $\mathcal{D}_1$ or $\mathcal{D}_2$ and since $\texttt{Conc}(C) \subseteq \texttt{Atoms}(\mathcal{B}_1)$, the $\texttt{TopRule}(C)$ needs to belong to $\mathcal{D}_1$.

   Let $C = C_1, \ldots, C_n \underset{u}{\Rightarrow} c$. By construction, $\texttt{depth}(C_i) \leqslant k$ and $\texttt{Atoms}(\texttt{Conc}(C_i)) \subseteq \texttt{Atoms}(\mathcal{B}_1)$ because $\texttt{TopRule}(C) \in \mathcal{D}_1$ for $(1 \leqslant i \leqslant n)$. By the induction hypothesis there exist arguments $C'_1$ $(1 \leqslant i \leqslant n)$ such that:

   a) $C_i \in Ar_1$,
   b) $\texttt{Conc}(C'_i) = \texttt{Conc}(C_i)$,
   c) $\texttt{DefRules}(C'_i) \subseteq \texttt{DefRules}(C_i)$,
   d) $\texttt{Concs}(\texttt{Sub}(C'_i)) \subseteq \texttt{Concs}(\texttt{Sub}(C_i))$.

   We apply the rule $\texttt{TopRule}(C)$ to construct an argument $C' = C'_1, \ldots, C'_n \Rightarrow \texttt{Conc}(C)$.

   The required properties of $C'$ follow straightforwardly from the properties of $C'_i$ $(1 \leqslant i \leqslant n)$:

   a) $C' \in Ar_1$ because
   $$\texttt{Atoms}(C') = \texttt{Atoms}(\texttt{TopRule}(C)) \cup \bigcup_{1 \leqslant i \leqslant n} \texttt{Atoms}(C'_i) \subseteq$$
   $\texttt{Atoms}(\mathcal{B}_1)$,
   b) $\texttt{Conc}(C') = \texttt{Conc}(C)$ because the arguments C and $C'$ share the same top rule,
   c) $$\texttt{DefRules}(C') = \{\texttt{TopRule}(C)\} \cup \bigcup_{1 \leqslant i \leqslant n} \texttt{DefRules}(C'_i) \subseteq \{\texttt{TopRule}(C)\} \cup \bigcup_{1 \leqslant i \leqslant n} \texttt{DefRules}(C_i) = \texttt{DefRules}(C),$$
   d) $$\texttt{Concs}(\texttt{Sub}(C')) = \{c\} \cup \bigcup_{1 \leqslant i \leqslant n} \texttt{Concs}(\texttt{Sub}(C'_i)) \subseteq \{c\} \cup \bigcup_{1 \leqslant i \leqslant n} \texttt{Concs}(\texttt{Sub}(C_i)) = \texttt{Concs}(\texttt{Sub}(C)).$$

2. $\texttt{TopRule}(C)$ is strict. By Lemma 145 we can assume the argument $C = C_1, ..., C_n \rightarrow c$ is flat. Therefore we can partition arguments $C_i$ into two groups $I_p \cup I_d = \{1, \ldots, n\}, I_p \cap I_d = \emptyset$ such that:

   a) for $i \in I_p$ an argument $C_i$ has an empty antecedent and a consequent that corresponds to a premise, i.e. $\texttt{TopRule}(C_i) = \rightarrow c_i$ and $\texttt{Conc}(C_i) \in \mathcal{P}_1 \cup \mathcal{P}_2$,

b) for $i \in I_d$ an argument $C_i$ has defeasible top rule, i.e. $\texttt{TopRule}(C_i) \in \mathcal{D}_1 \cup \mathcal{D}_2$.

Since knowledge bases $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint, $\texttt{Conc}(C_i)$ (for $i \in I_p$) and $\texttt{TopRule}(C_i)$ (for $i \in I_d$) belong to exactly one knowledge base. This allows for another partition of direct subarguments of C according to the atoms from which their conclusions are built. Let $I_1 \cup I_2 = \{1, \dots, n\}, I_1 \cap I_2 = \emptyset$ such that:

a) for $i \in I_1$ it holds that $\texttt{Atoms}(\texttt{Conc}(C_i)) \subseteq \texttt{Atoms}(\mathcal{B}_1)$,

b) for $i \in I_2$ it holds that $\texttt{Atoms}(\texttt{Conc}(C_i)) \subseteq \texttt{Atoms}(\mathcal{B}_2)$.

Let $I_p \cap I_1 = \{p_1, \dots, p_k\}$, $I_d \cap I_1 = \{d_1, \dots, d_l\}$ and $I_2 = \{o_1, \dots, o_m\}$. Consider an argument

$$C' = C_{p_1}, \dots, C_{p_k}, C'_{d_1}, \dots, C'_{d_l} \rightarrow c$$

where arguments $C'_i$ are the pure arguments corresponding to arguments $C_i$ for $i \in I_d \cap I_1$. Those arguments exist by the induction hypothesis because arguments $C_i$ for $i \in I_d \cap I_1$ have conclusions built from atoms of the first knowledge base and are of height at most k. Now consider the propositional inference

$$\begin{aligned}&\texttt{Conc}(C_{p_1}), \dots, \texttt{Conc}(C_{p_k}),\\ &\texttt{Conc}(C_{d_1}), \dots, \texttt{Conc}(C_{d_l}),\\ &\texttt{Conc}(C_{o_1}), \dots, \texttt{Conc}(C_{o_m}) \vdash c\end{aligned}$$

corresponding to the strict rule $\texttt{TopRule}(C)$. Since

$$\texttt{Atoms}(\{\texttt{Conc}(C_{o_1}), \dots, \texttt{Conc}(C_{o_m})\}) \cap \texttt{Atoms}(c) = \emptyset$$

it needs to hold that

$$\texttt{Conc}(C_{p_1}), \dots, \texttt{Conc}(C_{p_k}), \texttt{Conc}(C_{d_1}), \dots, \texttt{Conc}(C_{d_l}) \vdash c$$

or $\{\texttt{Conc}(C_{o_1}), \dots, \texttt{Conc}(C_{o_m})\}$ is inconsistent. The set $\{\texttt{Conc}(C_{o_1}), \dots, \texttt{Conc}(C_{o_m})\}$ cannot be inconsistent because argument C is consistent, therefore

$$\texttt{Conc}(C_{p_1}), \dots, \texttt{Conc}(C_{p_k}), \texttt{Conc}(C_{d_1}), \dots, \texttt{Conc}(C_{d_l}) \vdash c$$

needs to be a valid propositional inference rule. Since $\texttt{Conc}(C'_i) = \texttt{Conc}(C_i)$ for $i = d_1 \dots d_l$ it follows that $\texttt{TopRule}(C')$ is a valid rule. This finishes the demonstration that the argument $C'$ is valid.

It remains to show that $C'$ has the requested properties. It has the same conclusion as the argument C by construction. It is pure because it has a pure conclusion, subarguments $C_{p_1}, \dots, C_{p_k}$

are pure by construction and subarguments $C'_{d_1}, \ldots, C'_{d_l}$ are pure by induction hypothesis. The sets of defeasible rules used for construction of argument $C'$ is a subset of the set of defeasible rules used for construction of argument $C$, since we erased some of its subarguments and replaced others with ones where subset of defeasible rules were used. For the same reasons, excluding argument $C'$ itself which has same conclusion as $C$, the set of subarguments of $C'$ is a subset of the set of subarguments of $C$. From that follows $\mathtt{Concs}(\mathtt{Sub}(C')) \subseteq \mathtt{Concs}(\mathtt{Sub}(C))$.

□

The proof above is the only place where we use the fact that $\mathcal{A}$ is inconsistency-cleaned. This assures $C$ is consistent allows us to construct a pure argument $C'$. The argument JM from the leading example has a pure conclusion, but has no corresponding pure argument.

**Lemma 155.** *Let $\mathcal{B}$ be a knowledge base such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ where $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint. Let $\mathcal{A} = (Args, \rightharpoonup)$, $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ be the inconsistency-cleaned argumentation frameworks built from $\mathcal{B}$ and $\mathcal{B}_1$ respectively. For any complete extension $E \subseteq Args$ of $\mathcal{A}$ we have*

$$\mathtt{Concs}(E \cap Ar_1) = \mathtt{Concs}(E)_{|\mathtt{Atoms}(\mathcal{B}_1)}.$$

*Proof.* "$\subseteq$": We have $\mathtt{Concs}(E \cap Ar_1) \subseteq \mathtt{Concs}(E)$ because $\mathtt{Concs}$ is monotonic. Filtering both sides and noticing that it is vacuous on the left side, since not only conclusions but whole arguments in $Ar_1$ are built from the atoms $\mathtt{Atoms}(\mathcal{B}_1)$, we have $\mathtt{Concs}(E \cap Ar_1) = \mathtt{Concs}(E \cap Ar_1)_{|\mathtt{Atoms}(\mathcal{B}_1)} \subseteq \mathtt{Concs}(E)_{|\mathtt{Atoms}(\mathcal{B}_1)}$.

"$\supseteq$": For any $\alpha \in \mathtt{Concs}(E)_{|\mathtt{Atoms}(\mathcal{B}_1)}$ there exists argument $A \in E$ such that $\mathtt{Conc}(A) = \alpha$ and $\mathtt{Atoms}(\alpha) \subseteq \mathtt{Atoms}(\mathcal{B}_1)$. By Lemma 154 there exists an argument $A' \in Ar_1$ with the same conclusion, which is less or equally vulnerable than $A$, i.e. $A'^{-} \subseteq A^{-}$. Since $E$ is a complete extension $A' \in E$ and so $A' \in E \cap Ar_1$. Therefore $\alpha \in \mathtt{Concs}(E \cap Ar_1)$. □

Please recall Definition 4 of the characteristic function $F_{\mathcal{A}}$ associated with the argumentation framework $\mathcal{A} = (Args, \rightharpoonup)$, which for a given set of arguments $\mathcal{A}r \subseteq Args$ returns a set of arguments that are defended by $\mathcal{A}r$ in $\mathcal{A}$. The following lemma expresses the characteristic function $F_{\mathcal{A}_1}$ associated with the subframework $\mathcal{A}_1$ in terms of the characteristic function $F_{\mathcal{A}}$ associated with $\mathcal{A}$.

**Lemma 156.** *Let $\mathcal{B}$ be a knowledge base such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ where $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint. Let $\mathcal{A} = (Args, \rightharpoonup)$, $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ be the inconsistency-cleaned argumentation frameworks built from $\mathcal{B}$ and $\mathcal{B}_1$ respectively. For any set of arguments $\mathcal{A}r \subseteq Args_1$ we have*

$$F_{\mathcal{A}_1}(\mathcal{A}r) = F_{\mathcal{A}}(\mathcal{A}r) \cap Args_1.$$

*Proof.* "$\supseteq$": Let $A \in Args_1$ be defended by $\mathcal{A}r$ in $\mathcal{A}$. All defeaters of A from $\mathcal{A}_1$ defeat A in $\mathcal{A}$ as well, therefore $\mathcal{A}r$ defends against them.

"$\subseteq$": Let $A \in Args_1$ be defended by $\mathcal{A}r$ in $\mathcal{A}_1$. Take any argument $B \in Args$ defeating A in $\mathcal{A}$. By Lemma 154 there exists an argument $B' \in Ar_1$ with the same conclusion, which is less or equally vulnerable than B, i.e. $B'^{-} \subseteq B^{-}$. Since $B'$ has the same conclusion it is also a defeater of A. Since A is defended by $\mathcal{A}r$ in $\mathcal{A}_1$ there exists argument $C \in \mathcal{A}r$ defeating $B'$ and because B is at least as vulnerable as $B'$ it also defeats B. Therefore A is defended by $\mathcal{A}r$ in $\mathcal{A}$. □

Now we will use the above lemmas to prove that a complete extension of the argumentation framework constructed from the extended knowledge base restricted to the arguments of its subframework constructed from the initial knowledge base is a complete extension of that subframework.

**Lemma 157.** *Let $\mathcal{B}$ be a knowledge base such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ where $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint. Let $\mathcal{A} = (Args, \rightharpoonup)$, $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ be the inconsistency-cleaned argumentation frameworks built from $\mathcal{B}$ and $\mathcal{B}_1$ respectively. If E is a complete extension of $\mathcal{A}$ then $E \cap Ar_1$ is a complete extension of $\mathcal{A}_1$.*

*Proof.* $E \cap Ar_1$ is conflict-free in $\mathcal{A}_1$, because E is conflict-free in $\mathcal{A}$ and there are no new defeats in $\mathcal{A}_1$, since the construction step is monotonic.

From the monotony of $F_{\mathcal{A}}$ and the fact that E is a complete extension of $\mathcal{A}$ we have $F_{\mathcal{A}}(E \cap Ar_1) \subseteq F_{\mathcal{A}}(E) = E$. Intersecting both sides we obtain $F_{\mathcal{A}}(E \cap Ar_1) \cap Ar_1 \subseteq E \cap Ar_1$. Using equivalence $F_{\mathcal{A}}(E \cap Ar_1) \cap Ar_1 = F_{\mathcal{A}_1}(E \cap Ar_1)$ from Lemma 156, we obtain $F_{\mathcal{A}_1}(E \cap Ar_1) \subseteq E \cap Ar_1$.

It remains to show the reverse inclusion $E \cap Ar_1 \subseteq F_{\mathcal{A}_1}(E \cap Ar_1)$, i.e. the set $E \cap Ar_1$ is an admissible set of $\mathcal{A}_1$. Let $A \in Ar_1$ be an argument defeating some argument in $E \cap Ar_1$. This defeat is also present in $\mathcal{A}$ and E is admissible therefore there exists an argument $B \in E$ which defends E against A. By Lemma 154 there exists an argument $B' \in Ar_1$ defending against A such that $B'^{-} \subseteq B^{-}$. By Lemma 147 (taking $S = \{B\}$) $B' \in E$. Hence $E \cap Ar_1$ defends against A. Since A has been chosen arbitrarily, $E \cap Ar_1$ defends itself. □

**Lemma 158.** *Let $\mathcal{B}$ be a knowledge base such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ where $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint. Let $\mathcal{A} = (Args, \rightharpoonup)$, $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ be the inconsistency-cleaned argumentation frameworks built from $\mathcal{B}$ and $\mathcal{B}_1$ respectively. Let $\mathcal{A}r \subseteq Args_1$ be a set of arguments. If $\mathcal{A}r$ is an admissible set in $\mathcal{A}_1$ then $\mathcal{A}r$ is an admissible set in $\mathcal{A}$.*

*Proof.* The set of arguments $\mathcal{A}r$ is conflict-free in $\mathcal{A}$, because it is conflict-free in $\mathcal{A}_1$ and there are no defeats between arguments $\mathcal{A}r$ in $\mathcal{A}$ which where not present in $\mathcal{A}_1$. It remains to show that $\mathcal{A}r$ defends itself in $\mathcal{A}$.

Take any argument $A \in \mathit{Args}$ defeating the set $\mathcal{A}r$. It is only possible if $\mathtt{Atoms}(\mathtt{Conc}(A)) \subseteq \mathtt{Atoms}(\mathcal{B}_1)$. By Lemma 154 there exists an argument $A' \in \mathit{Args}_1$ such that $\mathtt{Conc}(A') = \mathtt{Conc}(A)$ and $A'^{-} \subseteq A^{-}$. The argument $A'$ defeats $\mathcal{A}r$ in $\mathcal{A}_1$ and since $\mathcal{A}r$ is admissible in $\mathcal{A}_1$ there exists an argument $B \in \mathcal{A}r$ defeating $A'$ and also $A$ since $A'^{-} \subseteq A^{-}$. Therefore $\mathcal{A}r$ defends itself in $\mathcal{A}$. □

Now we show that inconsistency-cleaned argumentation frameworks satisfy Non-interference under complete semantics.

**Theorem 159.** *The inconsistency-cleaned version of the* ASPICLite *system satisfies Non-interference under complete semantics.*

*Proof.* Let $\mathcal{B}$ be a knowledge base such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ where $\mathcal{B}_1$ and $\mathcal{B}_2$ are syntactically disjoint. Let $\mathcal{A} = (\mathit{Args}, \rightharpoonup)$ and $\mathcal{A}_1 = (\mathit{Args}_1, \rightharpoonup_1)$ be the inconsistency-cleaned argumentation frameworks built from $\mathcal{B}$ and $\mathcal{B}_1$ respectively.

Let $\{BE_1, \dots, BE_n\}$ and $\{SE_1, \dots, SE_m\}$ be sets of complete extensions of the argumentation frameworks $\mathcal{A}$ and $\mathcal{A}_1$ respectively. We need to show that $L = R$ where:

$$L = Cn_{complete}(\mathcal{B})_{|\mathtt{Atoms}(\mathcal{B}_1)} = \{\mathtt{Concs}(BE_1)_{|\mathtt{Atoms}(\mathcal{B}_1)}, \dots, \mathtt{Concs}(BE_n)_{|\mathtt{Atoms}(\mathcal{B}_1)}\},$$
$$R = Cn_{complete}(\mathcal{B}_1)_{|\mathtt{Atoms}(\mathcal{B}_1)} = \{\mathtt{Concs}(SE_1), \dots, \mathtt{Concs}(SE_m)\}.$$

By Lemma 155 $L = \{\mathtt{Concs}(BE_1 \cap Ar_1), \dots, \mathtt{Concs}(BE_n \cap Ar_1)\}$ and by Lemma 157 $BE \cap Ar_1$ is a complete extension of $\mathcal{A}_1$ for any complete extension $BE$ of $\mathcal{A}$. We need to prove that for each complete extension $SE$ of $\mathcal{A}_1$ there exists a complete extension $BE$ of $\mathcal{A}$, such that $BE \cap Ar_1 = SE$.

Let $SE$ be a complete extension of the argumentation framework $\mathcal{A}_1$, i.e. $F_{\mathcal{A}_1}(SE) = SE$. By Lemma 158 $SE$ is an admissible set in $\mathcal{A}$. Consider the set $BE = \bigcup_{n=1}^{\infty} F^n_{\mathcal{A}}(SE)$. It is a complete extension of $\mathcal{A}$ as the least fixed point of $F_{\mathcal{A}}$ containing $SE$. It remains to show that $BE \cap Ar_1 = SE$. Indeed by Lemma 156 we have

$$BE \cap Ar_1 = \left(\bigcup_{n=1}^{\infty} F^n_{\mathcal{A}}(SE)\right) \cap Ar_1 = \bigcup_{n=1}^{\infty} F^n_{\mathcal{A}}(SE) \cap Ar_1 =$$
$$= \bigcup_{n=1}^{\infty} F^n_{\mathcal{A}_1}(SE) = \bigcup_{n=1}^{\infty} SE = SE.$$

□

An inconsistency-cleaned argumentation framework is non-trivial under complete semantics.

**Theorem 160.** *The inconsistency-cleaned version of the* ASPICLite *system satisfies non-triviality under complete semantics.*

*Proof.* Let $\mathcal{A}t$ be a non-empty set of atoms. We have to prove that there exist two inconsistency-cleaned argumentation frameworks $\mathcal{A}_1 = (Args_1, \rightharpoonup_1)$ and $\mathcal{A}_2 = (Args_2, \rightharpoonup_2)$ built from $\mathcal{B}_1 = (\mathcal{P}_1, \mathcal{D}_1)$ and $\mathcal{B}_2 = (\mathcal{P}_2, \mathcal{D}_2)$ respectively such that $\mathtt{Atoms}(\mathcal{B}_1) = \mathtt{Atoms}(\mathcal{B}_2) = \mathcal{A}t$ and $\mathcal{C}n_{\mathtt{complete}}(\mathcal{A}_1) \neq \mathcal{C}n_{\mathtt{complete}}(\mathcal{A}_2)$.

Let $\mathcal{A}t = \{a_1, \ldots, a_n\}$ $(n \geqslant 1)$. Let $\mathcal{B}_1 = (\emptyset, \{\Rightarrow a_1, \ldots, \Rightarrow a_n\})$ and $\mathcal{B}_2 = (\emptyset, \{a_1 \Rightarrow a_1, \ldots, a_n \Rightarrow a_n\})$. Then $\mathcal{C}n_{\mathtt{complete}}(\mathcal{A}_1) = \{\{a_1, \ldots, a_n\}\}$ and $\mathcal{C}n_{\mathtt{complete}}(\mathcal{A}_2) = \{\emptyset\}$. $\mathcal{C}n_{\mathtt{complete}}(\mathcal{A}_1) \neq \mathcal{C}n_{\mathtt{complete}}(\mathcal{A}_2)$. □

From the fact that for any non-trivial formalism, Non-interference implies Crash-resistance (Theorem 140), it follows that without inconsistent arguments, argumentation frameworks satisfy crash resistance under complete semantics.

**Theorem 161.** *The inconsistency-cleaned version of the* ASPICLite *system satisfies Crash-resistance under complete semantics.*

*Proof.* It follows from Theorem 159 and Theorem 160. □

## 7.5 RELATED WORK

We have proposed to solve the problem of contamination by deleting all inconsistent arguments. There are other solutions proposed, for instance, deleting self-defeating arguments discussed by Pollock (1994) and Caminada and Amgoud (2007). We show that deleting self-defeating arguments does not solve the contamination problem of the ASPICLite system by the following example.

**Example 162.** *Consider the argumentation framework $\mathcal{A}$ in Figure 25. There are two self-defeating arguments $J_2$ and $M_2$ in $\mathcal{A}$. We obtain the argumentation framework in Figure 29 after deleting the self-defeating arguments $J_2$ and $M_2$ of $\mathcal{A}$. We can see that in Figure 29, the argument $W_1$ is still not*

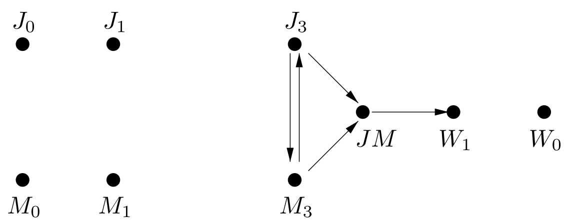



Figure 29: Deleting self-defeating arguments from the AF of Figure 25

*necessarily accepted. The weather forecast is still contaminated by completely irrelevant information.*

In general one has to be extremely careful when starting to remove arguments from an argumentation framework. To illustrate the perils, let us examine what happens when one starts deleting self-undercutting arguments.

**Example 163.** *Pollock (1994) Let* $\mathcal{D} = \{p \underset{u}{\Rightarrow} q\}$*,* $\{q \rightarrow \neg a\} \subseteq \mathcal{S}$*. Let* $a \in \mathcal{U}$ *and let* $\mathcal{S}$ *contain all propositional inferences. Then:*

$A_1 : p \quad A_2 : A_1 \Rightarrow q \quad A_3 : A_2 \rightarrow \neg a$

*(A possible interpretation:* $p$*: John says that he is unreliable and* $q$*: John is unreliable.)*

*In this argumentation framework (Figure 30),* $A_3$ *is a self-undercutting argument. If we deleted* $A_3$ *from the argumentation framework,* $A_1$ *and* $A_2$ *would be in the grounded extension and preferred extension. Then the postulate of Closure is violated because* $q \in \texttt{Concs}(E)$ *but* $\neg a \notin \texttt{Concs}(E)$*.*

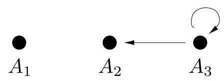



Figure 30: The argumentation framework of Example 163

What Example 163 illustrates is that if one starts deleting particular classes of arguments, then one may end up violating some of the rationality postulates (in this case: Closure). Our contribution is that we have proved that these problems do *not* occur when deleting inconsistent arguments. That is, unlike for instance deleting the class of self-undercutting arguments, deleting the class of inconsistent arguments does not cause any violations of for instance, the postulates of Closure and (Direct/Indirect) Consistency.

## 7.6 SUMMARY AND FUTURE WORK

In this chapter we have introduced the ASPICLite system, which is similar to the argumentation formalism treated by Caminada and Amgoud (2007), and we identified conditions under which it can avoid being affected by contaminating information.

ASPIC+ Prakken (2010); Modgil and Prakken (2013) is a framework for specifying argumentation systems. The specification includes an ordering on defeasible rules which is used to instantiate preferences over arguments according to one of two principles (last-link or weakest-link). The ASPICLite formalism can be seen as a system specified in ASPIC+ by setting the universal order on defeasible rules, i.e. every two rules are related, which then leads to equally preferred arguments irrespectively of the principle used. The following is an example (due to Leon van der Torre) showing that after applying the current solution to an ASPICLite framework with the last-link principle and an arbitrary ordering on defeasible rules, the postulate of Closure can be violated.

**Example 164.** *Given the knowledge base* $\mathcal{B} = (\mathcal{P}, \mathcal{D})$ *with* $\mathcal{P} = \emptyset$ *and* $\mathcal{D} = \{\Rightarrow p; p \Rightarrow q;$
$\Rightarrow \neg p \vee \neg q\}$ ASPICLite *will construct the arguments depicted in Figure 31*

*a). Further assume that* $\Rightarrow \mathsf{p}$ *has priority* 1 *(lowest),* $\Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ *has priority* 2 *(middle) and* $\mathsf{p} \Rightarrow \mathsf{q}$ *has priority* 3 *(highest).*

*The last-link principle*[5] *gives an ordering over arguments based on an ordering over defeasible rules by comparing sets of the last defeasible rules used in the construction of arguments, i.e. argument* $\mathsf{A}$ *is less or equally preferred to* $\mathsf{B}$ *if and only if there exists a last defeasible rule used in* $\mathsf{A}$ *that is less or equally preferred to any last defeasible rule used in* $\mathsf{B}$*. The set of last defeasible rules of an argument whose top rule is defeasible contains this single rule. If the top rule of an argument is strict and it has no antecedent the set of last defeasible rules is empty, otherwise it is the union of the last defeasible rules of its direct subarguments. For example, to compare argument* $\mathsf{A}_4$ *with* $\mathsf{A}_3$ *we need to consider rules* $\{\Rightarrow \mathsf{p}, \mathsf{p} \Rightarrow \mathsf{q}\}$ *and* $\{\Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}\}$ *respectively. Since* $\Rightarrow \mathsf{p}$ *is less preferred than* $\Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ *but* $\Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ *is not less or equally preferred to* $\Rightarrow \mathsf{p}$ *argument* $\mathsf{A}_4$ *is strictly less preferred than argument* $\mathsf{A}_3$*. In consequence argument* $\mathsf{A}_4$ *no longer rebuts arguments* $\mathsf{A}_3, \mathsf{A}_5, \mathsf{A}_6$ *(on* $\mathsf{A}_3$*). In the table below are listed arguments with associated last defeasible rules and preference, and the resulting framework is depicted in Figure 31 b).*

| *Argument* | | *Last Defeasible Rules* | *Preference* |
|---|---|---|---|
| $\mathsf{A}_1$: | $\Rightarrow \mathsf{p}$ | $\Rightarrow \mathsf{p}$ | (1) |
| $\mathsf{A}_2$: | $\mathsf{A}_1 \Rightarrow \mathsf{q}$ | $\mathsf{p} \Rightarrow \mathsf{q}$ | (3) |
| $\mathsf{A}_3$: | $\Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ | $\Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ | (2) |
| $\mathsf{A}_4$: | $\mathsf{A}_1, \mathsf{A}_2 \rightarrow \neg(\neg\mathsf{p} \vee \neg\mathsf{q})$ | $\Rightarrow \mathsf{p}; \mathsf{p} \Rightarrow \mathsf{q}$ | (1) |
| $\mathsf{A}_5$: | $\mathsf{A}_1, \mathsf{A}_3 \rightarrow \neg\mathsf{q}$ | $\Rightarrow \mathsf{p}; \Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ | (1) |
| $\mathsf{A}_6$: | $\mathsf{A}_2, \mathsf{A}_3 \rightarrow \neg\mathsf{p}$ | $\mathsf{p} \Rightarrow \mathsf{q}; \Rightarrow \neg\mathsf{p} \vee \neg\mathsf{q}$ | (2) |

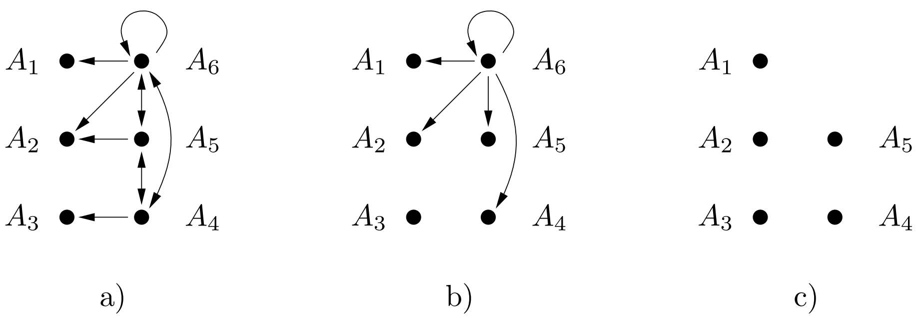



Figure 31: The argumentation frameworks illustrating problems with integration of preferences: a) ASPICLite, b) ASPICLite with Last Link, c) inconsistency-cleaned version of ASPICLite with Last Link

[5]We refer to last-link principle described in Prakken (2010) which corresponds to the last link principle with Elitist comparison in Modgil and Prakken (2013). The last-link principle considers also an ordering over premises which we omit for simplification, since we do not use premises in our example.

*The argument $A_6$ is an inconsistent argument, therefore according to the solution proposed in this chapter, we delete it. The resulting framework is depicted in Figure 31 c). There is one complete extension of arguments $\{A_1, A_2, A_3, A_4, A_5\}$[6] with conclusions that are inconsistent (arguments $A_3$ and $A_4$) and not closed under strict rules (arguments $A_2$ and $A_3$ without $A_6$).*

*Notice that it is combination of preferences with 'cleaning' that causes the problem since framework Figure 31 b) has a single extension $\{A_3\}$ which violates neither Closure nor consistency postulates.*

From Example 164 we can conclude that our solution does not work for any arbitrary reasonable argument ordering in the sense of Modgil and Prakken (2013); Prakken (2010).

In Section 7.3 we defined the postulates of Non-interference and Crash-resistance in the specific case of the ASPICLite formalism. Then we illustrated how these postulates were violated by that system. Then a solution, namely deleting inconsistent arguments, has been proposed. We showed that without inconsistent arguments, the ASPICLite system satisfies the postulates of Closure, Direct Consistency, Indirect Consistency, Non-interference and Crash-resistance under complete semantics.

Argumentation frameworks are built from knowledge bases. Some knowledge bases can be regarded as unions of syntactically disjoint knowledge bases. The consequences of each small argumentation framework should not influence other argumentation frameworks when they are logically unrelated. The union of the consequences of subframeworks should also be in the consequence of the whole argumentation framework. Inconsistent arguments can connect the graphs together and change the logical consequences of the frameworks. This leads to unrelated arguments affecting each other so that the consequences become unreasonable. The argumentation systems that have this problem violate the postulates of Non-interference and Crash-resistance. Several formalisms for argument-based entailment violate these two postulates, for instance, Pollock's OSCAR system Pollock (1995) and some instantiations of the ASPIC+ framework Modgil and Prakken (2013); Prakken (2010). The ASPICLite system also violates the postulates of Non-interference and Crash-resistance. In order to solve this problem in the ASPICLite system we delete inconsistent arguments from argumentation frameworks. Then the ASPICLite system under complete semantics satisfies the five postulates of Closure, Direct Consistency, Indirect Consistency, Non-interference and Crash-resistance. Those postulates ensure that the system does not crash and can produce logically reasonable results even when potentially contaminating information is being input into the system. The five postulates are satisfied by the inconsistency-cleaned version of the

[6] Some arguments corresponding to logical consequences of the arguments present were omitted.

ASPICLite system under complete semantics. So the troublesome behavior that occurs in formalisms like that of Prakken (2010) and that of Pollock (1994) is avoided in this particular argumentation system.

One important question is what is the price of our filtering. We can give just an informal answer. First, filtering of inconsistent arguments adds an additional step to the construction that possibly can increase complexity since verifying consistency is a hard problem. Second, inconsistent arguments represent in the framework information about the other arguments affected by inconsistency that is lost in the pruning step. Example 164 illustrates that adding preferences using the last-link principle doesn't work when inconsistent arguments are removed which suggests that this procedure relies on that information. The exact impact of the solution remains to be investigated in the future work.

A potential topic for future research would be how our solution (deleting inconsistent arguments) behaves in the context of semantics like grounded, preferred, semi-stable and ideal. Another topic would be the search for alternative solutions, for instance, forbidding strict rules feeding their consequents into the antecedents of other strict rules and disallowing the application of strict rules to an indirectly inconsistent set of formulas. Finally, another topic for investigation is to what extent is it possible to satisfy all five rationality postulates without having to delete all inconsistent arguments?

# 8 CONCLUSIONS

Below we summarise the answers found in this work to the research questions stated in the introductory chapter.

> **RQ 1:** How can we measure a distance between viewpoints represented by labellings in argumentation?

A distance is relevant to the problem of belief revision and judgement aggregation. Belief revision can be seen as finding a position consistent with the new information which is the closest to the original one. Judgement aggregation can be seen as finding a collective position which is closest to the positions of individuals in the group. There are generic distance measures like Hamming distance which can be used but they treat labellings as a vectors of labels. What makes measuring the distance between labellings challenging is to reflect the argumentation semantics. We used an axiomatic method and divided the above question into two subquestions.

> **RQ 1.1:** What are desirable properties of distance measures for labellings?

In Chapter 3, in addition to metric postulates, we proposed and analysed two compositionality postulates **(COM, COM$_{Sem}$)**, four equivalence postulates **(AUTO, ISO, LAB$_{Sem}$, IPI)** and three postulates based on betweenness and qualitative distance **(BTW, QDA, SQDA)** (each in two variants), which connect distance with argumentation framework and argumentation semantics.

The compositionality postulates state that for argumentation frameworks which can be divided into independent parts, measuring can be performed part by part and summed together. We consider two variants of independence.

The equivalence postulates state that the distance between the labellings of equivalent argumentation frameworks should be the same. Our postulates correspond to four notions of equivalence between frameworks and corresponding mappings between their labellings.

The betweenness postulate and qualitative distance postulates depend on the provided betweenness relation and partially ordered qualitative distance. We proposed two concrete betweenness relations (simple/refined) and two qualitative distances (Hamming set and Refined Hamming pair) which capture some approaches in literature, Miller and Osherson (2009); Duddy and Piggins (2012). Those postulates allow to specify in a systematic way a partial ordering over distances between pairs of labellings that should be preserved based on the way we interpret labelling.

Subsequently we addressed the following question:

**RQ 1.2:** Are those postulates jointly consistent?

In Chapter 4 we have analysed product distance measures, distance measures of a special form, which can be specified by a label distance function *diff* and a selection function $\mathfrak{S}$. We identified the properties of *diff* and $\mathfrak{S}$ on which depend satisfaction of the postulates from the previous chapter.

We have noted in Chapter 3 that the qualitative distance postulates based on Hamming set and Refined Hamming pair are incompatible. In Chapter 4 we proposed distance measures satisfying the qualitative distance postulates (separately each version) with all other postulates demonstrating they are jointly consistent.

Labels of arguments in a labellings of an argumentation framework are interdependent. The main challenge in the construction of the product distance measure is to select the set of arguments which captures all the differences and avoids redundancy. Hamming distance, which is one of product distance measures, fails **(IPI)** because it treats uniformly all arguments and therefore is sensitive to redundancy. Another generic distance, discrete metrics, fails compositionality, betweenness and strong qualitative distance postulates because it ignores too much information. We have introduced the notion of an issue which intuitively captures the arguments containing the same information. The proposed issue-based distance, which satisfies all postulates, selects exactly one argument from each issue. The full characterisation of issue-based distance is left for future work.

Our postulates (with exception of **(IPI)** whose generalisation we leave for future work) and distance measure constructions, and definition of issues in particular, depend on arbitrary set of labellings. This is important for two reasons. First, in this way our work can be combined with any of the argumentation semantics proposed. Second, it can be useful in cases when we are interested in a distance measure between a given set of labellings. The clustering of agents based on their opinions or judgement aggregation operators which select most 'central' opinion in the group are possible examples. The problem of calculating the set of labellings returned by multi-status semantics like complete usually has high computational complexity. But the problem of labelling verification is often easier. Therefore this class of problems is easier to apply. The relevant question to investigate in the future is: how issue-based distance changes when the set of known *Sem*-labellings used for issue calculation tends to the set of all *Sem*-labellings.

Whether our postulates are desirable remains an open question. Nevertheless the fact that they depend strongly on the argumentation semantics and can separate generic distance measures makes them properties worth considering. The application to judgement aggregation shows that **(IPI)** is desirable (more information follows). In the future work we would like to investigate the consequences of the other postulates in a similar way.

> **RQ 2:** How to use distance for aggregation of judgements represented by labellings?

> **RQ 2.1:** How can we 'repair' the collective outcome when it is not rational?

In Chapter 5 we used the distance methods for argumentation from Chapter 4, illustrating how they can be employed to address problems of aggregation in argumentation. We adapted the framework of Miller and Osherson from binary judgement aggregation to our setting, defining several operators for aggregating argument labellings. We illustrated informally that methods which fail **(IPI)** may lead to agenda manipulation. Finally we illustrated the generality of the resulting framework for aggregation by showing how the $\Uparrow\!\Downarrow$ aggregation method of Caminada and Pigozzi (2011) can be viewed as an instance of one of the MO methods.

> **RQ 3:** What type of dialogue can be associated with grounded semantics?

In Chapter 6, we have examined how the notion of grounded semantics can be specified in terms of persuasion dialogue. Unlike for instance the standard grounded game Prakken and Sartor (1997); Caminada (2004); Modgil and Caminada (2009) or the approach of Prakken (2005), our dialogue game does not depend on an implicit tree-like structure, in which the moves have to fit. Also, unlike the approach in for instance Parsons et al. (2002, 2003a,b), we do not merely *apply* the concept of grounded semantics (for instance for determining what moves an agent is allowed to make, depending on its acceptance attitude) but we *characterize* it.

In Chapter 7 we investigated the idea of instantiating ASPIC+ Prakken (2010); Modgil and Prakken (2013) with the set of strict rules generated by classical logic. We asked the following question:

> **RQ 4:** What are the consequences of generating strict rules by the entailment of classical logic?

We have introduced the ASPICLite system, which is similar to the argumentation formalism treated by Caminada and Amgoud (2007), and can be specified in ASPIC+. As expected after generating the strict rules with entailment of classical logic the obtained system fails consistency due to the principle of explosion. We adapted the postulates of Non-interference and Crash-resistance in the specific case of the ASPICLite formalism. Then we illustrated how these postulates were violated by that system. Then a solution, namely deleting inconsistent arguments, has been proposed. We showed that without inconsistent arguments, the ASPICLite system satisfies the postulates of Closure, Direct Consistency, Indirect Consistency, Non-interference and Crash-resistance under complete semantics. This provides the positive answer to the subquestion:

> **RQ 4.1:** Is it feasible to implement *Crash-resistance* and *Non-interference* in Argumentation-based logical formalism by removing inconsistent arguments?

Nevertheless, we found serious drawbacks in our implementation. The ability to deal with preferences is a strong side of the ASPIC+ framework. In ASPICLite this feature is lost. ASPICLite is no longer an instantiation of ASPIC+ due to the modified way in which argumentation framework is constructed. Moreover, it is not possible to keep preference handling as they are handled in ASPIC+ framework because the information contained in the inconsistent arguments removed in ASPICLite is used by ASPIC+ framework to maintain consistency.

The work started in this chapter was recently pushed forward by Grooters (2014); Grooters and Prakken (2014). In the proposed ASPIC* system they manage to combine preference handling and satisfy all the here mentioned postulates. In their approach they generate the set of strict rules with paraconsistent logic formalism - a family of logical formalisms which were studied to address the principle of explosion. It is worth noting that not all paraconsistent formalisms can be used and again the ASPIC* customise the construction of argumentation framework. This may suggest that Dung's style argumentation as a methodology to develop non-monotonic logic is difficult.

*Future work*

Concrete ideas about the future work have been pointed at the end of each chapter. Here we would like just to give the general line of research that we see as being important to this field.

First we would like to investigate the general notion of *position*. One of the points of real argumentation is to exchange and build positions. Therefore we need a system that will enable to express such position, search between expressed positions and monitor changes. This is a concern for the growing field of *web and social argumentation*, Leite and Martins (2011).

An agent is in an ongoing process of reasoning. The position can pre-formally be thought of as a fixed point in this process which determines agent's actions, opinions etc. This position can be elicited by presenting arguments to the agent which the agent assigns labels to. But this labelling is only an approximation of the position and cannot capture the whole knowledge or beliefs of an agent. The first idea may be to define a position as a function pos which assigns a labelling to every possible framework $\mathcal{A} \mapsto \mathrm{pos}(\mathcal{A}) \in \mathit{Labs}(\mathcal{A})$.

The future work is to formally define a notion of position and develop new kind of argumentation theory in which this notion of position plays a central role. The point is not to define what a rational position is but to define a communication tool so that agents can express their position with the requested precision. Thus semantics does not describe agents but the communication language.

Comparing to Dung, we propose to keep the formalism on an abstract level. In contrast to Dung we propose to bring argument construction to the picture through the notion of position and deliberately abandon the control over the way arguments are created. We see it as an advantage. First, there are cases where we simply have no control over arguments, e.g. text arguments created by humans. Second, in case where arguments are constructed from the knowledge base the issue of control simply shifts from argument construction into knowledge base construction. Third, it is questionable if knowledge can be expressed as a symbolic database. The advancements in non-symbolic artificial intelligence suggest the contrary. The challenge here is to define position through the interaction with an argumentation framework which need to be extended for that purpose.

# BIBLIOGRAPHY


ASPIC-consortium (2005). Deliverable D2.5: Draft formal semantics for ASPIC system.

Baigent, N. (1987). Preference proximity and anonymous social choice. *The Quarterly Journal of Economics*, 102(1):161–169.

Baroni, P., Caminada, M., and Giacomin, M. (2011). An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26:365–410.

Baroni, P. and Giacomin, M. (2007). On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700.

Besnard, P. and Hunter, A. (2001). A logic-based theory of deductive arguments. *Artif. Intell.*, 128(1-2):203–235.

Bex, F., Lawrence, J., Snaith, M., and Reed, C. (2013). Implementing the argument web. *Communications of the ACM*, 56(10):66–73.

Booth, R., Awad, E., and Rahwan, I. (2014). Interval methods for judgment aggregation in abstract argumentation. In *Proc. KR 2014*, pages 594–597.

Booth, R., Caminada, M., Podlaszewski, M., and Rahwan, I. (2012). Quantifying disagreement in argument-based reasoning. In *Proc. AAMAS 2012*, pages 493–500.

Booth, R. and Podlaszewski, M. (2014). Using distances for aggregation in abstract argumentation. In *Proceedings of the 26th Benelux Conference on Artificial Intelligence (BNAIC2014)*, pages 17–24.

Caminada, M. (2004). *For the sake of the Argument. Explorations into argument-based reasoning*. Doctoral dissertation Free University Amsterdam.

Caminada, M. (2005). Contamination in formal argumentation systems. In *Proceedings of the 17th Benelux Conference on Artificial Intelligence (BNAIC)*, pages 59–65.

Caminada, M. (2006a). On the issue of reinstatement in argumentation. *Lecture Notes in Computer Science*, 4160:111.

Caminada, M. (2006b). Semi-stable semantics. In *Proceedings of the 1st Conference on Computational Model of Arguments (COMMA)*, pages 121–130.

Caminada, M. (2010). Preferred semantics as socratic discussion. In Gerevini, A. E. and Saetti, A., editors, *Proceedings of the eleventh AI*IA symposium on artificial intelligence*, pages 209–216.

Caminada, M. (2014). Strong admissibility revisited. *COMMA*, pages 197–208.

Caminada, M. and Amgoud, L. (2005). An axiomatic account of formal argumentation. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI)*, pages 608–613.

Caminada, M. and Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310.

Caminada, M. and Gabbay, D. (2009). A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145.

Caminada, M. and Pigozzi, G. (2011). On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102.

Caminada, M., Pigozzi, G., and Podlaszewski, M. (2011). Manipulation in group argument evaluation. In *IJCAI-11*.

Caminada, M. and Podlaszewski, M. (2012a). Grounded semantics as persuasion dialogue. *COMMA*, 245:478–485.

Caminada, M. and Podlaszewski, M. (2012b). User-computer persuasion dialogue for grounded semantics. In *Proceedings of the 24th Benelux Conference on Artificial Intelligence (BNAIC2012)*, pages 343–344.

Caminada, M. W. A., Carnielli, W. A., and Dunne, P. E. (2012). Semi-stable semantics. *Journal of Logic and Computation*, 22(5):1207–1254.

Cayrol, C., de Saint-Cyr, F. D., and Lagasquie-Schiex, M. (2010). Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84.

Coste-Marquis, S., Devred, C., Konieczny, S., Lagasquie-Schiex, M., and Marquis, P. (2007). On the merging of dung's argumentation systems. *Artificial Intelligence*, 171(10-15):730–753.

Delgrande, J. (2004). Preliminary considerations on the modelling of belief change operators by metric spaces. In *Proc. NMR*, pages 118–125.

Deza, M. and Deza, E. (2009). *Encyclopedia of Distances*. Encyclopedia of Distances. Springer.

Duddy, C. and Piggins, A. (2012). A measure of distance between judgment sets. *Social Choice and Welfare*, 39(4):855–867.

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

Dung, P. M., Mancarella, P., and Toni, F. (2007). Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674.

Endriss, U. (2014). Social choice theory as a foundation for multiagent systems. In *Multiagent System Technologies*, pages 1–6. Springer.

Gabbay, D. (2009). Fibring argumentation frames. *Studia Logica*, 93(2-3):231–295.

Gelfond, M. and Lifschitz, V. (1988). The stable model semantics for logic programming. In *Proceedings of the 5th International Conference on Logic Programming/Symposium on Logic Programming (ICLP/SLP)*, pages 1070–1080.

Gelfond, M. and Lifschitz, V. (1991). Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–386.

Gorogiannis, N. and Hunter, A. (2011). Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, 175(9-10):1479–1497.

Governatori, G., Maher, M., Antoniou, G., and Billington, D. (2004). Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14(5):675–702.

Grandi, U. and Pigozzi, G. (2012). On compatible multi-issue group decisions. *Proc. LOFT-2012*.

Grooters, D. (2014). Paraconsistent logics in argumentation systems. Master's thesis, Utrecht University.

Grooters, D. and Prakken, H. (2014). Combining paraconsistent logic with argumentation. In *Computational Models of Argument. Proceedings of COMMA 2014*, volume 266, pages 301–312. IOS Press.

Hamblin, C. L. (1970). *Fallacies*. Methuen, London, UK.

Hamblin, C. L. (1971). Mathematical models of dialogue. *Theoria*, 37:130–155.

Koons, R. (2014). Defeasible reasoning. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition.

Lehmann, D., Magidor, M., and Schlechta, K. (2001). Distance semantics for belief revision. *Journal of Symbolic Logic*, pages 295–317.

Leite, J. and Martins, J. (2011). Social abstract argumentation. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three*, pages 2287–2292. AAAI Press.

Liao, B., Jin, L., and Koons, R. C. (2011). Dynamics of argumentation systems: A division-based method. *Artificial Intelligence*, 175(11):1790–1814.

Lin, F. and Shoham, Y. (1989). Argument systems: A uniform basis for nonmonotonic reasoning. In *Proceedings of the 1st International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 245–255.

List, C. and Puppe, C. (2009). Judgment aggregation: A survey. In *Handbook of Rational and Social Choice*. Oxford University Press.

Mackenzie, J. D. (1979). Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133.

Mackenzie, J. D. (1990). Four dialogue systems. *Studia Logica*, 51:567–583.

Miller, M. and Osherson, D. (2009). Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601.

Modgil, S. and Caminada, M. (2009). Proof theories and algorithms for abstract argumentation frameworks. In Rahwan, I. and Simari, G., editors, *Argumentation in Artificial Intelligence*, pages 105–129. Springer.

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

Modgil, S., Toni, F., Bex, F., Bratko, I., Chesñevar, C. I., Dvořák, W., Falappa, M. A., Fan, X., Gaggl, S. A., García, A. J., et al. (2013). The added value of argumentation. In *Agreement Technologies*, pages 357–403. Springer.

Parsons, S., Wooldridge, M., and Amgoud, L. (2002). An analysis of formal interagent dialogues. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems (AAMAS-02)*, pages 394–401.

Parsons, S., Wooldridge, M., and Amgoud, L. (2003a). On the outcomes of formal inter-agent dialogues. In Rosenschein, J. S., Sandholm, T., Wooldridge, M., and Yokoo, M., editors, *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2003), Melbourne, Australia*, New York City, NY, USA. ACM Press.

Parsons, S., Wooldridge, M., and Amgoud, L. (2003b). Properties and complexity of formal inter-agent dialogues. *Journal of Logic and Computation*, 13(3):347–376.

Peppas, P., Chopra, S., and Foo, N. (2004). Distance semantics for relevance-sensitive belief revision. *Proc. KR*, 4:319–328.

Pigozzi, G. (2006). Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2):285–298.

Pollock, J. (1987). Defeasible reasoning. *Cognitive Science*, 11(4):481–518.

Pollock, J. (1992). How to reason defeasibly. *Artificial Intelligence*, 57(1):1–42.

Pollock, J. (1994). Justification and defeat. *Artif. Intell.*, 67(2):377–407.

Pollock, J. (1995). *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA.

Prakken, H. (2005). Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15(6):1009–1040.

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

Prakken, H. and Sartor, G. (1997). Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75.

Rahwan, I., Larson, K., and Tohmé, F. (2009). A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*.

Rahwan, I. and Tohmé, F. (2010). Collective argument evaluation as judgement aggregation. In *Proc. AAMAS*, pages 417–424.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132.

S. Pan, K. L. and Rahwan, I. (2010). Argumentation mechanism design for preferred semantics. In *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA)*, pages 403–414.

Simari, G. R. and Loui, R. P. (1992). A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2-3):125–157.

Strasser, C. and Antonelli, G. A. (2014). Non-monotonic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2014 edition.

Thang, P. M., Dung, P. M., and Hung, N. D. (2009). Towards a common framework for dialectical proof procedures in abstract argumentation. *Journal of Logic and Computation*, 19(6):1071–1109.

Tintarev, N., Kutlak, R., Oren, N., Van Deemter, K., Green, M., Masthoff, J., and Vasconcelos, W. (2013). Sassy—scrutable autonomous systems. *Proceedings of The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, 2013.

Verheij, B. (1996). Two approaches to dialectical argumentation: admissible sets and argumentation stages. In Meyer, J.-J. and van der Gaag, L., editors, *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC'96)*, pages 357–368. Utrecht University.

Vreeswijk, G. (1993). *Studies in Defeasible Argumentation*. PhD thesis, Dept. of Mathematics and Computer Science,Vrije Universiteit Amsterdam.

Vreeswijk, G. (1997). Abstract argumentation systems. *Artificial Intelligence*, 90:225–279.

Walton, D. N. and Krabbe, E. C. W. (1995). *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Series in Logic and Language. State University of New York Press, Albany, NY, USA.

Wu, Y., Caminada, M., and Gabbay, D. (2009). Complete extensions in argumentation coincide with 3-valued stable models in logic programming. *Studia Logica*, 93(1-2):383–403. Special issue: new ideas in argumentation theory.

Wu, Y. and Podlaszewski, M. (2014). Implementing crash-resistance and non-interference in logic-based argumentation. *Journal of Logic and Computation*.

Yuan, T., Moore, D., and Grierson, A. (2003). Computational agents as a test-bed to study philosophical model "DE", a development of mackenzie's "DC". *Journal of Informal Logic*, 23(3):263–284.