# Abstract Dialectical Frameworks

Keshavarzi Zafarghandi, Atefeh

*DOI:*
10.33612/diss.211084817

# Abstract Dialectical Frameworks

Semantics, Discussion Games, and Variations

Atefeh Keshavarzi Zafarghandi

This research was funded by the Center of Data Science & Systems Complexity (DSSC) Doctoral Programme, at the University of Groningen.

university of groningen

# Abstract Dialectical Frameworks

Semantics, Discussion Games, and Variations

**PhD thesis**

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Tuesday 19 April 2022 at 11:00 hours

by

**Atefeh Keshavarzi Zafarghandi**

born on 8 March 1982
in Tehran, Iran

## Supervisors
Prof. H.B. Verheij
Prof. L.C. Verbrugge

## Assessment Committee
Prof. P. Baroni
Prof. G. Brewka
Prof. H.H. Hansen

# Acknowledgements

The journey towards a PhD has been a challenging and rewarding experience supported by many people. I have significantly progressed in this program compared to where I initially began four years ago and it would be impossible without the great relationship that I have had with my supervisors.

I would like to express my deepest gratitude to both my supervisors Bart Verheij and Rineke Verbrugge. Their knowledge and experience led to the significant improvement of my dissertation. They have always been motivating for pursuing my ideas and even when my drafts needed an enormous change, my supervisors had a positive point of view about my achievements. Further, I have always been appreciative for the freedom with which they guided me and for being open to all sorts of different approaches that did not always fit into our initial plans. I also want to thank them for their full support concerning my life over these years. Their kindness and advice helped me not to feel pressured or stressed during the project, especially in the tough times of the corona pandemic. I had the chance to freely share my feelings with them whenever I felt less motivated. In such conditions, by reminding of the priority of each person's life and health, they significantly helped me to go back to the main route of my work.

I would like to extend my sincere thanks to professors Gerhard Brewka, Pietro Baroni, and Helle Hansen, who kindly accepted to participate in the reading committee of this thesis. Their valuable comments led to the better development of the results. I am especially grateful to Helle Hansen for all her attention to this dissertation. I believe that it is almost impossible that even a tiny mistake would go unnoticed under her radar. I also like to thank professor Stefan Woltran, my master's degree supervisor, who recommended me to apply for this PhD program. He never stops supporting me with his valuable advice.

The financial support for this research was provided by the Center of

Data Science & Systems Complexity (DSSC) Doctoral Programme, at the University of Groningen which provides an excellent opportunity in a very friendly research environment.

I have to thank all my colleagues, who have supported me: Martin Diller, Yuri David Santos, Stipe Pandžić, Hamidreza Kasaei, Maaike Los, Hamed Ayoobi, Edoardo Baccini, Mahya Ameryan, Haibin Wen, Heng Zheng, Asmaa Haja, Cor Steging, and Koorosh Shomalzadeh. Thank you for creating a motivating working environment and for your daily encouragement and productive discussions. I wish to also kindly thank my paranymphs Koorosh and Yuri. I am especially indebted to Yuri for the many times that he commented on my work, especially the summary, as well as all his advice for finding a job. Thanks to Martin and Stipe who helped me to improve the language level of the last part of this dissertation, i.e., the discussion and conclusion chapter. I wish to thank Maaike for helping me in translating the summary into Dutch. During my time at the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, my office mates Hamed Ayoobi, Haibin Wen, and Heng Zheng made my life enjoyable and stress-free. Thank you for the cherished time that we spent together.

Although academic support is of great significance, the support of friends and family is equally important. Thanks to all my friends for their hospitality, kindness and all supports: Maryam Akbari, Ali Ghasab, Zahra Farahani, Fatemeh Ghorbani, Hamidreza Kasaei, Fatemeh Amirbeiki, Puran Seifi, Roza Mehrabi, Beyda Zerehband, Mahsa Zibaei, Ahlam Sadih, Reza Heydari, and all those that I might have forgotten. Additionally, I am extending my heartful thanks to Fatemeh Ghorbani and Hamidreza Kasaei for their enormous support during my PhD, for their advice to choose a fitting job and their dinner invitations. Thank you Roza, you are not only a good consultant but also a perfect fellow traveler. I am grateful to Fatemeh Amirbeiki, you are a perfect home-mate and we had great movie nights together. Special thanks to Maryam for all your support, kindness and your surprises. You are always available in hard situations of life with your kind heart and perfect advice. I would like to offer my special thanks to Ali Ghasab, not only for his idea about the cover of this dissertation and all his support in designing and providing the cover, but also for twenty years of friendship and all his support in different aspects of life.

This thesis would not exist without the great support from my family. I would like to show my gratitude to them for supporting me by assuming my

responsibilities and providing me with an opportunity to travel: especially my father Noori, my brothers Mohammad Reza, Hamid Reza and Ali, and Maryam, my aunt whose presence has helped us to cope with life after the departure of my mother. Additionally, I am extending my appreciation to Ali for his encouragement and his unwavering support throughout my studies. It was due to my family members' selfless support that I have been able to pursue my education abroad. Most importantly, I would like to commemorate my mother, Akram, who taught me to be perseverant, patient and hopeful in life. May her soul rest in peace.

From the bottom of my heart, I thank my Creator who gave me the blessing of life. Indeed, You are almighty over everything.

Dedicated to Noori, Maryam, Mohammad Reza, Hamid Reza, Ali, and the soul of my sweet mom, Akram.

# Contents

# Part I

# Introduction and Background

# Chapter 1

# Introduction

Argumentation is a human manner of making a decision, individually or collectively, even if the available information is incomplete or is inconsistent. Argumentation has recently received increased attention in artificial intelligence (AI for short), specifically, in the sub-field of artificial (computational) argumentation. AI supports the field of argumentation, with its formal and computational systems, for extracting arguments from a given knowledge representation to evaluating arguments and making a conclusion.

This work mainly concerns abstract dialectical frameworks, a systematic and flexible argumentation formalism. The formalism is explained in Section 2.5. In this section we deal with the motivations of studying argumentation theory and formal argumentation. That is, in Section 1.1 we are mainly concerned to responding to the query: 'Why is argumentation theory a significant and valuable topic to study?' Thus, first we clarify the importance of argumentation theory. To this end, in Section 1.1.1, we explain briefly the relevance of argumentation theory in our daily life. Then, in Section 1.1.2 we present a short history of argumentation, from Aristotle's time to the present. In addition, we shortly explain the historical trends of informal and formal argumentation.

Following the seminal work by Pollock, who has introduced the notion of defeasible logic, formal argumentation has received increasing attention in artificial intelligence. In Section 1.2 we present formal argumentation, generally. Then, to show that a discussion can be modeled via formal argumentation frameworks we present an example in Section 1.2.1. Then, in Section 1.2.2 we discuss the importance of one of the most prominent approaches in formal argumentation, namely abstract argumentation frame-

works (AFs for short) introduced by Dung. Dung's formalism focuses on attack among arguments, and investigates the evaluation of the acceptability of conflicting arguments. In the remainder of Section 1.2.2 we explain how the discussion example given in Section 1.2.1 can be modeled and evaluated via AFs. Despite the popularity of AFs for representing argumentation contexts and evaluating of the status of arguments therein, these frameworks are limited to model direct attack between arguments. Thus, various attempts have been done by researchers to generalize AFs to cover additional relevant relationships among arguments. In this work we are concerned with abstract dialectical frameworks (ADFs for short) as an expressive generalization of AFs. The importance and capabilities of ADFs have been presented in Section 1.2.3.

The remainder of the introduction is organised as follows. In Section 1.3, we present the main contributions of this thesis comprehensively, while also summarising the content of each chapter. In section 1.4, we present the list of our publications and we indicate which chapter contains the corresponding contribution.

## 1.1 Argumentation Theory

A motivating question for this section is; 'Why is argumentation theory a significant and valuable topic to study?' To clarify the role of argumentation in Section 1.1.1 we explain that argumentation plays a crucial role in our life. Then, in Section 1.1.2 we briefly present the historical importance of argumentation from Aristotle time to the present, and we discuss the future of argumentation in artificial intelligence. In Section 1.1.3 we investigate differences between classical logic and argumentation, to clarify the importance and role of argumentation in automated reasoning.

### 1.1.1 Argumentation is Everywhere

Arguing is so natural for all of us that we do it all the time either with ourselves or with other people. Argumentation is an essential part of our daily life both in our individual and social activities. Argumentation can be in the form of monologue, in our mind, by evaluating arguments and counterarguments, or it can be in the form of dialogue by entering a discussion or a debate in which arguments are exchanged between agents (van Eemeren et al., 2014). We argue with ourselves for and against a subject when we want to make a decision. Agents may engage in the exchange of arguments for a variety of purposes with several dialogue types having

been identified in the literature, such as inquiry, negotiation, information seeking, deliberation, and persuasion (Walton and Krabbe, 1995). For instance, we do argumentation when we discuss with our supervisors about a project, convince a job committee that we are the perfect person to hire, talk to our parents about where to go for holiday, or try to persuade colleagues why the topic of our study is an important, interesting and crucial subject to study.

In a complex task usually humans can handle incomplete and inconsistent information via argumentation for reasoning and making a decision. Human logical reasoning can be thought of as a process of argumentative discourse which consists of one or more parties supporting their ideas or opinions. Argumentative discourse offers agents a method to examine the discussed ideas and possibly make a choice among them. A good discussion may lead to a justifiable decision. At the moment of decision making, we are usually not certain of what is the consequence of our decisions, but we may know the set of possible consequences that our decision can lead to. That is, we usually make decisions based on argumentation under uncertainty. A rational decision maker prefers to make a decision with the least regret or the most satisfaction. This can be achieved via argumentation. A decision that can be justified by the participants in a discussion, might be taken after a process of argumentation and reasoning together. A decision may be improved by argumentative discussion among participants rather than making a decision via monologue. A right decision, may lead to a cure for a disease, to an investment in a project by a business person, to a judgment in a crime case, and to a fair debate.

However, at the moment of decision making we are faced with incomplete and uncertain information, that is, with new knowledge we may change our decision. This shows the non-monotonic nature of human reasoning, whereas reasoning in classical logic is monotonic, i.e., retraction of conclusions after adding premises is not possible.

Early non-monotonic logics were proposed in the Artificial Intelligence journal (Bobrow, 1980). One of the strengths of the argumentation approach is that it turns out to be powerful enough to model a wide range of formalisms for non-monotonic reasoning. A significant paper addressing logic, non-monotonic reasoning and logic programming is proposed by Dung (1995). Another key example that connects non-monotonic reasoning and argumentation is (Pollock, 1987). Furthermore, (García and Simari, 2004) contains the notions of logic programming, argumentation and non-monotonic reasoning, namely, defeasible logic programming (DeLP).

Formal argumentation can present the non-monotonic notion of logical consequence in the form of argument construction, argument relations and argument evaluation with the aim of resolving conflicts among arguments.

In the field of formal argumentation theory different argumentation frameworks have been proposed for modeling and evaluating arguments. Models of argumentation reflect how arguments relate to one another, and semantics of models of argumentation reflect how to use argumentation for making a decision under inconsistent, controversial and incomplete information (Bench-Capon and Dunne, 2007).

The interest in carrying out investigations connecting artificial intelligence and formal argumentation has been motivated by several reasons. In the following we present a list of the main reasons presented in this section.

1. Argumentation is everywhere.
2. Furthermore, argumentation has a crucial role in all aspects of life to make a good decision like in: legal reasoning, philosophy, psychology and politics.
3. Argumentation has vital connections to other fields of AI, in particular, knowledge representation, non-monotonic reasoning and game theory.
4. In the field of argumentation we are often faced with complex argumentation, i.e., argumentation with a big number of arguments and relations among arguments.
5. To address realistic decision problems, systematic reasoning methods for making a good decision are mandatory.
6. The existence of a wide variety of argumentation styles in real life leads to a variety of formal models of argumentation and semantics for evaluating arguments (Baroni et al., 2018b; van Eemeren et al., 2014).

**How can AI be used in argumentation?** If we consider the list of items that are presented as motivations for researchers in the AI domain to consider argumentation, it would be good to think about the relation between AI and argumentation and investigate how AI can be used in argumentation. AI supports different aspects of human reasoning. First, argument mining techniques can be used to extract arguments (Cabrio and Villata, 2018; Budzynska et al., 2014; Lippi and Torroni, 2016; Lawrence and Reed, 2020). Then, the structure of the arguments are used to explore

the relation among arguments (Prakken, 2010; Pollock, 1987). In the next step, based on the relations among arguments, sets of jointly acceptable arguments can be identified in an automated way. Then, a conclusion can be drawn via this set of arguments. Argumentation pervades artificial intelligence as a simulation of human reasoning. These encourages the development of computational models of argumentation with the aim of automated reasoning.

### 1.1.2 History of Argumentation Theory

Argumentation is deeply rooted in human history, and the academic study of argumentation goes back to the ancient Greece in theoretical philosophy. Reasoning via argumentation has been a specific topic in philosophy since the time of Aristotle. The extensive work on argumentation from Aristotle to today's computational argumentation in artificial intelligence shows how far research in argumentation has come (van Eemeren and Verheij, 2017; van Eemeren et al., 2014).

According to Leibniz, "the only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate [calculemus], without further ado, to see who is right" (Leibniz, 1685). Put differently, developing automated methods capturing the human ability of reasoning is an old, ambitious, and ongoing research goal.

Big dreams bring extraordinary results. According to Leibniz's point of view, human reasoning follows determinate axioms of logic, and conclusions are based on how the mind operates, implicitly following algorithmic procedures. In modern terms, one would rephrase his dream as the aim to design a formal system and a decision procedure for making a decision without any doubt. One could say that Leibniz was thinking about a machine that can do the following tasks: 1. arguing as a human, and 2. reasoning automatically and finding a correct conclusion, in the presence of conflicts among arguments. As he said he was looking for a method to investigate who is right in a dispute, thus, one may conclude that reasoning in debates and discussions has always been a central topic of automated reasoning in the legal domain. The study of argumentation and its role in human reasoning lies in the intersection of philosophy, logic and legal reasoning (Rissland et al., 2003).

However, realizing Leibniz's dream has proven to be a formidable task. Thanks to pioneering work in logic and the theory of computation, and

especially to the fundamental works of Kurt Gödel and Alan Turing, we now understand better what computers cannot do. Due to the complex and varied structure of argumentation, the attempt to develop a universal automated system to model and evaluate argumentation has as yet failed. The developments in computer science and artificial intelligence opened up the door to some of the most fascinating developments and ideas of the past centuries in automated reasoning. Currently researchers of the domain of formal argumentation are not as optimistic as Leibniz looking for a universal formal model of decision procedure. Instead, they are eager to present different formalisms for argumentation, each of which is tuned to the modeling of an aspect of the non-monotonic characterisation of argumentation. Over the last two decades, argumentation has become a fertile research area in artificial intelligence (AI for short) (Bench-Capon and Dunne, 2007). Formalisms of argumentation are used to model and evaluate argumentation, and AI tools are used for testing.

The landscape of studying argumentation in philosophy, AI, linguistics and elsewhere is wide. In the following we briefly present the historical trend of formal and informal argumentation in the late 20th century (see (Prakken, 2017; van Eemeren et al., 2014) for an overview).

Among the researchers in informal argumentation, we concentrate on Toulmin and his influential book *'The Uses of Argument'* (Toulmin, 2003) in which he presented the limitations of using classical logic for modeling human reasoning. He believed that deductive logic is not sufficient for the understanding of human reasoning and argumentation. According to Toulmin, deductive reasoning cannot cover all aspects of human argumentation, for instance, because of counterarguments, i.e., inconsistent information in argumentation.

Toulmin's work can be assumed as an early example of informal logic and argumentation research. He proposed his model of arguing based on discussion in the court room. Applications of his schemes are presented in *Introduction to Reasoning* (1979) (Toulmin et al., 1984). Toulmin's works can be considered as first steps towards the collecting of schemes of argumentation by Walton (2008). However, the work of Toulmin (2003) and the relevant works on argumentation schemes was rarely related to computational arguments until around 2000 (Reed and Norman, 2004; Verheij, 2009; Modgil and Caminada, 2009). The classification of arguments based on argument schemes by Walton has been investigated in computational argumentation (Verheij, 2003a; Bex et al., 2013).

Early systems for argumentation-based inference preceded the heyday

non-monotonic logic in the 1980s and 1990s. For instance, Lorenzen and Lorenz developed formal dialogue systems for argumentation by using a game formulation of disputes among agents in argumentation (Lorenzen and Lorenz, 1978). Such early work on dialogue logic reformulates existing monotonic notions of logical consequence. Non-monotonic logic had become fashionable around 1980. It is a power of non-monotonic logic that it helps finding a conclusion in reasoning with inconsistent and incomplete information. Thus, the idea arose in the field of argumentation that non-monotonic inference rules can be used to model argumentation. Current research in fields of non-monotonic logic, belief revision and computational argument shows that many features of non-mathematical reasoning can be formalised. The first *International Conference on Formal and Applied Practical Reasoning* (FAPR) in 1996, addresses the interdisciplinary area of practical reasoning in artificial intelligence. Moreover, various interdisciplinary collaborations, specifically, between formal and informal argumentation, have been reported in (Reed and Norman, 2004). Currently, the *International Conference on Computational Models of Argument* (COMMA) has been a regular forum for the exchange of the results computational argumentation, since 2006. Furthermore, an open access interdisciplinary journal of *Argument & Computation* (A&C for short) has provided a dedicated venue for papers in the field of computational argumentation.

Among the many works that present systems for formal argumentation, we now focus on the work of Pollock who can be thought of as a father of argumentation and AI. Pollock introduced the notable notion of defeasible reasoning in his work (Pollock, 1987). While Toulmin criticized that deductive inference rules did not fit well with the nature of argumentation and reasoning, Pollock proposed the philosophical notion of defeasible reasoning that better fits with the character of human argumentation and reasoning. By proposing the notion of defeasible reasoning Pollock rejected the point of view that all arguments in formal argumentation have to be deductively valid. In his research, Pollock assumes that reasoning outside of mathematics involves defeasible steps (Pollock, 1995, p. 41). Many ideas that are presented by Pollock are still important aspects of formal argumentation. For instance, Pollock considered the strength of arguments in his work and the notion of argument acceptability. He also distinguished kinds of defeat, in particular undercutting and rebutting defeat. He connected to AI by developing OSCAR, a software project which is an implementation of Pollock's idea on defeasible reasoning (Pollock, 1987).

Then in 1995 Dung presented his influential formalism of *abstract argumentation frameworks* in which argumentation is formalized based on arguments and attacks between them (Dung, 1995). A generalization of Dung's formalism is the basis of our work. In Section 1.2.2, we briefly explain why Dung's argumentation framework is a significant framework in AI and argumentation theory, and in Section 2.3, we discuss the formal definitions of Dung's framework in detail.

Recently formal and computational argumentation methods have been applied in a number of applications like in law (Prakken and Sartor, 2015), medicine (Hunter and Williams, 2012; Fox and Das, 2000), health promotion (Grasso et al., 2000), debating (Slonim et al., 2021), and dispute mediation (Janier et al., 2016) (see (Atkinson et al., 2017) for a survey). Moreover, several combinations of argumentation and machine learning have been studied (Cocarascu and Toni, 2016; Ayoobi et al., 2019). The developments of techniques of AI in argumentation theory have led to the design of machines in real-world situations. For instance, recently the autonomous debating system Project Debater has been developed that can perform a debate with a human expert debater (Slonim et al., 2021). This achievement was so notable and unique that it has been published in the top ranked science journal *Nature* on March 18, 2021.

Currently, having automated argumentation systems that can help people to make better choices is the goal of a field of human-machine interaction in AI. For instance, an *automated persuasion system* is a system for persuading agents to do (or not to do) an action via a persuasion dialogue (Potyka et al., 2019; Hunter, 2015, 2018; Hadoux and Hunter, 2018, 2019; Chalaguine and Hunter, 2020). Computational persuasion systems can for instance have the aim to convince people to change a habit. Having argumentation systems with the capability of formulating and evaluating complex human argumentation that leads to high-level human-machine interactions is a goal of argumentation theory and AI.

### 1.1.3 Distinction Between Logic and Argumentation

Automated reasoning is concerned with applying reasoning in the form of logic. On the other hand, formal argumentation is concerned with automated argumentation to evaluate arguments. As mentioned in Section 1.1.1, researchers of artificial intelligence-related areas are promoted in studying of formal argumentation. For instance, Dung's landmark paper in formal argumentation has been cited more that 4000 times, based on google scholar. In this section we clarify the distinction between logic and

formal argumentation in automated reasoning.

The development of formal logic played a significant role in automated reasoning (Davis et al., 1983; Davis, 2000), that led to artificial intelligence. Several computational models of argumentation have been proposed for automated reasoning. A prominent example of a software model for argumentative reasoning is the OSCAR system (Pollock, 1987), developed in 1987. A natural question arises: "How does argumentation in AI differ from logic?" A key difference between classical logic and argumentation is the monotonic nature of classical logic and the non-monotonic nature of argumentation/human reasoning. We often make a decision under inconsistent and incomplete information, where deduction of classical logic is not a very useful reasoning model, as Toulmin argued in (Toulmin, 2003) and we presented it in Section 1.1.2.

**What is the meaning of an inconsistent piece of information**? Our argumentation contains arguments and counterarguments, thus a piece of information for a reasoning may be inconsistent. However, a set of sentences has a model in classical logic if it is a consistent set. In other words, when a set of sentences is inconsistent, then anything is deductively implied in classical logic. Although the mathematical proof of an argument may not be possible, we would like to know whether an argument is reasonable or persuasive via argumentation. The idea of argumentation is that whether or not an agent believes an argument depends on whether or not this argument can be defended against the counterarguments. In human reasoning and argumentation, in order to draw a conclusion in an inconsistent piece of information we focus on a consistent subset of a given information.

**What is the meaning of incomplete information**? In our daily life information in any reasoning task is incomplete, since always there exists a new piece of information that can be added. That is, argumentation involves incomplete information. We argue and make a decision under incomplete information and an argumentation may not be convincing anymore in the light of new information. However, in classical logic when information is incomplete, then nothing is derivable deductively.

All in all, since our reasoning involves uncertain and inconsistent information, new information may cause a change in the conclusions drawn. This reflects the fact that argumentation is a non-monotonic process. In contrast, in classical logic one proves statements. If the proof exists, then a queried statement is not refutable. That is, if the correctness of a statement is proven, it remains correct, even in presence of new information. Thus,

classical logic is a monotonic form of reasoning.

Furthermore, a formal logical proof is a proof in which every logical inference has been checked back to the fundamental axioms of logic, while the main aim of argumentation is reasoning based on arguments, either in single-agent systems or multi-agent systems. Argumentation can be used in reasoning for persuasion, deliberation, dispute, and discussion with or without formal proof. In general, argumentation and argumentative dialectic reasoning is closer to human reasoning than strict classical logic and deductive inference rules of logic. Thus, human reasoning can be automated through argumentation in which non-monotonic inference can be modelled as the competition between arguments. Indeed in argumentation one has to argue why some conclusions can/should be considered and others not. However, recently in (Besnard et al., 2020) the connections between a formalism of argumentation which is called abstract argumentation frameworks and logic have been considered. This paper is a survey for investigating where logic has been used to capture different aspects of abstract argumentation frameworks.

## 1.2 Formal Argumentation

There exist various formalisms for the modeling of argumentation, evaluating of arguments, and drawing of conclusions. First in Section 1.2.1, we present an example of discussion and we show how this discussion can be represented formally. In Section 1.2.2, we informally introduce abstract argumentation frameworks (AFs for short), presented in Dung's influential paper (Dung, 1995). Then, in Section 1.2.3 we discuss abstract dialectical frameworks (ADFs for short), an expressive generalization of AFs, first introduced in (Brewka and Woltran, 2010), and further refined in (Brewka et al., 2013, 2017a, 2018a).

### 1.2.1 A Discussion Example

We start this section with a simple discussion example to present the intuition behind formal argumentation,

**Example 1.1** *Ali and Maryam came to the Netherlands to do their PhD in the beginning of March. They are looking for a proper health insurance with a good coverage and a good price. After a research about the terms and conditions of different insurances they are going to share their points of view with each other.*

*Ali says, "I think Menzis is the best insurance for us to buy, since it has a good coverage and it is the cheapest insurance among other insurances with the same coverage per month."*
*Maryam answers: "But I do not agree with you, because coverage of Univé is the same as Menzis, but it is cheaper per month." It seems Maryam's statement defeats Ali's statement.*
*Ali continues the discussion by presenting a new piece of information that Maryam did not pay attention to. Ali says, "Do you know that based on Univé's terms, one has to buy the insurance from the first of February. But we arrived here the first of March!" Thus, if we choose Univé we have to pay one month extra. That is, we would have to pay more in a year by buying Univé.*
*By this new piece of information Ali defeats Maryam's statement and also supports his first statement. There is a saying "The one who laughs last, laughs best." The discussion between Ali and Maryam is in Figure 1.1.*

- $a_1$: Menzis is the best insurance
- $a_2$: You have to pay Univé from 1st Feb. You arrived 1st March. Right?!
- $m_1$: Univé is cheaper per month with the same coverage!

**Figure 1.1:** Discussion between two agents for buying an insurance

*The discussion between Ali and Maryam can be illustrated formally by a directed graph, as in Figure 1.2. In the associated graph each node indicates an argument and each vertex shows attack between arguments. In Figure 1.2 a directed arrow from $m_1$ to $a_1$ represents that there is a conflict between these two arguments and argument $m_1$ attacks argument $a_1$. The graph in Figure 1.2 is a formal way of presenting the discussion between Ali and Maryam, presented in Figure 1.1.*

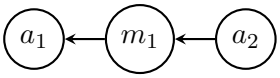


**Figure 1.2:** The graph illustration of the discussion of Example 1.1

The formal model of abstract argumentation can be used to study argumentative reasoning. This formal model abstracts away from the content of argument, and focuses on the attack relation among arguments. The formal model of the discussion Example 1.1 presents that there is an attack from $a_2$ to $m_1$, and there is an attack from $m_1$ to $a_1$. While each argument can be considered as believable/acceptable as an isolated argument, it is not reasonable/rational to accept or believe all arguments together. For instance, in Example 1.1, it is not rational to choose all arguments together, i.e., $\{a_1, m_1, a_2\}$.

It is common in each argumentation that the acceptance of some of the arguments is incompatible with the acceptance of other arguments. A goal of each debate/discussion is to find which positions are acceptable via argumentation. The idea of argumentative reasoning is that an argument or a statement is believable/acceptable if it can be defended successfully against its counterarguments. In any argumentation, we are eager to identify the set of jointly acceptable arguments. In other words, a key question in formal argumentation is "How can we choose believable/acceptable arguments in a given formalism?" The formal model of the discussion of Example 1.1 implies that argument $a_2$ is acceptable, since it is not attacked by any arguments. Furthermore, argument $m_1$ is not acceptable since it is attacked by $a_2$ and there is no defence for $m_1$. Although argument $a_1$ is attacked by argument $m_1$, the set $\{a_1, a_2\}$ is acceptable, since $a_1$ is defended by $a_2$.

Roughly speaking, in a formal argumentation process we can distinguish the following steps: 1. We consider a given knowledge base. 2. From the knowledge base we construct the abstract representation of arguments and model the relation among arguments. 3. After modeling an argumentation in an abstract way, we evaluate the arguments.

Based on (Atkinson et al., 2017) argumentation consists of five layers: structural, relational, dialogical, assessment, and rhetorical, although the distinction among these layers may not be strict in some contexts. Each layer is concerned with answering a query, as follows.

- The structure layer deals with the question 'How are arguments constructed?' There are a number of ways to produce arguments from a given knowledge base. One can consider an argument as a pair $(\phi, \alpha)$ where $\phi$ is a minimal set of consistent formulas that logically entails $\alpha$, with respect to a given logical system. There are other approaches to extract arguments in a knowledge base, for instance, assumption-based argumentation (ABA for short) (Toni, 2012, 2014),

ASPIC+ (Modgil and Prakken, 2014), and defeasible logic programming (DeLP) (García and Simari, 2004; Verheij, 2003b). The field of argument mining, which studies the extraction of arguments from natural language, has recently received increased attention (Cabrio and Villata, 2018; Lippi and Torroni, 2015). The structure of arguments is beyond the topic of this work, and not further discussed.The interested reader on argument mining techniques and applications can see (Budzynska et al., 2014; Lippi and Torroni, 2016).

- The relational layer is concerned with the question 'What are the relations among arguments?' After extracting arguments from a knowledge base, it is necessary to clarify the relation among arguments to see how an argument relates to other arguments in favor and against the argument (Pollock, 1987). To this end, one can use the structure of arguments (Prakken, 2010).

- The dialogical layer addresses the question 'How can argumentation be undertaken in dialogues?' Basically, the term 'dialectical method' refers to a discussion among two or more people who have different points of view about a subject but are willing to find a reasonable conflict resolution by argumentation. In classical philosophy, dialectic is a method of reasoning based on arguments and counterarguments (Krabbe, 2006; Macoubrie, 2003). There are several types of dialogue: inquiry, negotiation, information seeking, deliberation, and persuasion. Based on the rules of a dialogue, agents argue for or against an argument. In a discussion we argue in a cooperative or competitive manner to reach an agreement.

- The assessment layer answers the question 'How can arguments be evaluated and conclusions drawn?' Answering this question leads to the introduction of several types of semantics in each argumentation formalism. Evaluation of arguments in argumentation formalisms has received increased attention in the two last decades. We present two powerful formalisms of argumentation in Chapter 2, namely, abstract argumentation frameworks (AFs), introduced informally in 1.2.2, and abstract dialectical frameworks (ADFs). Then, we present a set of semantics of AFs and ADFs in Section 2.3 and Section 2.5, respectively. For the purpose of automated reasoning, a number of solvers for formalisms of argumentation has been proposed. Furthermore, the International Competition on Computational Models of

Argumentation (ICCMA) was organized to evaluate the solvers.[1]

- The rhetorical layer clarifies the question 'How can the argumentation be adapted to convince agents?' This layer may be absent in some contexts, for instance, when arguments are built from a knowledge base. However, when an agent tries to persuade another agent to do something, then it seems that some rhetorical device is used (Chalaguine and Hunter, 2020; Hunter, 2018). Rhetorical aspects of argumentation are not a topic of our work.

### 1.2.2 Dung's Argumentation Frameworks

Since the landmark paper by Dung (1995) has been published in 1995, abstract argumentation frameworks (AFs for short) have gained more and more significance in the AI domain. AFs have become a base for formal and computational argumentation (Baroni et al., 2020).The reader may find the definition and examples of AFs in the background chapter in Section 2.3. Some reasons to show that AFs are significant frameworks of argumentation are as follows.

- First of all, AFs have proven useful to capture the essence of different non-monotonic formalisms. It is shown in (Dung, 1995) that several non-monotonic reasoning formalisms from the AI domain, such as Pollock's defeasible reasoning (Pollock, 1987), Reiter's default logic (Reiter, 1980), and logic programming (Gabbay et al., 1998) can be regarded as instances of AFs.

- Further, in (Dung, 1995) it is shown that AFs can capture the solutions of some well-known practical problems, namely, the theory of $n$-person games and the well-known stable marriage problem. Recently, in (Bistarelli and Santini, 2020) the connection between several forms of AFs and several kinds of stable matching problems has been studied.

- In addition, compared to other non-monotonic formalisms (which are built on top of classical logical syntax), AFs are a much simpler formalism indeed, they are just directed graphs in which nodes present argument and directed edges indicate attack relation among arguments.

[1] `http://argumentationcompetition.org`

- Moreover, AFs are nowadays an integral concept in several advanced argumentation-based formalisms in the sense that their semantics are defined based on a formal connection to Dung AFs, for instance by a translation, instantiation or extension.

- Furthermore, the simplicity of the syntax of AFs together with the powerful semantics of AFs have made them an attractive modeling and evaluating tool in diverse areas, like multi-agent systems (McBurney et al., 2012), multi-agent negotiation (Amgoud et al., 2007) and legal reasoning (Bench-Capon and Dunne, 2005).

- Based on Google Scholar there are more than 4000 citations to (Dung, 1995).

- As mentioned in Section 1.1.2, automated reasoning methods in the legal domain have a long history (see (Sergot et al., 1986)). Since then many connections to abstract argumentation have been made. In (Bench-Capon, 2020) the role of AFs in AI and Law has been discussed.

- The complexity of reasoning problems that can be defined for several semantics for AFs is well understood (Dvořák and Dunne, 2018) and ranges from tractability up to the second level of the polynomial hierarchy. Furthermore, the analysis of complexity for restricted classes of AFs has also been studied. Such restrictions can make decision problems easier from a complexity perspective (Dvořák et al., 2012).

- Due to the fact that AFs have become the centerpiece of higher-level argumentation systems, there is a growing interest in efficient solving techniques for reasoning tasks within AFs. This is witnessed by efficient algorithms for reasoning tasks in AFs in terms of answer-set-programming (Egly et al., 2010), and software systems for solving reasoning tasks in AFs (Cerutti et al., 2017; Charwat et al., 2015).

- Finally, the relevance of AFs is witnessed by the *International Competition on Computational Models of Argumentation* (ICCMA), where systems for solving different problems on AFs compete on different tracks (Thimm and Villata, 2017).

The fundamental contribution of Dung is to abstract away from the content of particular arguments and to focus only on conflicts among arguments,

where each argument is viewed as an atomic item. The only information AFs take into account is whether an argument attacks another one or not. Conflicts among arguments is a key factor for having an argumentative discussion, since otherwise there is nothing to argue about. The discussion of Example 1.1, depicted in Figure 1.2 is an instance of an AF.

A key query is "Which sets of arguments fit together, or which set of arguments are acceptable together." Semantics single out coherent subsets of arguments which "fit" together, according to specific criteria (Baroni et al., 2011). In other words, each semantics clarifies a point of view of accepting a set of arguments together, thus, there exist several types of semantics. More formally, an AF semantics takes an argumentation framework as input and produces as output a collection of sets of arguments, called extensions. An extension is a jointly acceptable set of arguments.

In abstract argumentation, the most basic concept underlying nearly all semantics is conflict-freeness: a set of arguments is called conflict-free if it does not contain any conflicting arguments. Different semantics provide different ways to solve the inherent conflicts between statements. Furthermore, in the presence of conflicts an argument cannot be accepted just because it exists, but it has to be defended against possible counter-arguments. In AFs, this intuition is captured by the notion of admissibility, which also plays an important role with respect to rationality postulates (Caminada and Amgoud, 2007). In the AF of Figure 1.2, it is not reasonable to accept conflicting arguments. Thus, argument $m_1$ can be chosen with neither $a_1$ nor $a_2$. There is no doubt on the acceptance of $a_2$, since no counterargument was proposed against $a_2$. Furthermore, we can accept $a_1$ and $a_2$ jointly, since $a_1$ is defended by $a_2$ against the attack of $m_1$. The set $\{a_1, a_2\}$ would be an admissible extension of this AF, since there is no conflict among $a_1$ and $a_2$ (there is no direct edge among them in Figure 1.2) and for any attacker of this set, there is a defender inside of the set. In other words, set $\{a_1, a_2\}$ is admissible since not only there are no conflicts among the elements of this set, but also this set can defend its arguments against its attackers. Often a new semantics is an adaptation of an already existing one by introducing further restrictions on the set of accepted arguments (that are chosen together) or possible attackers.

Although AFs are popular in the modeling of argumentation and are widely used and studied within AI, AFs are limited to model an elementary attack relation among arguments. However, the relation among arguments might be more diverse than simple attack. For instance, an argument may not be strong enough to defeat another argument, but jointly with another

argument it may do so. However, this cannot be modeled explicitly via AFs. Also one cannot directly model the support relation between arguments via AFs. To overcome such deficiencies of AFs and still utilize the capabilities of AFs, several generalizations of AFs have been proposed to present different types of relations among arguments beyond simple attack. For instance, 1. the notion of collective attack is presented in (Dvořák et al., 2020), to model the notion that a set of arguments together attack an argument and a single argument within this set is not powerful enough to attack the argument. 2. In (Cayrol and Lagasquie-Schiex, 2009) the notion of support relation among arguments is presented. 3. Also a formalism presented by Verheij (2003b) is expressive enough to model nested support and attack ( i.e., support/attack of the support/attack relation) relations among arguments. 4. Further, the notion of preference among arguments that allows to rank arguments is presented in (Amgoud and Cayrol, 2002). This framework allows to evaluate arguments based on their values and their relation with other arguments. See (Brewka et al., 2014; Baroni et al., 2020) for an overview of generalizations of AFs. In Section 1.2.3 we present an outline of an expressive logical generalization of AFs which is the basis of this PhD thesis.

### 1.2.3 Abstract Dialectical Frameworks

Abstract dialectical frameworks (ADFs for short) are generalizations of Dung argumentation frameworks where arbitrary logical relationships among arguments can be formalized with so-called acceptance conditions which are attached to the arguments (Brewka and Woltran, 2010; Brewka et al., 2017b). These acceptance conditions are usually in the form of propositional logic formulas. This allows to express notions of support, collective attacks, and other complex relations which bring more modeling capacity for ADFs. For instance, the acceptance condition of an argument receiving several individual attacks from other arguments would be the conjunction of negated atoms, one for each argument. Or, if two arguments jointly attack an argument this can be presented by the disjunction of the negation of those atoms. It has been shown that ADFs unify several generalizations of AFs, namely AFs with collective attacks (Nielsen and Parsons, 2006), and bipolar AFs (Cayrol and Lagasquie-Schiex, 2005). Due to their flexibility in formalizing relations between arguments, ADFs have recently been used in several applications; in legal reasoning (Al-Abdulkarim et al., 2014; Collenette et al., 2020; Al-Abdulkarim et al., 2016), online dialog systems (Neugebauer, 2017, 2019; Pührer, 2017), and

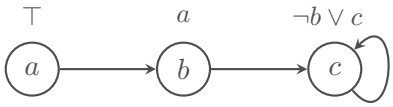


**Figure 1.3:** ADF of Example 1.2

text exploration (Cabrio and Villata, 2016). Example 1.2 presents a simple instance of an ADF.

**Example 1.2** *An instance of an ADF is depicted in Figure 1.3. This ADF contains three arguments* $\{a, b, c\}$*. Dependencies between arguments are shown by the directed edges in the associated graph. Here edges are interpreted not necessarily as attacks, but as abstract relations. The edges in the associated graph are called links. If there is a link from argument* $a$ *to* $b$*, we say that* $a$ *is a parent of* $b$*. The concrete relation between an argument and its parents is specified in the acceptance condition. Each argument has an acceptance condition shown as propositional formula attached to each node. Given the status of each parent of an argument in the attached propositional formula, the acceptance condition indicates under which condition an argument is acceptable. For instance, argument* $a$ *is always acceptable, denoted by the acceptance condition* $\top$*. By the acceptance condition of* $b$*, this argument is acceptable if and only if* $a$ *is also accepted. This can be interpreted as a support relation. Further, the acceptance condition of* $c$*, namely* $\neg b \vee c$ *says that* $c$ *is acceptable if and only if either* $b$ *is denied or* $c$ *is accepted. The relation between* $b$ *and* $c$*, i.e.,* $(b, c)$ *is an AF like attack. However, the acceptance of* $c$ *is more complex than an AF attack, since it combines a support from* $c$ *and an attack from* $b$*.*

The semantics of ADFs are defined based on three-valued interpretations that assign each argument to true (**t**), false (**f**) or undecided (**u**). For instance, in Example 1.2, the interpretation that assigns argument $a$ to **t**, argument $b$ to **t** and argument $c$ to **f** would be an admissible interpretation, presented formally in Definition 2.47. In the following we present some of the reasons that clarify why ADFs are expressive formalisms of argumentation.

- ADFs are a proper generalisation of AFs. Thus, ADFs are at least as expressive as AFs. Hence, the problems that can be presented in AFs are also presentable in ADFs (Strass, 2013b).

- The notions of relation between arguments in ADFs are more flexible than AFs, and can represent simple attack or support, joint attack or support, or any logical mix of these.

- ADFs provide nearly all standard semantics of AFs (Brewka and Woltran, 2010; Brewka et al., 2013; Polberg et al., 2013; Strass, 2013b).

- ADFs are expressive enough to unify several generalisations of AFs, for instance, SETAFs (Nielsen and Parsons, 2006) well-studied in (Dvořák et al., 2020; Polberg, 2016; Linsbichler et al., 2016).

- The distinction between supported and stable models from logic programming is present in ADFs but is missing in AFs. Specifically, ADFs allow cyclic support dependencies among arguments.

- ADFs are expressive enough to model non-monotonic knowledge representation languages, as investigated in (Alcântara and Sá, 2018; Heyninck et al., 2020).

- ADFs are proposed in a number of applications, for instance, in legal reasoning (Al-Abdulkarim et al., 2016, 2014), online dialog systems (Neugebauer, 2017, 2019), the instantiation of defeasible theories (Strass, 2014), and text exploration (Cabrio and Villata, 2016).

- The additional expressiveness of ADFs comes with the price of typically a higher computational complexity (Strass and Wallner, 2015). Specifically, reasoning in ADFs spans the first three (rather than the first two, as for AFs) levels of the polynomial hierarchy.

Furthermore, other research investigating ADFs (Brewka et al., 2011; Brewka and Gordon, 2010; Ellmauthaler, 2012; Strass and Wallner, 2015; Strass, 2013a, 2018; Wallner, 2020) can be found in the literature. In this thesis, ADFs are further investigated. The main contributions of this thesis are explained comprehensively in Section 1.3.

## 1.3 Main Contributions and Thesis Outline

This thesis is broadly concerned with studying different aspects of a formalism of argumentation called abstract dialectical frameworks (ADFs),

first introduced in (Brewka and Woltran, 2010) and then further explored in (Brewka et al., 2013, 2017a, 2018a). The main objectives are as follow:

1. Presenting semantics for ADFs that have not been introduced before (Part II of the thesis);
2. Presenting discussion games for already existing semantics of ADFs (Part III); and
3. Investigating subclasses and a superclass of ADFs (Part IV).

Taking into account the new set of semantics for ADFs, we introduce the concept of strong admissibility semantics, in Chapter 3, and semi-stable semantics, in Chapter 5. Then, in Chapter 6, we introduce discussion games for grounded semantics, in Chapter 7, we introduce discussion games for preferred semantics of ADFs to answer the credulous decision problems of ADFs under these semantics. Discussion games presented in this thesis show a dialogical proof procedure behind these two semantics of ADFs. In order to study the expressiveness of ADFs we focus on subclasses and a generalization of ADFs. With regards to the subclasses of ADFs, first we introduce the subclasses of ADFs, in Chapter 8, and we study whether these subclasses fulfill the same properties of the similar subclasses in AFs. Then, in Chapter 9 we study the relation between a generalization of AFs, namely SETAFs and a subclass of ADFs. Moreover, we compare the expressiveness of subclasses of ADFs. At the end, in Chapter 10, we present a generalization of ADFs and show how it can be used to model and evaluate a practical problem. The main contribution of each chapter is explained with further details in the following. Note that this overview contains undefined terms that will be explained in the corresponding chapters.

**Part II: Semantics.** The first major topic is the introduction of a new type of semantics for ADFs: strong admissibility semantics. A reason of presenting new types of semantics is that the issue of argumentation semantics of formalisms of argumentation has been the subject of much recent study, (see (Baroni et al., 2011) as an overview of semantics of AFs). Much research on the topic of formal argumentation is based on abstract argumentation frameworks (AFs) of Dung (1995) and the generalizations of AFs. One central question in AFs is "Which sets of arguments can be accepted jointly?," where each such set is called an extension. Different answers to this question correspond to different definitions of argumentation

semantics. The question of how to define argumentation semantics is a critical one in any formalism of argumentation.

**Chapter 3: Strong Admissibility Semantics of ADFs** We first introduce a major new type of semantics for ADFs, namely, strong admissibility semantics. Similar to AFs the concept of grounded semantics is an important point of view of acceptance of arguments in ADFs. Each ADF has a unique grounded interpretation which essentially consists in an unquestionable assignment of truth values to arguments. Thus, it is critical to investigate the truth value of a queried argument in the grounded interpretation of an ADF. While, the grounded interpretation only presents the truth value of arguments which are unquestionable, it is required to explain why a queried argument has a specific truth value in the grounded interpretation.

Thus, the first contribution of this work is to present the notion of strong admissibility semantics for ADFs, in Chapter 3. We show that the notion of strong admissibility semantics of ADFs presented in this work will satisfy the following conditions which are akin to the properties of the notion of strong admissibility semantics of AFs.

1. Strong admissibility is defined in terms of strongly justified arguments.
2. Strongly justified arguments are recursively reconstructed from their strongly justified parents.
3. Each ADF has at least one strongly admissible interpretation.
4. The set of strongly admissible interpretations of ADFs forms a lattice with as least element the trivial interpretation and as maximum element the grounded interpretation.
5. The strong admissibility semantics can be used to answer whether an argument is justifiable under grounded semantics.
6. The strong admissibility semantics of ADFs is different from the admissible, conflict-free, complete and grounded semantics of ADFs.
7. The strong admissibility semantics for ADFs is a proper generalization of the strong admissibility semantics for AFs.

**Chapter 4: Complexity of Strong Admissibility Semantics** Computational complexity of strong admissibility semantics of AFs is studied in (Dvořák and Wallner, 2020; Caminada and Dunne, 2020). However, the

computational complexity analysis under the strong admissibility semantics of ADFs has not beed studied. Thus, the second contribution of Part II of this work is studying the complexity of reasoning tasks under the strong admissibility semantics of ADFs, in Chapter 4, as follows.

1. The credulous decision problem, i.e., whether there exists a strongly admissible interpretation that satisfies the queried argument, is coNP-complete.
2. The skeptical decision problem, i.e., whether all strongly admissible interpretations satisfy a queried argument, is trivial.
3. The verification problem, i.e., whether a given interpretation is a strongly admissible interpretation of an ADF, is coNP-complete.
4. The strong justification problem for an argument in an interpretation, i.e., whether an argument is strongly justified in an interpretation is coNP-complete.
5. The problem of finding a smallest witness of strong justification of an argument, i.e, whether there exists a minimal strongly admissible interpretation that satisfies a queried argument, is $\Sigma_2^{\mathsf{P}}$-complete.

**Chapter 5: Semi-Stable Semantics of ADFs** The next contribution of Part II is to define the notion of semi-stable semantics of ADFs. Stable semantics reflect the 'black-and-white' character of the classical logic in non-monotonic frameworks. Although it is possible that a non-monotonic framework does not have any stable extension, researchers in these domains sometimes preferred to have no outcome as opposed to an imperfect one, like a preferred extension. To overcome the possibility that some AFs do not have any stable extension, the concept of semi-stable semantics have been introduced for AFs, first in (Verheij, 1996) (under a different name) then further investigated in (Caminada, 2006).

Semi-stable semantics of AFs is a way of approximating stable semantics when a given AF does not have any stable extension. Key characteristics of semi-stable semantics in AFs are as follow.

1. It is placed between stable semantics and preferred semantics;
2. If an AF has at least one stable extension, then the set of stable extensions and semi-stable extensions coincide;
3. Each finite AF has at least one semi-stable extension.

As the main contribution in Chapter 5, we introduce the notion of semi-stable semantics of ADFs, to approximate stable semantics in the cases that a given ADF does not have any stable interpretation. Stable semantics of AFs are generalised to ADFs in two ways, that is, to two-valued semantics and stable semantics. In ADFs the notion of stable model is defined based on the notion of two-valued model. Thus, in ADFs the user can choose whether support cycles should be accepted or rejected, by choosing two-valued models or stable models as semantics. An ADF may have no stable model because of the following reasons.

- On the one hand, if a given ADF does not have any two-valued model, then it does not have any stable model.
- On the other hand, an ADF may have two-valued models, while none of them is a stable model.

To present the notion of semi-stable semantics of ADFs, we focus on the first issue in this thesis, i.e., when a given ADF does not have a two-valued model. To define the notion of semi-stable semantics for ADFs, we follow the same method as for stable semantics of ADFs. That is, first we introduce the notion of semi-two-valued semantics. Then, we pick semi-stable models among semi-two-valued models of a given ADF. The idea is detecting cycle supports via semi-stable semantics when an ADF does not have a two-valued model.

We will show that the semi-stable semantics and semi-two-valued model presented in Chapter 5 satisfy the following conditions which are akin to the properties of the notion of semi-stable semantics of AFs.

1. A semi-stable model and a semi-two-valued model of a given ADF should maximize the union of the sets of the accepted and of the rejected/denied arguments among all complete interpretations.
2. Each semi-stable model and each semi-two-valued model is a preferred interpretation;
3. Each stable model is a semi-stable model and a semi-two-valued model;
4. Each finite ADF has at least one semi-two-valued model;
5. If an ADF has a stable model, then the set of stable models coincides with the set of semi-stable models;
6. The notion of semi-stable semantics and semi-two-valued semantics for ADFs is a proper generalization of semi-stable semantics for AFs.

**Part III: Discussion Games.** In the realm of discussion games for semantics we present two games for two of the semantics of ADFs; grounded semantics, in Chapter 6, and preferred semantics, in Chapter 7. In ADFs semantics have been introduced to indicate points of view of evaluating of arguments, defined based on three-valued interpretations. In ADFs, an *admissible* interpretation does not contain any unjustifiable information about the arguments and a *preferred* interpretation presents maximum information about the arguments without losing admissibility. Furthermore, an interpretation is *grounded* if it collects all the information that is beyond any doubt.

Answering whether there exists an interpretation of a particular type of semantics in which an argument is justifiable, i.e., has a given value, is a fundamental issue: this decision problem is called *credulous decision problem*. Answering credulous decision problems under semantics of ADFs is a significant issue. In application it is significant not only to answer whether a queried argument is justifiable under a type of semantics but also to explain why it is so. Although dialectical methods have a role in determining semantics of both AFs and ADFs, the roles are not obvious in the definition of semantics. To cover this gap, quite a number of works have been presented to show that semantics of AFs can be interpreted in terms of structural discussion (Jakobovits and Vermeir, 1999; Prakken and Sartor, 1997; Caminada, 2018; Dung and Thang, 2007). The idea is that these discussion games can be used as proof procedures for the semantics of AFs .

Despite the fact that the essence of argumentation is dialogue, semantics of ADFs specify the truth values of arguments, without indicating how interpretations are to be constructed. This raises the question whether semantics of ADFs are expressible in terms of discussion games (Barth and Krabbe, 1982). Because of the special structure of the ADFs, the existing methods used to interpret semantics of AFs cannot be reused in ADFs. This motivates us to study whether there is a discussion game and a winning strategy for justification of an argument under a specific semantics.

**Chapter 6: Grounded Discussion Games** As the first contribution of this part of the thesis we consider grounded semantics of ADFs. Answering the credulous decision problem of ADFs under grounded semantics is a critical issue, since each ADF has a unique grounded interpretation and no one has any doubt on the truth values of arguments in the grounded inter-

pretation. Furthermore, it is required to explain why a queried argument has a specific truth value in the grounded interpretation. In Chapter 6, we present a grounded discussion game for ADFs, to show that grounded semantics of ADFs are interpretable in terms of structural discussion. That is, a queried argument is justifiable under grounded semantics of a given ADF iff it is possible to win the associated discussion game. This makes it possible to use the discussion games for the purpose of explanation "why is an argument justifiable in the grounded interpretation?"

On the one hand, a grounded discussion game, presented in Chapter 6, explains why a queried argument is justifiable under grounded semantics of a given ADF. On the other hand, strong admissibility semantics of ADFs, presented in Chapter 3, is a point of view to explain a reason of a truth value of a queried argument in the grounded interpretation. In Section 6.4 we clarify the relation between the notion of strong admissibility semantics of ADFs and the grounded discussion game.

**Chapter 7: Preferred Discussion Games** As the second contribution of Part III of the thesis we consider preferred semantics of ADFs. Answering decision problems of preferred semantics has a higher computational complexity than other semantics in ADFs (Strass and Wallner, 2015). In this chapter we present preferred discussion games to show that preferred semantics of ADFs are interpretable in terms of structural discussion.

Similar works, with the purpose of showing the proof procedure of justification of arguments in a preferred extension of AFs, have been done via dialectical games (Vreeswijk and Prakken, 2000; Dung and Thang, 2007; Modgil and Caminada, 2009; Caminada et al., 2014; Cayrol et al., 2003).

The main contributions of Chapters 6 and 7 are:

1. Presenting the discussion games which provide proof procedures to answer credulous decision problems under preferred and grounded semantics of ADFs.
2. Showing that our methods are sound and complete.

Based on the methods of discussion games, which have been presented in chapters 6 and 7, algorithms can be provided not only to answer credulous decision problems of ADFs under grounded and preferred semantics but also to be used in a human-machine dialogue.

**Part IV: Variations.** In this part of this thesis we focus on variations of ADFs to show the expressiveness of ADFs.

**Chapter 8: Investigating Subclasses of ADFs** As the main contribution of this chapter, we prove several results about subclasses of ADFs. In the following, we first explain the motivation of presenting and investigating subclasses of ADFs. To this end, we explain the state of the art of introducing subclasses of AFs.

The analysis of restricted classes of AFs has been done, since the complexity of the reasoning problems that can be defined for the several semantics for AFs ranges from tractability up to the second level of the polynomial hierarchy (Dvořák and Dunne, 2018). In (Dung, 1995), Dung already showed that the class of acyclic (also known as well-founded) AFs leads to a collapse of the different semantics. Further studies include symmetric AFs (Coste-Marquis et al., 2005) and AFs under other graph-driven restrictions (Dunne, 2007). Symmetric AFs have been proven to satisfy the property of coherence (preferred and stable semantics coincide) and relatively-groundedness (the grounded extension is given by the intersection of the preferred extensions). These restrictions make decision problems often easier from a complexity perspective.

ADFs are more flexible than AFs in formalizing relations between arguments, however, this additional expressiveness comes with the price of higher computational complexity (Strass and Wallner, 2015). It is thus natural to investigate subclasses of ADFs. Hence, the first contribution of Chapter 8 is to do a systematic investigation of subclasses of ADFs. Thus, first we define several subclasses of ADFs and investigate how the restrictions we define influence the semantic evaluation of such ADFs.

As a first contribution of Chapter 8, we introduce and study the following subclasses.

1. We introduce acyclic ADFs (i.e., the link-structure forms an acyclic graph) and we show that—analogously to well-founded AFs—the main semantics, namely grounded, complete, preferred, and two-valued model/stable semantics, coincide for this class.
2. We further investigate the concept of symmetric ADFs. In contrast to the case of AFs, we will see that properties as coherence and relatively-groundedness do not carry over to symmetric ADFs.
3. We find that the class of symmetric ADFs requires further restrictions which leads us to the classes of acyclic support symmetric ADFs

(ASSADFs) and support-free symmetric ADFs (SFSADFs). For both classes we show that they satisfy a weaker form of coherence.

4. We also show that these two classes differ in the sense that odd-cycle free SFSADFs are coherent while odd-cycle free ASSADFs are not.

As a second contribution of Chapter 8, following the work of Dunne et al. 2015, we investigate the expressiveness of our ADF subclasses in terms of signatures, i.e. the set of possible outcomes that can be achieved by ADFs (of a particular class) under the different semantics. Some of the main results of this part of the chapter are as follows:

1. We complement the results which have been obtained for general (Pührer, 2015; Strass, 2015) and bipolar ADFs (Linsbichler et al., 2016) and also compare our ADF subclasses to abstract argumentation frameworks in terms of expressiveness.

2. In particular, we show that the expressiveness of SFSADFs, ASSADFs and bipolar ADFs is equal for some of the semantics, but different for admissibility-based semantics.

**Chapter 9: Expressiveness of SETAFs and Support-Free ADFs under 3-valued Semantics** In this chapter we investigate two of the generalizations of AFs. The first formalism we consider are SETAFs as introduced by Nielsen and Parsons (2006). SETAFs extend Dung AFs by allowing for collective attacks such that a set of arguments $B$ attacks another argument $a$ but no proper subset of $B$ attacks $a$. SETAFs have received increasing interest in the last years. For instance, (Yun et al., 2018) observed that for particular instantiations, SETAFs provide a more convenient target formalism than Dung AFs. The second formalism we consider are support-free abstract dialectical frameworks (SFADFs), a subclass of ADFs. The main contributions of Chapter 9 are as follows.

1. We embed SETAFs under 3-valued labeling based semantics (Flouris and Bikakis, 2019) in the more general framework of ADFs.

2. We investigate the expressiveness of SETAFs under 3-valued semantics by providing exact characterizations of the signatures for preferred, stable, grounded and conflict-free semantics.

3. At the end, we study the relations between SETAFs and SFADFs.

**Chapter 10: A Generalization of ADFs** As the last contribution in Chapter 10 we show how ADFs can be used to model and evaluate a practical problem. To this end, we generalize ADFs to a formalism which is called numerical abstract dialectical frameworks (nADFs for short) in our work. We show that nADFs are expressive enough to formalize standard decision problems, namely expected utility problems. Then, we show how the nADF semantics can be used to choose the best set of decisions.

## 1.4 Publications

Most results in this thesis have been published in international peer reviewed workshops, conferences and journals, as follows.

**Part II: Semantics** Chapter 3 is based on (Keshavarzi Zafarghandi et al., 2021d), which is an extended version of (Keshavarzi Zafarghandi et al., 2021b).

Chapter 4 is based on (Keshavarzi Zafarghandi et al., 2021a).

Chapter 5 is based on (Keshavarzi Zafarghandi et al., 2021c).

**Part III: Discussion Games** Chapter 6 is based on (Keshavarzi Zafarghandi et al., 2020).

Chapter 7 is adapted from (Keshavarzi Zafarghandi et al., 2019a).

**Part IV: Variations** Chapter 8 is based on (Diller et al., 2020), which is an extended version of (Diller et al., 2018).

Chapter 9 is based on (Dvořák et al., 2020).

Chapter 10 is based on (Keshavarzi Zafarghandi et al., 2019b).

The full list of publications is as follows.

1. (Keshavarzi Zafarghandi et al., 2021b) Atefeh Keshavarzi Zafarghandi, Bart Verheij and Rineke Verbrugge. Strong Admissibility for Abstract Dialectical Frameworks. In: Chih-Cheng Hung, Jiman Hong, Alessio Bechini and Eunjee Song, editors, *The 36th ACM/SIGAPP Symposium on Applied Computing SAC '21.* pages 873–880. ACM, 2021. **Chapter 3**.
2. (Keshavarzi Zafarghandi et al., 2021d) Atefeh Keshavarzi Zafarghandi, Bart Verheij and Rineke Verbrugge. Strong Admissibility for Abstract Dialectical Frameworks. *Journal of Argument*

*& Computation.* pages (online first). IOS press, 2021. **Chapter 3**.

3. (Keshavarzi Zafarghandi et al., 2021a) Atefeh Keshavarzi Zafarghandi, Wolfgang Dvořák, Bart Verheij and Rineke Verbrugge. Computational Complexity of Strong Admissibility for Abstract Dialectical Frameworks. In: Leila Amgoud and Richard Booth, editors, *19th International Workshop on Non-Monotonic Reasoning (NMR).* pages 295–304. NMR, 2021. **Chapter 4**.

4. (Keshavarzi Zafarghandi et al., 2021c) Atefeh Keshavarzi Zafarghandi, Bart Verheij and Rineke Verbrugge. Semi-Stable Semantics for Abstract Dialectical Frameworks. In: Meghyn Bienvenu and Gerhard Lakemeyer, editors, *Proceedings 18th International Conference on Principles of Knowledge Representation and Reasoning (KR).* pages 422–431, KR, 2021. **Chapter 5**.

5. (Keshavarzi Zafarghandi et al., 2020) Atefeh Keshavarzi Zafarghandi, Bart Verheij and Rineke Verbrugge. A Discussion Game for the Grounded Semantics of Abstract Dialectical Frameworks. In: Henry Prakken, Stefano Bistarelli, Francesco Santini and Carlo Taticchi, editors, *Proceedings of Computational Models of Argument COMMA 2020.* Volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 431–442. IOS press, 2020. **Chapter 6**.

6. (Keshavarzi Zafarghandi et al., 2019a) Atefeh Keshavarzi Zafarghandi, Bart Verheij and Rineke Verbrugge. Discussion Games for Preferred Semantics of Abstract Dialectical Frameworks. In: Gabriele Kern-Isberner and Zoran Ognjanovic, editors, *European Conference on Symbolic and Quantitative Approaches with Uncertainty.* Volume 11726 of *Lecture Notes in Computer Science*, pages 62–73. Springer, 2019. **Chapter 7**.

7. (Diller et al., 2020) Martin Diller, Atefeh Keshavarzi Zafarghandi, Thomas Linsbichler, Stefan Woltran. Investigating Subclasses of Abstract Dialectical Frameworks. In: Special Issue: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games: 25 years later. Pietro Baroni, Francesca Toni, and Bart Verheij, editors, *Journal of Argument & Computation.* Volume 11, pages 191–219. IOS press, 2020. **Chapter 8**.

8. (Diller et al., 2018) Martin Diller, Atefeh Keshavarzi Zafarghandi,

Thomas Linsbichler, Stefan Woltran. Investigating Subclasses of Abstract Dialectical Frameworks. In: Sanjay Modgil, Katarzyna Budzynska, and John Lawrence, editors, *Computational Models of Argument - Proceedings of COMMA*. Volume 305, pages 61–72. IOS press, 2018. **Chapter 8**.

9. (Dvořák et al., 2020) Wolfgang Dvořák, Atefeh Keshavarzi Zafarghandi and Stefan Woltran. Expressiveness of SETAFs and Support-Free ADFs Under 3-Valued Semantics. In: Henry Prakken, Stefano Bistarelli, Francesco Santini and Carlo Taticchi, editors, *Computational Models of Argument - Proceedings of COMMA*. Volume 326, pages 191–202. IOS press, 2020. **Chapter 9**.

10. (Keshavarzi Zafarghandi et al., 2021b) Atefeh Keshavarzi Zafarghandi, Bart Verheij and Rineke Verbrugge.Embedding Probabilities, Utilities and Decisions in a Generalization of Abstract Dialectical Frameworks. In: Jasper De Bock, Cassio P. de Campos, Gert de Cooman, Erik Quaeghebeur and Gregory R. Wheeler, editors, *International Symposium on Imprecise Probabilities: Theories and Applications, ISIPTA*. Volume 103, pages 246–255. PMLR, 2019. **Chapter 10**.

# Chapter 2

# Background

In this chapter, we introduce the formal background for our work. We will first, in Section 2.1, give some preliminaries on propositional logic. Furthermore, in Section 2.2 we recall the basics of order theory. Then, in Section 2.3, we introduce the syntax and semantics of abstract argumentation frameworks (AFs) (Dung, 1995). Subsequently, we present two generalisations of Dung's AFs. In Section 2.4, we consider SETAFs as introduced by by Nielsen and Parsons (2006). In Section 2.5, we present the syntax and semantics of abstract dialectical frameworks (ADFs) (Brewka et al., 2018a, 2014; Polberg, 2017).

## 2.1 Propositional Logic

We assume basic knowledge of the syntax and semantics of propositional logic. For a comprehensive introduction we refer to (Enderton, 2001). Let $\mathscr{P}$ be a set of propositional atoms. Sentences, expressions and formulas are built in the language of propositional logic using logical connectives. For propositional formulas we make use of the standard connectives negation ($\neg$), logical and ($\wedge$), logical or ($\vee$), implication ($\rightarrow$), and equivalence ($\leftrightarrow$) and evaluate formulas with respect to standard semantics of propositional logic. Let $\varphi$ be a formula and let $S$ be a set of atoms. We say that $S$ is a model of $\varphi$, denoted by $S \models \varphi$, if $\varphi$ evaluates to true when atoms in $S$ are considered true and all other atoms are considered false.

A formula is in conjunctive normal form (CNF) if it is of the form $\bigwedge_{c \in C} \bigvee_{x \in c} x$, where $C$ is a set of clauses, and a clause $c$, with $c \in C$, is a disjunction of literals. The semantics of propositional formulas is defined in terms of interpretation functions. While in classical logic there are

two truth values, namely *true*, denoted by $\mathbf{t}$, and *false*, denoted by $\mathbf{f}$, in *three-valued logic* there are three truth values indicating true, false and undecided, denoted by $\mathbf{u}$.

**Definition 2.1** *Let $\mathscr{P}$ be a set of atoms. A three-valued interpretation $v$ is a function $v : \mathscr{P} \mapsto \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$, that maps each atom to one of the three truth values true ($\mathbf{t}$), false ($\mathbf{f}$), or undecided ($\mathbf{u}$). If $v$ assigns atoms to $\{\mathbf{t}, \mathbf{f}\}$, then it is a two-valued interpretation.*

**Definition 2.2** *Let $v$ be a three-valued interpretation on atoms of $A$. We write $v^{\mathbf{t}}$ for $\{a \in A \mid v(a) = \mathbf{t}\}$, $v^{\mathbf{f}}$ for $\{a \in A \mid v(a) = \mathbf{f}\}$, and $v^{\mathbf{u}}$ for $\{a \in A \mid v(a) = \mathbf{u}\}$.*

An interpretation $v$ over a propositional formula $\varphi$ indicates a particular logical point of view on the propositions of formula $\varphi$, that is, $v$ shows that each proposition of $\varphi$ is assigned to either $\mathbf{t}$, $\mathbf{f}$, or $\mathbf{u}$. Furthermore, according to the standard evaluation of the formulas of propositional logic, we consider two-valued interpretations that assign each atom to either $\mathbf{t}$ or $\mathbf{f}$, that is $v(\varphi) \in \{\mathbf{t}, \mathbf{f}\}$, based on the standard evaluation of $\varphi$, as follows.

**Definition 2.3** *Let $\varphi$ be a propositional formula over atoms of $A$ and let $v$ be a two-valued interpretation over the atoms of $\varphi$, i.e., $v : A \mapsto \{\mathbf{t}, \mathbf{f}\}$. The evaluation of $\varphi$ under $v$, denoted by $v(\varphi)$, is defined recursively as follows.*

- *if $\varphi = a$ and $a \in A$, than $v(\varphi) = v(a)$,*
- *if $\varphi = \top$, then $v(\varphi) = \mathbf{t}$;*
- *if $\varphi = \bot$, then $v(\varphi) = \mathbf{f}$;*
- *if $\varphi = (\neg\psi)$, then*

$$v(\varphi) = \begin{cases} \mathbf{t} & \text{if } v(\psi) = \mathbf{f}, \\ \mathbf{f} & \text{if } v(\psi) = \mathbf{t} \end{cases}$$

- *if $\varphi = (\psi \wedge \theta)$, then*

$$v(\varphi) = \begin{cases} \mathbf{t} & \text{if } v(\psi) = \mathbf{t} \text{ and } v(\theta) = \mathbf{t}, \\ \mathbf{f} & \text{otherwise} \end{cases}$$

- *if* $\varphi = (\psi \vee \theta)$, *then*

$$v(\varphi) = \begin{cases} \mathbf{f} & \text{if } v(\psi) = \mathbf{f} \text{ and } v(\theta) = \mathbf{f}, \\ \mathbf{t} & \text{otherwise} \end{cases}$$

- *if* $\varphi = (\psi \rightarrow \theta)$, *then*

$$v(\varphi) = \begin{cases} \mathbf{f} & \text{if } v(\psi) = \mathbf{t} \text{ and } v(\theta) = \mathbf{f}, \\ \mathbf{t} & \text{otherwise} \end{cases}$$

- *if* $\varphi = (\psi \leftrightarrow \theta)$, *then*

$$v(\varphi) = \begin{cases} \mathbf{f} & \text{if } v(\psi) \neq v(\theta), \\ \mathbf{t} & \text{otherwise} \end{cases}$$

Thus, a propositional formula can be evaluated by the two-valued interpretation defined over its propositional atoms. A formula evaluates to a unique truth value via a given interpretation. Next we define some concepts of two-valued interpretation $v$ with respect to a formula $\varphi$.

**Definition 2.4** *Let* $\varphi$ *be a propositional formula and let* $v$ *be a two-valued interpretation over the atoms of* $\varphi$.

- $v$ *satisfies* $\varphi$, *denoted by* $v \models \varphi$, *if and only if* $v(\varphi) = \mathbf{t}$. *It is said that* $v$ *is a model of* $\varphi$.

- $\varphi$ *is satisfiable if and only if there exists an interpretation* $v$ *over the variables of* $\varphi$ *where* $v \models \varphi$.

- $\varphi$ *is a valid formula (or a tautology), denoted by* $\models \varphi$ *if for each interpretation* $v$ *over the variables of* $\varphi$ *it holds that* $v \models \varphi$.

- $\varphi$ *is unsatisfiable if and only if no interpretation makes the formula true, that is, for each* $v$ *over the atoms of* $\varphi$, *it holds that* $v(\varphi) = \mathbf{f}$, *denoted by* $v \not\models \varphi$.

- $\varphi$ *is falsifiable if and only there exists an interpretation* $v$ *over the set of variables of* $\varphi$ *where* $v \not\models \varphi$.

For instance, let $\varphi = \neg b \vee a$ be a propositional formula and let $v$ be an interpretation that assigns both $a$ and $b$ to $\mathbf{t}$, then $v(\varphi) = v(\neg b \vee a) = \mathbf{t}$, that is, $v$ satisfies $\varphi$. On the other hand, let $v'$ be an interpretation that assigns $a$ to $\mathbf{f}$ and $b$ to $\mathbf{t}$, then $v'(\varphi) = v'(\neg b \vee a) = \mathbf{f}$, that is, $v'$ does not satisfy $\varphi$. Thus, $\varphi$ is a satisfiable formula, but not a valid formula of propositional logic.

Furthermore, a propositional formula $\varphi$ can be evaluated under three-valued interpretations $v$, for which we write $\varphi^v$.[1]

**Definition 2.5** *Let $\varphi$ be a propositional formula, let $A$ be the set of atoms of $\varphi$, let $a \in A$, and let $v$ be a three-valued interpretation over the set of atoms of $\varphi$, i.e., $v : A \mapsto \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. The* partial evaluation *of $\varphi$ under a three-valued interpretation $v$ is the formula $\phi^v$ which is defined recursively as follows.*

- *if $\varphi = a$ and $v(a) = \mathbf{t}$, then $\varphi^v = \top$,*
- *if $\varphi = a$ and $v(a) = \mathbf{f}$, then $\varphi^v = \bot$,*
- *if $\varphi = a$ and $v(a) = \mathbf{u}$, then $\varphi^v = a$,*
- *if $\varphi = \neg\psi$, then $\varphi^v = \neg\psi^v$,*
- *if $\varphi = (\psi \wedge \theta)$, then $\varphi^v = (\psi^v \wedge \theta^v)$,*
- *if $\varphi = (\psi \vee \theta)$, then $\varphi^v = (\psi^v \vee \theta^v)$,*
- *if $\varphi = (\psi \rightarrow \theta)$, then $\varphi^v = (\neg\psi^v \vee \theta^v)$,*
- *if $\varphi = (\psi \leftrightarrow \theta)$, then $\varphi^v = (\neg\psi^v \vee \theta^v) \wedge (\psi^v \vee \neg\theta^v)$.*

Intuitively, the partial evaluation of $\varphi$ with a three-valued interpretation $v$ replaces variable $a$ of $\varphi$ with $\top$ or $\bot$ if $v(a)$ is equal to $\mathbf{t}$ or $\mathbf{f}$, respectively; and if $v(a) = \mathbf{u}$, then $a$ remains unchanged. For instance, let $\varphi = \neg a \wedge b$ and let $v = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}\}$. Since $v(b) = \mathbf{u}$, variable $b$ will remain unchanged in the evaluation of $\varphi$ under $v$, while variable $a$ will be replaced by $\bot$ so we get $\varphi^v = \neg\bot \wedge b$, which is logically equivalent to $b$. That is, the partial evaluation of $\varphi$ under $v$ takes the two-valued part of $v$ and replaces the evaluated variables with their truth values:

$$\varphi^v = \varphi[a/\top \ : v(a) = \mathbf{t}][a/\bot \ : \ v(a) = \mathbf{f}]$$

[1]Note that the notation of $\varphi^v$ which we use in this work does not exactly correspond to the evaluation of formulas in three-valued Kleene logics as presented in (Priest, 2008). We use this notation based on substitutions as it is used in literature of argumentation theory (Brewka et al., 2018a, page 15).

## 2.2 Ordering Relations

Moreover, we use standard mathematical concepts such as functions, preorders, and lattices. Let $S$ be a set of variables and $\leq S \times S$ be a binary relation.

- $\leq$ is called a *partial order* if it is reflexive, antisymmetric, and transitive.
- A set $S$ that is equipped with a partial order $\leq$ is called a *partially ordered set (or poset)*, denoted as a pair $(S, \leq)$.
- Let $(S, \leq)$ be a poset and let $S' \subseteq S$. An element $x \in S$ is said to be an *upper bound* of $S'$ if $s \leq x$, for each $s \in S'$.
- An upper bound $x$ of $S'$ is said to be its *least upper bound*, (or join, or supremum), if $x \leq a$ for each upper bound $a$ of $S'$.
- Dually, the notions of *lower bound* and *the greatest lower bound*, (or meet, or infimum) are defined for a poset.
- A poset $(S, \leq)$ is called a *join-semilattice* if each two-element subset $\{a, b\} \subseteq S$ has a join. Furthermore, it is called a meet-semilattice if each two-element subset has a meet.
- A poset $(S, \leq)$ is called a *lattice* if it is both a join-semilattice and a meet-semilattice.
- A poset $(S, \leq)$ is called a *complete lattice* if every $S'$ such that $S' \subset S$ has both a greatest lower bound and a least upper bound in $S$.
- A poset $(S, \leq)$ is called a *complete meet-semilattice* if every non-empty subset $S' \subseteq S$ has a greatest lower bound in $S$ and every ascending chain in $S$ has a least upper bound in $S$.

Truth values can be ordered via the information content.

**Definition 2.6** *The truth values, i.e.,* $\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ *are strictly ordered by* $<_i$ *such that* $\mathbf{u} <_i \mathbf{t}$ *and* $\mathbf{u} <_i \mathbf{f}$ *and no other pair of truth values are related by* $<_i$*. Relation* $\leq_i$ *is the reflexive closure of* $<_i$*.*

The pair $(\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}, \leq_i)$ is a complete meet-semilattice with the meet operator $\sqcap_i$, such that $\mathbf{t} \sqcap_i \mathbf{t} = \mathbf{t}$ and $\mathbf{f} \sqcap_i \mathbf{f} = \mathbf{f}$, while it returns $\mathbf{u}$ otherwise.

The meet of two interpretations $v$ and $w$ is defined pointwise as $(v \sqcap_i w)(a) = v(a) \sqcap_i w(a)$ for all $a \in A$. The notion of meet operator between

two interpretations can be extended to the meet of several interpretations. Let $v_1, \ldots v_n$ be interpretations, the meet of these interpretations is denoted by $\bigsqcap_{i=1}^{n} v_i$ and it is presented as follows, for each $a \in A$.

$$\prod_{i=1}^{n} v_i(a) = \begin{cases} \mathbf{t} & \text{if for each } i, \text{ it holds that } v_i(a) = \mathbf{t}, \\ \mathbf{f} & \text{if for each } i, \text{ it holds that } v_i(a) = \mathbf{f}, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

Interpretations can be ordered via $\leq_i$-ordering with respect to their information content, presented formally in Definition 2.7.

**Definition 2.7** *Let $v$ and $w$ be interpretations over a set of atoms $A$, interpretation $v$ is at least as informative as $w$, denoted by $w \leq_i v$, if and only if*

$$w(a) \leq_i v(a) \text{ for all } a \in A$$

For instance, let $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$ and $w = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}\}$ be two interpretations over a set of atomes $A = \{a, b, c\}$. Since $w(a) = \mathbf{u} <_i v(a) = \mathbf{t}$, $w(b) = \mathbf{u} <_i v(b) = \mathbf{f}$, and $w(c) = \mathbf{f} \leq_i v(c) = \mathbf{f}$, it holds that $w \leq_i v$. It is well-known that this definition ensures that $\leq_i$ a partial order on interpretations.

## 2.3 Abstract Argumentation Frameworks

Abstract argumentation frameworks provide a formal way of presenting arguments as abstract entities with direct conflict (attack) among them. Every abstract argumentation framework (AF for short), as introduced in the landmark paper by Dung (Dung, 1995), is composed of two components 1. a set of arguments, and 2. a binary relation on this set, interpreted as attack. The formal definition of abstract argumentation frameworks is presented in Definition 2.8. In the following we assume as given a set $\mathscr{P}$ of propositional variables or atoms, which serves as the universe of arguments. That is, we assume that each argument in AF can be associated with a propositional variable.

**Definition 2.8** *(Dung, 1995) An abstract argumentation framework (AF) is a pair $(A, R)$ in which $A \subseteq \mathscr{P}$ is a set of arguments and $R \subseteq A \times A$ is a binary relation representing attacks (conflicts) among arguments.*

An AF can be represented as a directed graph in which nodes indicate the set of arguments and directed edges show conflict among arguments. A

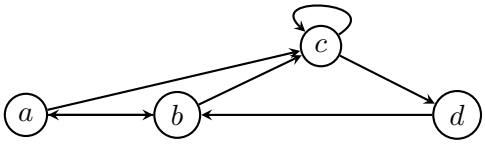


**Figure 2.1:** AF of Example 2.9

directed edge from argument $a$ to $b$ means that there is an attack from $a$ to $b$.

**Remark 2.8.1** *Let $F = (A, R)$ be an AF.*

- *For each $a, b \in A$, the notation $R(a, b)$ or $(a, b) \in R$ is used to represent that there is an attack from $a$ to $b$.*
- *$F$ is named finite argumentation framework when $A$ is a finite set of arguments. Note that in this work we assume that all AFs are finite.*

**Example 2.9** *Let $F = (A, R)$ be an AF with $A = \{a, b, c, d\}$ and $R = \{(a, b), (b, a), (a, c), (b, c), (c, c), (c, d), (d, b)\}$. The associated graph of this AF is depicted in Figure 2.1. In this AF argument $d$ attacks $b$ since $(d, b) \in R$. Plus, $(c, c) \in R$ says that $c$ attacks itself. Furthermore, $(a, b) \in R$ and $(b, a) \in R$ show that there is a symmetric attack between $a$ and $b$.*

Note that we say set $S \subseteq A$ attacks an argument $a$ if there is a $s \in S$ such that $(s, a) \in R$. For instance, in AF of Example 2.9, set $S = \{b, d\}$ attacks argument $c$, since $(b, c) \in R$. In AF $F = (A, R)$, it is said that argument $a$ is *undefeated* in $F$ if there is no $b$ such that $(b, a) \in R$, that is, if there is no input edge over $a$ in the associated graph of $F$. That is, argument $a$, whatever it is, is acceptable/believable by everyone, since there is no counterargument against $a$. In AFs arguments not only attack but also defend one another. For instance, in an AF if argument $a$ attacks $b$, and argument $b$ attacks $c$, then $a$ defends (or supports) argument $c$ against the attack of $b$.

AFs are a formalism not only for modeling argumentation but also for evaluating arguments. Thus, a topic to study is 'which argument is believable/acceptable?' or 'which sets of arguments are believable/acceptable together?' In the field it is more common and rational to say that we are collecting a set of arguments that are acceptable together rather than believable because we may accept an argument even if we do not believe

in it. For instance, a child, after arguing with her/his parents, accepts to do his/her homework whether or not the child believes that it will be good for herself/himself.

More formally, a key question is that 'how is it possible to choose arguments that fit together in an argumentation that contains conflicts?' Answering this question leads to the introduction of several types of semantics. That is, conflicts among arguments in an argumentation are resolved using appropriate semantics. Different semantics in AFs reflect the fact that there is no unique way of evaluating arguments. For an overview about argumentation semantics we refer to (Baroni et al., 2011). Another account of abstract argumentation semantics has been provided in (Baroni and Giacomin, 2007) by introducing certain principles and studying their fulfillment by the different semantics.

There are two main approaches for defining semantics of AFs extension-based and labelling-based (see (Dung, 1995; Baroni et al., 2011) for an overview). In Section 2.3.1 we present central semantics of AFs based on extensions. Then, in Section 2.3.2 we present two of the semantics of AFs based on labelling, namely admissible semantics and strongly admissible semantics. We do not present the whole set of semantics that are presented in Section 2.3.1 in the form of labelling, since we do not use those in our work. The reader interested in further information on labelling based-semantics of AFs is referred to (Baroni et al., 2011). We wrap up in Section 2.3.2 by presenting functions that map extension-based semantics of AFs to labelling-based semantics of AFs and vice versa.

### 2.3.1 Extension-based Semantics of AFs

As presented in Section 2.3, an issue of argumentation is to determine acceptable sets of arguments, meaning informally, sets able to defend themselves collectively while avoiding internal attacks. We aim to remind the reader of the formal definitions of these notions, but first let us present them informally. An extension is a set of jointly chosen arguments. A set of semantics based on extensions presents a point of view of collecting and accepting arguments together. Formally, an extension-based semantics is a function that takes an AF as an input and produces a set of extensions as the output. In this section we present the relevant extension-based semantics to our work as they are presented in (Dung, 1995). That is, we present the notions of conflict-free, admissible, preferred, complete, grounded, and stable semantics based on extensions, introduced in (Dung,

1995)[2]. Furthermore, we present the notion of ideal semantics, introduced in (Dung et al., 2007). Moreover, we present the notion of semi-stable semantics for AFs, first introduced in (Verheij, 1996) (under a different name) then further investigated in (Caminada, 2006). Finally, we present the extension-based notion of strongly admissible semantics for AFs, first introduced in the work of Baroni and Giacomin (2007) based on the notion of strongly defended of arguments, and later in (Caminada, 2014) without having to refer to strong defence of arguments. In this section we present semi-stable semantics and strongly admissible semantics of AFs with further explanations, since we introduce these two notions for ADFs in our work, Chapters 3 and 5.

Based on extension semantics of AFs to reach a coherent conclusion it is not rational to choose a set of argument that are conflicting with one another in an extension. Furthermore, we do not accept an argument only because it exists but it is necessary that the argument is defended against counterarguments. Thus, the semantics of AFs are typically defined based on two important concepts, namely conflict-freeness and admissibility. Intuitively, conflict-freeness states that if there is a conflict between two arguments, then cannot be jointly accepted. Furthermore, the basic concept of admissibility specifies two main factors 1. arguments within the set do not attack one another, i.e., conflict-freeness, and 2. accepted arguments must defend themselves against attacks. In Definition 2.10 we present the notion of conflict-freeness and in Definition 2.11 we present the notion of an argument defended by a set.

**Definition 2.10** *Let $F = (A, R)$ be an AF. The set $S \subseteq A$ is a conflict-free set (extension) in $F$ if there is no $a, b \in S$ such that $(a, b) \in R$. The set of conflict-free sets of $F$ is denoted by $cf(F)$.*

Let us consider the AF, presented in Example 2.9. It holds that $cf(F) = \{\{\}, \{a\}, \{b\}, \{d\}, \{a, d\}\}$. Note that $c$ does not belong to any conflict-free extensions because of self attack over $c$.

**Definition 2.11** *Let $F = (A, R)$ be an AF. An argument $a \in A$ is defended by a set $S \subseteq A$ of arguments (alternatively, we say that $a$ is acceptable with respect to $S$ or $a$ is justifiable with respect $S$)(in $F$) if for each argument $c \in A$, it holds that if $(c, a) \in R$ then there is a $s \in S$ such that $(s, c) \in R$ ($s$ is called a defender of $a$).*

[2]In the literature, conflict-freeness and admissibility are often regarded as properties rather than semantics. We will use the properties, but at the same time treat them as semantics, i.e., conflict-free and admissible extensions.

Intuitively, a set $S$ defends an argument $a$ if $S$ can argue successfully against attacking arguments of $a$. Consider again our AF from Example 2.9. In this AF argument $d$ is defended by set $S = \{b\}$, since $b \in S$ attacks the attacker of $d$, namely $c$. Thus, $d$ is acceptable with respect to $S$. Furthermore, argument $a$ is defended by $S = \{a\}$, since there is only one attack over $a$ from $b$ and $a \in S$ attacks $b$ as well. Thus, $a$ is acceptable with respect to $S = \{a\}$. On the other hand, argument $b$ is not defended by set $S = \{b\}$. Although, it holds that $S$ defends argument $b$ against the attack from $a$, argument $b$ is not defended against the attack from $d$ by $S$. Thus, $b$ is not acceptable with respect to $S$.

In AFs the idea of collecting of arguments that can defend themselves against counterarguments is captured in the notion of an admissible set (extension). An admissible set is a conflict-free set of arguments, where each argument in the set is defended by the set, formally presented in Definition 2.12.

**Definition 2.12** *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is an admissible set in $F$ if*

- *$S$ is conflict-free in $F$; and*
- *each $s \in S$ is defended by $S$ in $F$.*

*The set of all admissible extensions of $F$ is denoted by $adm(F)$.*

Intuitively speaking, by choosing an admissible set in a given AF we disagree/reject all the attackers over the arguments of this set. Since this admissible set can defend all of its arguments against the attackers. In AF of Example 2.9, it holds that $adm(F) = \{\{\}, \{a\}, \{a, d\}\}$. Set $S = \{a, d\}$ is an admissible extension in $F$, since 1. it is conflict-free, 2. $S$ is defended all of its arguments, i.e., $a$ is defended against the attack from $b$ by $a$, and $d$ is defended against the attack from $c$ by $a$. Note that set $\{d\}$ is a conflict-free set in $F$ but it is not an admissible extension in $F$, since it is not defended against the attack from $c$ by $\{d\}$. Also, set $\{b\}$ is not an admissible extension in $F$, since $b$ is not defended against the attack from $d$ by $\{b\}$.

In AFs, several kinds of admissible extension are distinguished, expressing different point of views on accepting arguments together. In Definition 2.13 we present the notion of the characteristic operator over sets of arguments. The characteristic operator is a function that takes a set of arguments of an AF as an input and returns the set of arguments

that are defended by that set. Then, we define semantics of AFs via the characteristic operator.

**Definition 2.13** *Let $F = (A, R)$ be an AF. The characteristic operator, denoted by $\Gamma_F$, is defined over the power set of $A$, as follows. Let $S \subseteq A$. $\Gamma_F(S) = \{a \in A \mid a \text{ is defended by } S\}$.*

If we consider for an AF $F = (A, R)$ the partial order over the power set of $A$ with respect to set inclusion, i.e., $(2^A, \subseteq)$, then $\Gamma_F$ is a monotonic function with respect to $\subseteq$-ordering (subset-ordering). That is, if $S, S' \subseteq A$ with $S \subseteq S'$, then $\Gamma_F(S) \subseteq \Gamma_F(S')$. Most of the semantics of Dung's framework can be given via the characteristic operator, namely admissible, preferred, complete and grounded semantics. These semantics are then certain fixed points of the characteristic function. In Definition 2.14 we have the definitions of semantics of AFs.

**Definition 2.14** *Let $F = (A, R)$ be an AF. A set $S \in cf(F)$ is*

- *admissible in $F$ if $S \subseteq \Gamma_F(S)$;*
- *preferred in $F$ if $S = \Gamma_F(S)$ and there is no $T \in cf(F)$ with $S \subset T$ and $T = \Gamma_F(T)$, in other words, $S$ is $\subseteq$-maximal admissible;*
- *complete in $F$ if $S = \Gamma_F(S)$;*
- *grounded in $F$ if $S$ is the $\subseteq$-least fixed point of $\Gamma_F$;*
- *stable in $F$ if $\forall a \in A \setminus S$: $\exists b \in S$ s.t. $(b, a) \in R$;*
- *ideal in $F$ if it is a maximal admissible extension included in each preferred extension, presented in (Dung et al., 2007).*

We refer to the set of all preferred, complete, grounded, ideal, and stable extensions of AF $F$ as $prf(F)$, $com(F)$, $grd(F)$, $idl(F)$, and $stb(F)$, respectively. Note that an admissible set contains only arguments that are defended by that set. In addition, a preferred extension represents maximum information about arguments without losing admissibility. An extension being preferred means that after adding any additional argument to the set, the set is no longer admissible. Preferred semantics present a way to solve as many conflicts as possible among arguments.

An interpretation is complete if it exactly contains justifiable arguments. The grounded extension can be constructed by choosing unattacked

arguments and each argument that can be iteratively defended by these unattacked arguments. More informally speaking, the grounded extension collects the set of all arguments that are beyond any doubt. Stable semantics reflect the 'black-and-white' character of classical logic in AFs, that an argument is either accepted or rejected, i.e., each argument which is not in a stable extension is attacked by an accepted argument. An ideal extension collects arguments which are in all preferred extensions without losing of admissibility. Ideal semantics is a skeptical point of view of collecting arguments which are accepted via preferred point of view of semantics.

**Example 2.15** *Considering again AF of Example 2.9, depicted in Figure 2.1. As we presented earlier* $adm(F) = \{\{\}, \{a\}, \{a, d\}\}$. *It holds that the empty set is an admissible extension, since it is conflict-free and it does not contain any argument to defend. Plus, set* $\{a\}$ *is also an admissible extension in* $F$*, since it is conflict-free and* $a$ *defends itself against an attack from* $b$*. Furthermore, set* $\{a, d\}$ *is admissible, since* $d$ *is also defended by* $a$ *against an attack from* $c$*. The only preferred extension in* $F$ *is* $\{a, d\}$*, since it a maximal admissible extension with respect to set inclusion. Moreover,* $com(F) = \{\{\}, \{a, d\}\}$*, since* $\Gamma_F(\{\}) = \{\}$ *and* $\Gamma_F(\{a, d\}) = \{a, d\}$. *Note that the admissible set* $\{a\}$ *is not a complete extension since* $a$ *defends* $d$*, but* $d$ *does not belong to* $\{a\}$*, i.e.,* $\Gamma(\{a\}) = \{a, d\}$. *The unique grounded extension of this AF is the empty set. Further, since* $prf(F) = \{\{a, d\}\}$*, the ideal extension in* $F$ *is also* $\{a, d\}$. *Moreover, set* $\{a, d\}$ *is a stable extension in* $F$ *since this set attacks all the arguments that does not belong to this set.*

Theorem 2.16 presents the main relations among semantics of AFs, presented in Definition 2.14. Theorem 2.16 also shows distinctions among the semantics of AFs.

**Theorem 2.16** *(Dung, 1995) Let* $F$ *be an AF. The following properties hold in* $F$*.*

- *Each stable extension is a preferred extension, but not vice versa.*
- *Each preferred extension is a complete extension, but not vice versa.*
- *Each complete extension is an admissible extension, but not vice versa.*
- *Each admissible extension is a conflict-free extension (set), but not vice versa.*

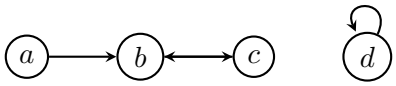


**Figure 2.2:** AF of Example 2.18 that does not have any stable extensions

- *Each ideal extension is an admissible extension, but not vice versa.*
- *Each grounded extension is a complete extension, but not vice versa.*

Note that stable extensions do not always exist, thus other semantics may be used for resolving conflicts. One of the crucial lemmas in AFs proven by Dung in 1995 is the Fundamental Lemma, presented in Lemma 2.17.

**Lemma 2.17** *Let $F = (A, R)$ be an AF and $S \subseteq A$ be an admissible set of arguments for $F$, and $a$ and $a'$ be arguments that are acceptable with respect to $S$ in $F$. Then,*

- *$S' = S \cup \{a\}$ is an admissible extension for $F$, and*
- *$a'$ is acceptable with respect to $S'$ in $F$.*

### Semi-stable semantics

A basic property of semantics of AFs is that each AF has at least one admissible, preferred, complete extension. This follows from the fact that the empty set is an admissible extension in any AF. Furthermore, each AF has a unique grounded extension and a unique ideal extension. However, it is possible that an AF does not have any stable extension. In Example 2.18 we present an instance of AF that does not have any stable extension.

**Example 2.18** *Let $F = (\{a, b, c, d\}, \{(a, b), (b, c), (c, b), (d, d)\})$ be an AF, depicted in Figure 2.2. It holds that $adm(F) = \{\{\}, \{a\}, \{c\}, \{a, c\}\}$, and $prf(F) = grd(F) = com(F) = idl(F) = \{\{a, c\}\}$. However, $F$ does not have any stable extension. Since by the definition of stable semantics a conflict-free set is a stable extension if it attacks all the arguments that do not belong to this set. In $F$ argument $d$ attacks itself, thus it cannot belong to any conflict-free sets and stable extensions, as well. Further, no argument different from $d$ attacks $d$, that is none of the conflict-free subsets of arguments of $F$ satisfies the second condition of stable extensions. Thus, $F$ does not have any stable extensions.*

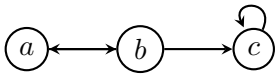


**Figure 2.3:** AF of Example 2.19

Because stable extensions do not always exist, in order to pick at least a set of arguments, preferred and grounded semantics have become popular in argumentation. In contrast, stable semantics still enjoys a strong support in logic programming (Gelfond and Lifschitz, 1988) and answer set programming (Gelfond and Lifschitz, 1991), since it can be preferred to have no outcome as opposed to an imperfect one. On the one hand, in argumentation a grounded extension presents the least amount of information about the acceptance of arguments. That is, a grounded extension collects a set of arguments about which there is no doubt. In other words, the grounded extension of a given AF is too skeptical. On the other hand, it is possible that an AF has a stable extension but the set of preferred extensions and stable extensions are not equal. Example 2.19, presents an AF that contains a stable extension but the set of stable extensions and preferred extensions do not coincide. Furthermore, the grounded extension in AF of Example 2.19 is a subset of a stable extension, i.e., $stb(F) \neq grd(F)$.

**Example 2.19** *Let* $F = (\{a, b, c\}, \{(a, b), (b, a), (b, c), (c, c)\})$ *be an AF, depicted in Figure 2.3 It holds that* $prf(F) = \{\{a\}, \{b\}\}$*, however, the only stable extension of* $F$ *is* $\{b\}$ *and the unique grounded extension of* $F$ *is* $\{\}$*.* $F$ *is an instance of AF that contains a stable extension, however, the set of preferred extensions of* $F$ *is not equal to the set of stable extensions of* $F$*. Furthermore, the grounded extension of* $F$ *is a subset of the stable extension of* $F$*.*

To overcome the possibility of non-existence of any stable extension, the semi-stable semantics has been proposed in AFs. Semi-stable semantics is a way of approximating stable semantics when a given AF does not have any stable extension. A key characteristic of the semi-stable semantics is that if an AF has a stable extension, then the semi-stable semantics coincides with the stable semantics. In contrast, this property does not hold for preferred and grounded semantics, as it was shown in Example 2.19.

The notion of semi-stable semantics for AFs was first introduced in (Verheij, 1996) (under a different name) then further investigated in (Caminada, 2006). In Definition 2.20, we recall the definition of semi-stable

semantics of AFs, as it is presented in (Caminada, 2006).

**Definition 2.20** *(Caminada, 2006) Let $F = (A, R)$ be an argumentation framework and let $S$ be an extension of $F$. For $a \in A$, we write $a^+ = \{b \mid (a, b) \in R\}$ and $S^+ = \cup\{a^+ \mid a \in S\}$. Set $S$ is called a semi-stable extension iff $S$ is a complete extension and $S \cup S^+$ is maximal among complete extensions of $F$, with respect to set inclusion. That is, set $S$ is a semi-stable extension of $F$ iff $S \in com(F)$ and for all $S' \in com(F)$ such that $S' \neq S$, we have that $S \cup S^+ \not\subseteq S' \cup S'^+$.*

The set of semi-stable extensions of $F$ is denoted by *semi-stb*$(F)$ in this work. Note that in Definition 2.20, $S^+$ collects all the arguments of $A$ that are attacked by $S$. Some alternative definitions of semi-stable semantics of AFs are also presented in (Caminada, 2006), as follows.

- An extension $S$ of $F$ is semi-stable if $S$ is a preferred extension where $S \cup S^+$ is maximal among all preferred extensions of $F$, with respect to set inclusion. That is, $S \in$ *semi-stb*$(F)$ iff $S \in$ *prf*$(F)$ and for all $S' \in$ *prf*$(F)$ such that $S' \neq S$, we have that $S \cup S^+ \not\subseteq S' \cup S'^+$.

or alternatively,

- An extension $S$ of $F$ is semi-stable if $S$ is an admissible extension where $S \cup S^+$ is maximal among all admissible extensions of $F$, with respect to set inclusion. That is, $S \in$ *semi-stb*$(F)$ iff $S \in$ *adm*$(F)$ and for all $S' \in$ *adm*$(F)$ such that $S' \neq S$, we have that $S \cup S^+ \not\subseteq S' \cup S'^+$.

The intuition of a semi-stable extension is that it maximizes the set of evaluated arguments, that is, the set of arguments that have been accepted or rejected, without losing admissibility. Key characteristics of semi-stable semantics in AFs are presented in Theorem 2.21. This theorem says that semi-stable semantics are placed between stable semantics and preferred semantics. It also holds that if an AF has at least one stable extension, then the set of stable extensions and semi-stable extensions coincide, and that each finite AF has at least one semi-stable extension.[3]

**Theorem 2.21** *Let $F = (A, R)$ be an AF, and let $S$ be an subset of $A$.*

- *if $S \in$ semi-stb$(F)$, then $S \in$ prf$(F)$;*

[3](Verheij, 2003b, Example 5.8) shows that existence is not guaranteed for infinite AFs. See also (Caminada and Verheij, 2010).

- *if* $S \in stb(F)$*, then* $S \in semi\text{-}stb(F)$*;*
- *if* $stb(F) \neq \emptyset$*, then* $stb(F) = semi\text{-}stb(F)$*;*
- *if* $|A|$ *is finite, then* $semi\text{-}stb(F) \neq \emptyset$*.*

Consider the AF of Example 2.18, as we have shown that AF does not have any stable extension. The reason is an isolated-self-attack argument $d$. An argument is called an isolated-self-attack argument if it has no parent and no child except itself, and it has a self-attack. Semi-stable semantics of AFs can be used as a remedy in the cases that there is no stable extension. The semi-stable extension of AF of Example 2.18 is $\{a, c\}$.

**Strongly admissible semantics**

All the semantics introduced until now may have multiple extensions, except the grounded semantics and the ideal semantics. Among all semantics the grounded semantics has a specific popularity, some reasons of which are as follows. 1. Each AF has a unique grounded extension. 2. The elements of the grounded extension usually belong to other semantics of AFs. Specifically, the grounded extension is the least complete extension. 3. The grounded extension collects all unattacked (undoubted) arguments and each argument that can be iteratively supported by these unattacked arguments. Thus, all things considered, no one has any doubt on the acceptance of the arguments that are in the grounded extension. Then, an important reasoning task for the grounded semantics is to verify whether a queried argument is part of the grounded extension.

However, if an agent asks 'Why do you think that no one has any doubt on the acceptance of a specific argument in an AF?' The only answer is that the queried argument belongs to the grounded extension of that AF. But the grounded semantics does not have any explanation for that, i.e. why a queried argument has to be accepted by anyone. Furthermore, to answer this query not all arguments within the grounded extension are necessary. That is, there is no need of constructing the grounded extension to answer the query. To handle this issue the notion of strong admissibility semantics of AFs has been introduced. A strongly admissible extension in an AF explains why an argument is acceptable without any doubt, without presenting all arguments in the grounded extension.

In AFs the concept of strong admissibility semantics has first been defined in the work of Baroni and Giacomin (2007), based on the notion of strong defence. Later in (Caminada, 2014) this concept was introduced

without referring to strong defence. Further, in 2019 Caminada and Dunne presented a labelling account of strong admissibility to answer the credulous decision problem of AFs under grounded semantics. In the following we present the notion of strong admissibility semantics of AFs based on the notion of strong defence, as it is presented in (Baroni and Giacomin, 2007).

**Definition 2.22** *(Baroni and Giacomin, 2007) Given an argumentation framework $F = (A, R)$, $a \in A$ and $S \subseteq A$, it is said that $a$ is* strongly defended *by $S$ if and only if each attacker $c \in A$ of $a$ is attacked by some $s \in S \setminus \{a\}$ such that $s$ is strongly defended by $S \setminus \{a\}$.*

Note that this definition is well-defined since the number of arguments in $F$ is finite and in each step an argument is excluded from the set of ancestors of a given argument. In Example 2.18, depicted in Figure 2.2, argument $c$ is strongly defended by set $S = \{a, c\}$, since the attacker of $c$, namely $b$ is attacked by $a \in S \setminus \{c\}$ and $a$ is strongly defended by $S \setminus \{c\}$. Actually, $a$ is strongly defended by $S = \emptyset$, since $a$ is not attacked by any argument. Note that in this example, although $c$ is defended by $S = \{c\}$, it is not strongly defended by $S = \{c\}$. Because there is no argument except $c$ in $S$ that defends $c$ against the attack of $b$.

**Definition 2.23** *Given an AF $F = (A, R)$ and set $S \subseteq A$, it is said that $S$ is a* strongly admissible *extension of $F$ if every $s \in S$ is strongly defended by $S$.*

The set of strongly admissible extensions of $F$ is denoted by $sadm(F)$ in this work. In Example 2.18, sets $S_1 = \emptyset$, $S_2 = \{a\}$, and $S_3 = \{a, c\}$ are strongly admissible extensions of $F$; all of them are subsets of the grounded extension of $F$. However, set $S' = \{c\}$ is not a strongly admissible extension of $F$. Since, $c \in S'$ is not strongly defended by $S'$. Because argument $c$ is attacked by $b$, however, no argument in $S' \setminus \{c\}$ attacks $b$. Thus, although $S' = \{c\}$ is an admissible extension of $F$, it is not a strongly admissible extension of $F$.

In (Caminada and Dunne, 2019), the concept of strongly admissible semantics of AFs are defined without having to refer to strong defence; we rephrase it in Definition 2.24.

**Definition 2.24** *(Caminada and Dunne, 2019) Let $F = (A, R)$ be an argumentation framework. We say that $S \subseteq A$ is a* strongly admissible *extension of $F$ if and only if every $a \in S$ is defended by some $S' \subseteq S \setminus \{a\}$ which in its turn is a strongly admissible extension.*

It is shown in (Baroni and Giacomin, 2007) that strongly admissible extensions of an AF forms a lattice with the empty set as the least element and the grounded extension as the maximum element; we recall it in Theorem 2.25.

**Theorem 2.25** *(Baroni and Giacomin, 2007) Let $F$ be an AF. The set of strongly admissible extensions of $F$ forms a lattice with the empty set as the least element and the grounded extension as the maximum element.*

Theorem 2.25 presents a significant result that leads to the distinction between strongly admissible semantics and admissible and complete semantics of AFs. In (Dung, 1995) it is shown that admissible extensions of a given AF form a meet-semilattice with the empty set as the least element and the preferred extensions as its maximal elements. Furthermore, it is shown in (Dung, 1995) that the complete extensions of a given AF form a complete meet-semilattice with the grounded extension as its least element and the preferred extensions as its maximal elements. In contrast, the strongly admissible extensions form a lattice with the empty set as the least element and the grounded extension as the maximum element. These different lattices show a distinction between strongly admissible semantics and admissible/complete semantics of AFs.

Finally, in (Caminada and Dunne, 2019) and (Baroni and Giacomin, 2007), the relation between strongly admissible semantics of an AF and its admissible, conflict-free and grounded semantics is clarified; let us recall the properties in Proposition 2.26.

**Proposition 2.26** *(Caminada and Dunne, 2019; Baroni and Giacomin, 2007) Let $F$ be an AF. The following properties hold:*

- *Each strongly admissible extension in $F$ is an admissible extension, however, the other direction does not hold.*

- *Each strongly admissible extension of $F$ is an admissible extension and it is a subset of the grounded extension of $F$, however, the other direction does not hold. That is, AF $F$ may have an admissible extension that is a subset of the grounded extension of $F$ but that is not a strongly admissible extension of $F$.*

**Proof** Let $F$ be an AF.

- In *Theorem* 4 in (Caminada and Dunne, 2019) it has been proved that each strongly admissible extension in $F$ is an admissible extension. We provide a proof that the other direction does not hold, by giving a counter-example. Let $F = (\{a, b\}, \{(a, b), (b, a)\})$. Now the set $S = \{a\}$ is a conflict-free and admissible extension of $F$, however, it is not a strongly admissible extension in $F$. This is because there is no $S' \subseteq S \setminus \{a\}$ that defends $a$ against the attack of $b$.

- It has been proved in (Baroni and Giacomin, 2007) that the strongly admissible semantics of an AF form a lattice with the grounded extension as the maximum element. However, it does not hold that any admissible extension of a given AF that is a subset of the grounded extension is a strongly admissible extension. We provide a proof by giving a counter-example. Let $F = (\{a, b\}, \{(a, b), (b, c), (c, d)\})$. The grounded extension of $F$ is $\{a, c\}$; furthermore, $S = \{c\}$ is an admissible extension of $F$. However, $S$ is not a strongly admissible extension of $F$, since there is no $S' \subseteq S \setminus \{c\}$ that defends $c$ against the attack of $b$.

□

In the following we present the distinction between strongly admissible semantics of AFs and ideal semantics of AFs. The notion of the ideal extension is presented in Definition 2.14. In Definition 3.48 in (Baroni and Giacomin, 2007) first the notion of ideal set is defined, then the notion of ideal semantics is presented; we rephrase this definition here. An admissible extension $S$ is called *ideal* set iff it is a subset of each preferred extension of $F$. The ideal extension of $F$ is a maximal (with respect to set-inclusion) ideal set. Note that an ideal set is not necessarily an ideal extension. We show that the strongly admissible extensions differ from the ideal sets and the ideal extension of a given AF.

**Proposition 2.27** *The notion of strongly admissible semantics of AFs differs from the notion of ideal semantics of AF.*

**Proof** We provide a proof by an example. Let $F = (\{a, b\}, \{(a, b), (b, a), (b, b)\})$, as depicted in Figure 2.4. The unique grounded extension of $F$ is the empty set. The set of strongly admissible extensions of $F$ is $\{\emptyset\}$. However, the set of ideal sets of $F$ is $\{\emptyset, \{a\}\}$. Thus, $\{a\}$ is the ideal extension of $F$. As we see, the set of strongly admissible extensions of $F$ is

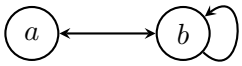


**Figure 2.4:** AF of Proposition 2.27

neither equal with the set of ideal sets of $F$ nor is it equal with the ideal extension of $F$.

□

The relations between the different semantics of AFs, presented in this section, are depicted in Figure 2.5, studied in (Baroni et al., 2011; Baroni and Giacomin, 2005, 2008; Caminada, 2007b; Caminada et al., 2012; Dung, 1995; Dung et al., 2007; Verheij, 1996). Let $\sigma, \gamma \in \{cf, adm, prf, com, grd, idl, stb, semi\text{-}stb, sadm\}$. Each box in Figure 2.5 presents $\sigma$ semantics. In this figure, for reasons of brevity in each box we remove 'semantics' from $\sigma$ semantics and we have only the name of the associated semantics in a box. For instance, instead of writing stable semantics we have stable in a box. We show a $\sigma$-extension is a $\gamma$-extension by drawing an arrow from a box contains $\sigma$ to a box contains $\gamma$. That is, an arrow from $\sigma$ box to $\gamma$ box means that $\sigma(F) \subseteq \gamma(F)$ for a given AF $F$. The red dashed line from strong admissibility semantics to grounded semantics means that any strongly admissible extension is a subset of the grounded extension.

**Proposition 2.28** *In accordance with Figure 2.5, for any AF $F$ the following relation holds among the semantics of AFs.*

- $stb(F) \subseteq semi\text{-}stb(F) \subseteq prf(F) \subseteq com(F) \subseteq adm(F) \subseteq cf(F)$;
- $grd(F) \subseteq com(F)$;
- $idl(F) \subseteq com(F)$;
- *each strongly admissible extension of $F$ is a subset of the grounded extension of $F$.*

### 2.3.2 Labelling-based Semantics of AFs

AF semantics can also be presented in terms of labelling functions, giving a fine-grained view of the acceptance status of arguments. Already Pollock (1995) used labeling functions for structured arguments, and Verheij (1996)

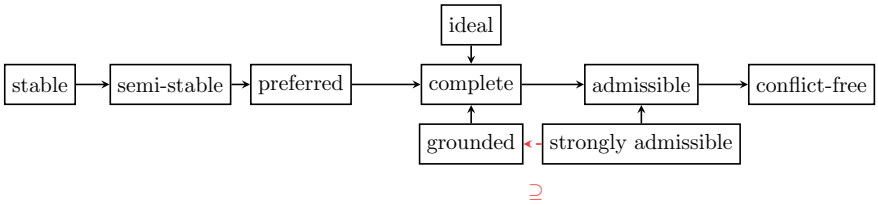


**Figure 2.5:** Relation among semantics of AFs

applied labeling to abstract argumentation, while introducing the stage and semi-stable extensions (using another name). Caminada (2006) investigated three-valued labelings using labels `in`, `out`, or `undec`. The concept of labelings for AFs has been used for argument acceptability/deniability (e.g., (Verheij, 2007)), and further labeling-based semantics have been presented by Caminada and Gabbay (2009). Here we focus on three-valued labellings of an AF, assigning to each argument either `in`, `out`, or `undec`. In this section we only present the notions of admissible semantics and strongly admissible semantics of AFs based on labelling, as they are the only ones needed for our work.

**Definition 2.29** *(Caminada and Dunne, 2019, Definition 4) Let* $F = (A, R)$ *be an argumentation framework. An* argument labelling *is a function* $\mathscr{L} : A \mapsto \{\texttt{in}, \texttt{out}, \texttt{undec}\}$*. An argument labelling is called an* admissible *labelling if and only if* $\mathscr{L}$ *is a total function and for each* $a \in A$ *it holds that:*

- *if* $\mathscr{L}(a) = \texttt{in}$*, then for each* $b$ *that attacks* $a$ *it holds that* $\mathscr{L}(b) = \texttt{out}$*,*
- *if* $\mathscr{L}(a) = \texttt{out}$*, then there exists a* $b$ *that attacks* $a$ *such that* $\mathscr{L}(b) = \texttt{in}$*.*

A labelling account of strong admissibility semantics of $F$ is presented in (Caminada and Dunne, 2019); we recall this definition in Definition 2.32, to use it in the proofs of theorems of Section 3.3 of Chapter 3, in order to show that strongly admissible semantics of ADFs form a generalization of strongly admissible semantics of AFs. To this end, let us first rephrase the concept of min-max numbering in Definition 2.30. [4] Note that, if

[4]Caminada and Dunne (Caminada and Dunne, 2019) describe the intuition behind the min-max number of an argument as follows: 'The game-theoretic length of the path (consisting of alternately `in` and `out` labelled arguments) from the argument back to an unattacked ancestor argument. The player selecting the `in` labelled arguments aims to make the path as short as possible whereas the player selecting the `out` labelled arguments aims to make the path as long as possible.'

$\mathscr{L}$ is a labelling, we write $\texttt{in}(\mathscr{L})$ for $\{a \in A \mid \mathscr{L}(a) = \texttt{in}\}$, $\texttt{out}(\mathscr{L})$ for $\{a \in A \mid \mathscr{L}(a) = \texttt{out}\}$, and $\texttt{undec}(\mathscr{L})$ for $\{a \in A \mid \mathscr{L}(a) = \texttt{undec}\}$.

Specifically, in Lemma 3.40 we use the notion of strongly admissible labelling of AFs, which is defined in terms of min-max numbering (Definition 2.32) to show that the map of each strongly admissible labelling in a given AF $F$ is a strongly admissible interpretation of the associated ADF $D_F$, and vice versa. That is, we show that there is a one-to-one relation between the set of strongly admissible labellings in a given AF $F$ and the set of strongly admissible interpretations of the associated ADF $D_F$.

**Definition 2.30** *Let $\mathscr{L}$ be an admissible labelling of argumentation framework $F = (A, R)$. A* min-max numbering *is a total function $\mathscr{MM}_{\mathscr{L}} : \texttt{in}(\mathscr{L}) \cup \texttt{out}(\mathscr{L}) \mapsto \mathbb{N} \cup \infty$ such that for each $a \in \texttt{in}(\mathscr{L}) \cup \texttt{out}(\mathscr{L})$ it holds that:*

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

- *if $\mathscr{L}(a) = \texttt{out}$, then*
  *$\mathscr{MM}_{\mathscr{L}}(a) = min(\{\mathscr{MM}_{\mathscr{L}}(b) \mid (b, a) \in R$ and $\mathscr{L}(b) = \texttt{in}\}) + 1$ (with $min(\emptyset)$ defined as $\infty$).*

**Theorem 2.31** *(Caminada and Dunne, 2019, Theorem 6) Every admissible labelling has a unique min-max numbering.*

**Definition 2.32** *(Caminada and Dunne, 2019, Definition 10) A strongly admissible labelling is an admissible labelling whose min-max numbering yields natural numbers only (so no argument is numbered $\infty$).*

**Example 2.33** *Let $F = (\{a, b, c, d\}, \{(a, b), (c, d), (d, c)\})$. By Definition 2.30, admissible labelling $\{a \mapsto \texttt{in}, b \mapsto \texttt{out}, c \mapsto \texttt{in}, d \mapsto \texttt{out}\}$ has a unique min-max numbering $\{(a : 1), (b : 2), (c : \infty), (d : \infty)\}$. However, this admissible labelling is not a strongly admissible labelling in $F$, since the $\mathscr{MM}_{\mathscr{L}}(c) = \mathscr{MM}_{\mathscr{L}}(d) = \infty$. On the other hand, the admissible labelling $\{a \mapsto \texttt{in}, b \mapsto \texttt{out}, c \mapsto \texttt{undec}, d \mapsto \texttt{undec}\}$ has a unique min-max numbering $\{(a : 1), (b : 2)\}$, since both $\mathscr{MM}_{\mathscr{L}}(a)$ and $\mathscr{MM}_{\mathscr{L}}(b)$ are finite, so by Definition 2.32, it is a strongly admissible interpretation in $F$.*

In (Baroni et al., 2018a), both extension-based and labelling-based approaches of semantics of AFs are presented. Moreover, two functions are

defined to map extension-based semantics of a given AF $F$ to labelling-based semantics and vice versa, to show that each extension-based semantics of AF has a labelling-based semantics reformulation and vice versa. Namely, *Ext2Lab*($\lambda$) is used to present the extension form of labelling $\lambda$ of $F$, and *Lab2Ext*($e$) is used to present the labelling form of the extension $e$ of $F$; we recall them in Definitions 2.34–2.35.

**Definition 2.34** *(Baroni et al., 2018a, Definition 3.6) Let $F = (A, R)$ and let $S$ be an extension of $F$. For $a \in A$, we write $a^+ = \{b \mid (a, b) \in R\}$ and $S^+ = \cup\{a^+ \mid a \in S\}$. If $S$ is a conflict-free set of $F$, then the corresponding labelling is defined as Ext2Lab$(S) = \{S, S^+, A \setminus (S \cup S^+)\}$.*

Function *Ext2Lab*(.) in Definition 2.34 is such that arguments of $S$ are labelled `in`, elements of $S^+$ are labelled `out` and all other arguments of $A$ are labelled `undec`. The alternative way of presenting *Ext2Lab*(.) is as follows.

$$\textit{Ext2Lab}(S)(a) = \begin{cases} \texttt{in} & \text{if } a \in S, \\ \texttt{out} & \text{if } a \in S^+, \\ \texttt{undec} & \text{otherwise.} \end{cases}$$

**Definition 2.35** *(Baroni et al., 2018a, Definition 2.7)) Given an argumentation framework $F = (A, R)$ and a labelling $\lambda$, the corresponding set of arguments Lab2Ext($\lambda$) is defined as Lab2Ext($\lambda$) = $\texttt{in}(\lambda)$. That is, Lab2Ext($\lambda$) is the set of all arguments that are labelled `in` in $\lambda$.*

**Proposition 2.36** *(Baroni et al., 2018a, Proposition 3.14) For any argumentation framework $F = (A, R)$, it holds that:*

- *if $S$ is a strongly admissible set of $F$, then Ext2Lab($S$) is a strongly admissible labelling of $F$;*
- *if $\lambda$ is a strongly admissible labelling of $F$, then Lab2Ext($\lambda$) is a strongly admissible set of $F$.*

## 2.4 SETAFs: Argumentation Frameworks with Collective Attacks

Since abstract argumentation frameworks have been introduced by Dung 1995 as a core formalism in formal argumentation, a popular line of research investigates extensions of Dung AFs that allow for a richer syntax (see,

e.g. (Brewka et al., 2014)). A generalisation of Dung AFs that allow for a more flexible attack structure, but do not consider support between arguments, are SETAFs as introduced by Nielsen and Parsons 2006. SETAFs extend Dung AFs by allowing for collective attacks that a set of arguments $B$ attacks another argument $a$ but no proper subset of $B$ attacks $a$, as it is presented formally in Definition 2.37.

Here we formally present the notion SETAFs as a generalization of AFs, since we work on this class in Chapters 8 and 9.

**Definition 2.37** *A set argumentation framework(SETAF) is an ordered pair $F = (A, R)$, where $A$ is a finite set of arguments and $R \subseteq (2^A \setminus \{\emptyset\}) \times A$ is the attack relation.*

Given a SETAF $(A, R)$, we write $S \mapsto_R b$ if there is a set $S' \subseteq S$ attacking b, i.e. $(S', b) \in R$. We say that in this case also $S$ attacks b. Moreover, we write $S' \mapsto_R S$ if $S' \mapsto_R b$ for some $b \in S$. We drop the subscript in $\mapsto_R$ if the attack relation is clear from the context.

Notions of conflict and defense can be defined for SETAFs in analogy to these notions in the context of AFs. Given a SETAF $F = (A, R)$, a set $S \subseteq A$ is *conflicting* in $F$ if $S \mapsto_R S$; $S \subseteq A$ is *conflict-free* in $F$, if $S$ is not conflicting in $F$, i.e. if $S' \cup \{a\} \not\subseteq S$ for each $(S', a) \in R$. An argument $a \in A$ is *defended* (in $F$) by a set $S \subseteq A$ if for each $B \subseteq A$, such that $B \mapsto_R a$, also $S \mapsto_R B$. A set $T$ of arguments is hence defended (in $F$) by $S$ if each $a \in T$ is defended by $S$ (in $F$).

The semantics of SETAFs can now also be defined similarly to AFs via a characteristic operator. With a slight abuse of notation (because of the use of the same notation for the characteristic operator), we thus define first of all also for a SETAF $F = (A, R)$, $\Gamma_F(S) = \{a \in A \mid a \text{ is defended by } S \text{ in } F\}$; here the notion of "defense" clearly being that defined for SETAFs. For completeness we detail the definitions of all semantics we consider in this work for SETAFs, although the definitions are exactly as those for AFs (modulo the use of the more general notions of attack and that the characteristic operator referenced therein is the characteristic operator defined for SETAFs):

**Definition 2.38** *Let $F = (A, R)$ be a SETAF. A set $S$ which is conflict-free in $F$ is*

- admissible *in $F$ iff $S \subseteq \Gamma_F(S)$;*
- complete *in $F$ iff $S = \Gamma_F(S)$;*

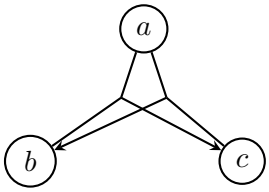


**Figure 2.6:** The SETAF of Example 2.39.

- grounded *in $F$ iff $S$ is the $\subseteq$-least fixed-point of $\Gamma_F$;*
- preferred *in $F$ iff $S$ is $\subseteq$-maximal admissible (resp. complete) in $F$;*
- stable *in $F$ iff for all $a \in A \setminus S$, $S$ attacks $a$.*

We will use the same abbreviations for SETAFs as for AFs for denoting the sets of arguments obtained when applying the semantics on SETAFs. We also recall that several important properties of Dung AFs carry over to SETAFs; we refer to (Nielsen and Parsons, 2006; Flouris and Bikakis, 2019) for details.

In the following we provide an example of a SETAF and also illustrate the concept of extensions and semantics for SETAFs.

**Example 2.39** *The SETAF $F = (\{a, b, c\}, \{(\{a, b\}, c), (\{a, c\}, b)\})$ is depicted in Figure 2.6. In $F$, $(\{a, b\}, c) \in R$ says that there is a joint attack from $a$ and $b$ to $c$, and $(\{a, c\}, b) \in R$ says that there is a joint attack from $a$ and $c$ to $b$. The former attack represents that neither $a$ nor $b$ are strong enough to attack $c$ by themselves. The latter attack indicates that neither $a$ nor $c$ are strong enough to attack $b$ by themselves. The conflict-free extensions of $F$ are $cf(F) = \{\{\}, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$, the admissible extensions $adm(F) = \{\{\}, \{a\}, \{a, b\}, \{a, c\}\}$, the complete extensions $com(F) = \{\{a\}, \{a, b\}, \{a, c\}\}$, the unique grounded extension $grd(F) = \{\{a\}\}$, and the preferred extensions $prf(F) = stb(F) = \{\{a, b\}, \{a, c\}\}$. Note that, for instance, $\{b, c\}$ is a conflict-free extension. However, it is not an admissible extension, since $\{b, c\} \not\subseteq (\Gamma_F(\{b, c\}) = \{\})$. Further, $\{a\}$ is an admissible and a complete extension, since $\Gamma_F(\{a\}) = \{a\}$. On the other hand $\{a\}$ is not a preferred extension because it is not a $\subseteq$-maximal admissible extension.*

SETAFs have received increasing interest in the last years. For instance, semi-stable, stage, ideal, and eager semantics have been adapted to SETAFs in (Dvořák et al., 2019; Flouris and Bikakis, 2019); translations between

SETAFs and other abstract argumentation formalisms are studied in (Polberg, 2017); (Yun et al., 2018) observed that for particular instantiations, SETAFs provide a more convenient target formalism than Dung AFs.

## 2.5 Abstract Dialectical Frameworks

While being popular and simple, AFs can only beused to model argumentation contexts with simple attack relations among arguments, depicted by directed edges in the associated graph. Thus, there exist a number of generalizations for AFs, as many researchers felt a need to cover additional relevant relationships among arguments. For instance, to model group attacks among arguments (Nielsen and Parsons, 2006), or to model preference over the arguments (Bench-Capon, 2003; Bench-Capon and Atkinson, 2009), or to model support relation among arguments (Cayrol and Lagasquie-Schiex, 2009), or to model nested support and attack (Verheij, 2003b).

Abstract dialectical frameworks (ADFs) were first introduced in (Brewka and Woltran, 2010), and further refined in (Brewka et al., 2013, 2017a, 2018a) Among the generalizations of AFs, ADFs allow for a systematic and flexible generalization of AFs in which the logical relations among arguments can be represented. In particular, arguments can not only attack one another, but also support each other and interact in logically composite ways.

In AFs acceptance of an argument depends on the rejection of its attacker (parents). However, in ADFs any logical combination of accepted and denied parents may lead to the acceptance of the argument in question. This leads to the concept of acceptance condition of arguments, presented formally in Definition 2.40. Again as for AFs we assume that $\mathscr{P}$ which is a set of propositional variables or atoms is a universe of arguments.

**Definition 2.40** *An abstract dialectical framework (ADF) is a tuple* $F = (A, L, C)$ *where:*

- $A \subseteq \mathscr{P}$ *is a set of arguments (statements, positions), denoted by letters;*
- $L \subseteq A \times A$ *is a set of links among arguments;*
- $C = \{\varphi_a\}_{a \in A}$ *is a collection of propositional formulas over arguments, called acceptance conditions.*

**Remark 2.40.1** *ADF $D = (A, L, C)$ is called a* finite *abstract dialectical framework if $A$ is a finite set of arguments. Note that in this work, we assume that all ADFs are finite.*

An ADF can be represented by a graph in which nodes indicate arguments and links show the relation among arguments. Each argument $a$ in an ADF is labelled by a propositional formula, called acceptance condition, $\varphi_a$ over $par(a)$ where $par(a) = \{b \mid (b, a) \in L\}$. An argument $a$ is called an *initial argument* if $par(a) = \{\}$. The acceptance condition of each argument clarifies under which condition the argument can be accepted (Brewka and Woltran, 2010; Brewka et al., 2018a). Furthermore, acceptance conditions indicate the set of links implicitly. Thus, in a concrete example of ADFs, we oftentimes only define acceptance conditions explicitly and implicitly define links via the variables of the propositional formulas. That is, for the reason of brevity we avoid presenting the set of links of ADFs in our examples. Also, there is an alternative notion for ADFs which is more compact, which is also used in the literature (see e.g., (Pührer, 2020a)), presented in Definition 2.41.

**Definition 2.41** *An abstract dialectical framework (ADF) $D$ is a set of tuples $\{(a, \varphi_a)\}_{a \in A}$ where $A$ is the set of arguments and $\varphi_a$ is a propositional formula over $par(a)$, called the acceptance condition of $a$.*

**Definition 2.42** *Let $D = (A, L, C)$ be an ADF and let $a$ be an argument.*

- *$a$ is called an initial argument if $par(a) = \emptyset$;*
- *$a$ is called an isolated argument if it is an initial argument and it does not have any child, i.e., $a$ does not have any outgoing links.*

Note that if an argument $a$ is initial or isolated, then its acceptance condition $\phi_a$ must be either $\top$ or $\bot$.

**Example 2.43** *An example of an ADF $D = (A, L, C)$ is shown in Figure 2.7, which contains four arguments, i.e., $A = \{a, b, c, d\}$. Dependencies between arguments are shown by the directed edges in the associated graph, and acceptance conditions are shown as propositional formulas attached to each node. The acceptance condition of an argument clarifies the set of parents of the argument, thus, there is no need of presenting $L$ in $D$ explicitly. Furthermore, an acceptance condition of an argument indicates under which condition the argument can be accepted/denied. For instance,*

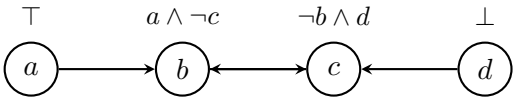


**Figure 2.7:** ADF of Examples 2.43

*the acceptance condition of c, namely $\varphi_c : \neg b \wedge d$, says that c depends on b and d, i.e., $par(c) = \{b, d\}$, and it states that c can be accepted if b is denied and d is accepted. In this ADF, the arguments a and d are initial arguments, since $par(a) = par(d) = \{\}$. The acceptance condition of a, namely $\varphi_a : \top$, says that a is always acceptable and the acceptance condition of d, namely $\varphi_d : \bot$, says that d is always deniable.*

### 2.5.1 Semantics of ADFs

In this section we present the semantics of ADFs. A more comprehensive account of ADF semantics and their origins in approximation fixpoint theory (Denecker et al., 2004) can be found in (Strass, 2013b). Since the acceptance condition of each argument in ADFs is a propositional formula, interpretations are proper tools to present the semantics of ADFs. Thus, semantics of ADFs are defined based on *three-valued interpretations*, presented in Section 2.1. A three-valued interpretation function assigns each proposition, namely argument in ADFs, to a *truth value*; true, false or undecided, which we denote with **t**, **f** and **u**, respectively.

Let $D = (A, L, C)$ be an ADF, a *three-valued interpretation* $v$ (for $D$) is a function $v : A \mapsto \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ that maps arguments to one of the three truth values. Furthermore, $v$ is called a two-valued interpretation if for each $a \in A$ either $v(a) = \mathbf{t}$ or $v(a) = \mathbf{f}$. Moreover, $v$ is called *trivial* interpretation (of $D$), and $v$ is denoted by $v_{\mathbf{u}}$, if $v(a) = \mathbf{u}$ for each $a \in A$. In the current work we say that the truth value of *a is presented* in $v$, if $v(a) \in \{\mathbf{t}, \mathbf{f}\}$. As it is presented in Section 2.1, interpretations can be ordered via $\leq_i$ with respect to their information content.

**Definition 2.44** *Let $\mathcal{V}$ be the set of all three-valued interpretations for an ADF $D = (A, L, C)$. We say for interpretations $v, w \in V$, interpretation $v$ extends $w$ ($v$ is an extension of $w$), if $w(a) \leq_i v(a)$ for each $a \in A$, denoted by $w \leq_i v$. Note that $\leq_i$ is a partial order, in particular, it is antisymmetric.*

- *Interpretations $v$ and $w$ are* incomparable *if neither $w \leq_i v$ nor $v \leq_i w$, denoted by $w \bowtie v$.*

- *The set of all two-valued interpretations that extend a given interpretation $v$ is denoted by $[v]_2$, i.e.,*

$$[v]_2 = \{v' \mid v' \text{ is a two-valued interpretation and } v \leq_i v'\}$$

*Each $w \in [v]_2$ is called a completion of $v$.*

For reasons of brevity, we will sometimes shorten the notion of three-valued interpretation $v = \{a_1 \mapsto t_1, \ldots, a_m \mapsto t_m\}$ with arguments $a_1, \ldots, a_m$ and truth values $t_1, \ldots, t_m$ as follows: $v = \{a_i \mid v(a_i) = \mathbf{t}\} \cup \{\neg a_i \mid v(a_i) = \mathbf{f}\}$. For instance, $v = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}\} = \{\neg a, b\}$.

Semantics for ADFs can be defined via the *characteristic operator* $\Gamma_D$ which maps interpretations to updated interpretations. Given an interpretation $v$ (for $D$), the partial valuation of $\varphi_a$ by $v$, is $v(\varphi_a) = \varphi_a^v = \varphi_a[b/\top : v(b) = \mathbf{t}][b/\bot : v(b) = \mathbf{f}]$, for $b \in par(a)$.

**Definition 2.45** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Applying $\Gamma_D$ on $v$ leads to $v'$ such that for each $a \in A$, $v'$ is as follows:*

$$v'(a) = \begin{cases} \mathbf{t} & \text{if } \varphi_a^v \text{ is irrefutable (i.e., } \varphi_a^v \text{ is a tautology),} \\ \mathbf{f} & \text{if } \varphi_a^v \text{ is unsatisfiable (i.e., } \varphi_a^v \text{ is a contradiction),} \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

*An argument $a$ is called justifiable with respect to interpretation $v$ if $v(a) \in \{\mathbf{t}, \mathbf{f}\}$ and $\Gamma_D(v)(a) = v(a)$.*

Intuitively, the characteristic operator $\Gamma_D$ assigns an argument $a$ to $\mathbf{t}$ ($\mathbf{f}$) if the partial evaluation of the acceptance condition of $a$, namely $\varphi_a$, by a given interpretation $v$ is irrefutable (unsatisfiable), respectively. In other words, the characteristic operator $\Gamma_D$ assigns an argument $a$ to $\mathbf{t}$ ($\mathbf{f}$) if all completions of the given interpretation $v$ satisfy (do not satisfy) the acceptance condition of $a$. Note that the operator $\Gamma_F$ is $\leq_i$-monotonic, that is, when $v \leq_i w$ for interpretations $v$ and $w$, then $\Gamma_D(v) \leq_i \Gamma_D(w)$. The idea of the proof is as follows; let $a$ be an argument such that $\Gamma_D(v)(a) = \mathbf{t}$. Then, by the definition of the characteristic operator, $\varphi_a^v$ is irrefutable. Since $v \leq_i w$, it holds that $\varphi_a^w$ is a tautology. Thus, $\Gamma_D(w)(a) = \mathbf{t}$. By the similar proof method it holds that if $\Gamma_D(v)(a) = \mathbf{f}$ for an $a$, then $\Gamma_D(w)(a) = \mathbf{f}$. Hence, $\Gamma_D(v) \leq_i \Gamma_D(w)$.

Example 2.46 illustrates the definition of the characteristic operator.

**Example 2.46** *Consider ADF $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$ from Example 2.43 and the trivial interpretation $v_{\mathbf{u}}$ of*

*D. We calculate* $v' = \Gamma_D(v_{\mathbf{u}})$ *over all the arguments of D. Consider argument a; since* $\varphi_a : \top$*, it holds that* $\varphi_a^{v_{\mathbf{u}}} : \top$*, that is,* $\varphi_a^{v_{\mathbf{u}}}$ *is irrefutable. Thus, a is assigned to* $\mathbf{t}$ *in* $v'$*. However, since both of the parents of b, namely, a and c are assigned to* $\mathbf{u}$ *in* $v_{\mathbf{u}}$*, it holds that* $\varphi_b^{v_{\mathbf{u}}} = \varphi_b$*, which is neither a tautology nor unsatisfiable. Thus,* $v'(b) = \Gamma_D(v_{\mathbf{u}})(b) = \mathbf{u}$*. By the same reason, it holds that* $v'(c) = \Gamma_D(v_{\mathbf{u}})(c) = \mathbf{u}$*. However, it holds that* $\varphi_d^{v_{\mathbf{u}}} : \bot$*, that is,* $\varphi_d^{v_{\mathbf{u}}}$ *is unsatisfiable. Thus,* $v'(d) = \Gamma_D(v_{\mathbf{u}})(d) = \mathbf{f}$*. Hence,* $v' = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, d \mapsto \mathbf{f}\}$*.*

*If we apply the characteristic operator* $\Gamma_D$ *over* $v'$*, then since* $\Gamma_D$ *is a monotonic operator, the truth values of a and d in* $\Gamma_D(v')$ *are equal to* $v'(a)$ *and* $v'(d)$*, respectively. Since* $\varphi_c^{v'} = \neg b \wedge \bot \equiv \bot$*, it holds that* $\Gamma_D(v')(c) = \mathbf{f}$*. However, since* $\varphi_b^{v'} = \top \wedge \neg c = \neg c$*, it holds that* $\Gamma_D(v')(b) = \mathbf{u}$*. Thus,* $\Gamma_D(v') = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$*.*

The operator-based semantics of ADFs can be routed back to the work of Denecker, Marek, and Truszczynski (2000; 2003; 2004) on approximation fix point theory (AFT) (for a detailed analysis of the relationship between ADFs and AFT see, e.g. (Strass, 2013b)). The characteristic operator for ADFs generalizes the characteristic function for AFs. The operator $\Gamma_D$ over interpretation $v$ returns, the consensus truth value of the evaluation of the acceptance formula $\varphi_a$ under each two-valued interpretation extending $v$. Intuitively, $\Gamma_D$ checks which truth values can be justified based on the information in $v$ and the acceptance conditions. The semantics of ADFs, as defined by (Brewka et al., 2018a), are based on (collections of) three-valued interpretations. The semantics of ADFs are defined via the characteristic operator as in Definition 2.47.

**Definition 2.47** *Given an ADF D, an interpretation v is:*

- *conflict-free iff* $v(s) = \mathbf{t}$ *implies* $\varphi_s^v$ *is satisfiable and* $v(s) = \mathbf{f}$ *implies* $\varphi_s^v$ *is unsatisfiable;*[5]
- *admissible in D iff* $v \leq_i \Gamma_D(v)$*;*
- *preferred in D iff v is* $\leq_i$*-maximal admissible;*
- *complete in D iff* $v = \Gamma_D(v)$*;*

[5]Note that the notion of conflict-free semantics for three-valued semantics, presented in Definition 2.47 is based on the given notion in (Strass, 2014; Gaggl et al., 2021). However, the notion of conflict-free semantics can be proposed as it is given in (Strass and Wallner, 2015), where an argument can be assigned to false if the partially evaluated acceptance condition is refutable.

- *two-valued model in $D$ iff $v = \Gamma_D(v)$ and $v$ is a two-valued interpretation;*

- *the grounded interpretation of $D$ iff $v$ is the least fixed point of $\Gamma_D$.*

The set of all $\sigma$ interpretations for an ADF $D$ is denoted by $\sigma(D)$, where $\sigma \in \{cf, adm, com, grd, prf, mod\}$ abbreviates the different semantics in the obvious manner. Conflict-free interpretations are defined by weakening a condition of admissible interpretations. However, this definition differs from admissibility by requiring satisfiability instead of a tautology. Thus, each admissible interpretation is a conflict-free interpretation, (based on the definition of conflict-freeness rewritten in this work).

Since the characteristic operator is monotonic, via the fix point theorem for monotone operators in complete partial orders (see, e.g. (Davey and Priestley, 2002, Theorem 8.22)) we have the existence of the least fix point of $\Gamma_D$, i.e., the grounded interpretation, in each ADF $D$. Similar to AFs, admissible interpretations of a given ADF form a semi-lattice with the trivial interpretation as the least element and the preferred interpretations as its maximum elements. Moreover, complete interpretations of a given ADF form a semi-lattice with the grounded interpretation as its least element and the preferred interpretations as its maximum elements.

The intuitions behind the semantics of ADFs are as follows. In ADFs an interpretation is called *admissible* if it does not contain any unjustifiable information. An interpretation is called *preferred* if it represents maximum information about arguments without losing admissibility. Thus, each admissible interpretation is contained in a preferred interpretation. That is, to answer the credulous decision problem under preferred semantics (i.e., to investigate whether there is a preferred interpretation that contains the truth value of a given argument), it is sufficient to answer the problem under admissible semantics. An interpretation is *complete* if it exactly contains justifiable information. An interpretation is *two-valued model* if it exactly contains justifiable information and it is a two-valued interpretation. Finally, an interpretation is *grounded* if it collects all the information that is beyond any doubt.

**Example 2.48** *Let us consider again ADF $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$ from Example 2.43, depicted in Figure 2.7. Furthermore, consider several three-valued interpretations $\{v_0, v_1, v_2, v_3, v, v'\}$ as they are in Table 2.1. We investigate whether they are part of certain semantics. Since for each $i$ with $0 \leq i \leq 3$ it holds that $v_i \leq_i \Gamma_D(v_i)$, it*

| Interpretation | $a$ | $b$ | $c$ | $d$ | $\Gamma_D$ | part of semantics |
|---|---|---|---|---|---|---|
| $v_0 = v_{\mathbf{u}}$ | **u** | **u** | **u** | **u** | $\Gamma_D(v_{\mathbf{u}}) = v_1$ | *cf, adm* |
| $v_1$ | **t** | **u** | **u** | **f** | $\Gamma_D(v_1) = v_2$ | *cf, adm* |
| $v_2$ | **t** | **u** | **f** | **f** | $\Gamma_D(v_2) = v_3$ | *cf, adm* |
| $v_3$ | **t** | **t** | **f** | **f** | $\Gamma_D(v_3) = v_3$ | *cf, adm, prf, grd, com* |
| $v$ | **u** | **t** | **u** | **u** | $\Gamma_D(v) = v_2$ | *cf* |
| $v'$ | **u** | **u** | **u** | **t** | $\Gamma_D(v') = v_1$ | – |

**Table 2.1:** Interpretations for ADF from Example 2.48

*holds that each $v_i$ for $i$ with $0 \leq i \leq 3$ is an admissible interpretation of $D$. In general, each admissible interpretation is a conflict-free interpretation, since conflict-free interpretations are defined by weakening the condition of admissible interpretations. Thus, each $v_i$ for $i$ with $0 \leq i \leq 3$ is a conflict-free interpretation of $D$. The interpretation $v_3$ is a complete interpretation of $D$, since $\Gamma_D(v_3) = v_3$. Further, $v_3$ is the grounded interpretation of $D$, since it the least fixed point of $\Gamma_D$. In addition, $v_3$ is a maximal admissible interpretation of $D$, then $v_3$ is a preferred interpretation of $D$. Moreover, since $v_3$ is a two-valued interpretation, it is a two-valued model of $D$.*

*Note that the interpretation $v = \{b\}$ is not a(n) admissible/preferred/complete/the grounded interpretation of $D$, since $\Gamma_D(v) = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$, that is, $v \not\leq_i \Gamma_D(v)$. However, $v$ is a conflict-free interpretation of $D$. The truth value of $b$ is assigned to $\mathbf{t}$ in $v$. Thus, to show that $v$ is a conflict-free interpretation of $D$, it is enough to check whether $\varphi_b^v$ is satisfiable. Since $\varphi_b^v$ is indeed satisfiable, it holds that $v$ is a conflict-free interpretation of $D$.*

*Furthermore, for $v' = \{d\}$ it holds that $\Gamma_D(v') = \{a, \neg d\}$. Thus, since $v' \not\leq_i \Gamma_D(v')$, it holds that $v'$ is not a(n) admissible/preferred/complete/the grounded interpretation of $D$. To check whether $v'$ is a conflict-free interpretation, we have to check whether $\varphi_d^{v'}$ is satisfiable. Since $\varphi_d^{v'} \equiv \bot$ (i.e., it is unsatisfiable), it holds that $v'$ is not a conflict-free interpretation of $D$.*

Another semantics that we consider for ADFs in this work is stable semantics. The notion of stable semantics for ADFs is defined following the same ideas from logic programming. Stable models extend the concept of minimal model in logic programming by excluding self-justifying cycles of atoms. The notion of stable semantics of ADFs is defined over the two-valued model semantics of ADFs. Roughly speaking, a two-valued model is a stable model if it does not contain any support cycle. Thus, a user may use stable semantics to detect support cycle in a given ADFs.

Formally, to investigate whether a two-valued model $v$ of an ADF $D$ is a stable model of $D$ first we evaluate the reduction of $D$, namely $D^v$, introduced in Definition 2.49.

**Definition 2.49** *Let $D = (A, L, C)$ be an ADF and let $v$ be a two-valued model of $D$. The reduction of $D$, denoted by $D^v$ which is called a stb-reduct of $D$, is evaluated via following steps:*

1. *eliminate all nodes that are assigned to $\mathbf{f}$ in $v$ and their corresponding links from $D$;*
2. *replace the eliminated nodes with $\bot$ in the acceptance conditions of their children.*

To check whether a two-valued model $v$ (of ADF $D$) is a stable model, after evaluating ADF $D^v$, i.e., a *stb*-reduct (reduct) of $D$ for $v$, check whether the arguments that are assigned to $\mathbf{t}$ in $v$ are in the grounded interpretation of $D^v$. The concept of stable semantics of ADFs has been presented in ((Brewka et al., 2013), Definition 6) and in ((Brewka et al., 2018a), Definition 18), we recall it in Definition 2.50. Note that, $v^{\mathbf{t}}$ contains those arguments that $v$ maps to true, as it is presented in Definition 2.2.

**Definition 2.50** *Let $D = (A, L, C)$ be an ADF and let $v$ be a two-valued model of $D$. Then $v$ is a stable model of $D$ if $v^{\mathbf{t}} = w^{\mathbf{t}}$, where $w$ is the grounded interpretation of the stb-reduct $D^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}}$, $L^v = L \cap (A^v \times A^v)$, and $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$. The set of all stable models of $D$ is denoted by $stb(D)$.*

The grounded interpretation collects all the information that is beyond any doubt, thus, it is called that there is a constructive proof for all arguments presented in the grounded interpretation. Hence, intuitively. a two-valued model $v$ of $D$ is a stable interpretation (model), if there exists a constructive proof for all arguments assigned to true in $v$, if all arguments which are assigned to false in $v$ are actually false. Example 2.51 clarifies the notion of stable semantics of ADFs.

**Example 2.51** *Let $D = (\{a, b, c\}, \{\varphi_a : \neg b, \varphi_b : b \vee \neg c, \varphi_c : \neg a \vee \neg b\})$ be an ADF, depicted in Figure 2.8. $D$ has two two-valued models, namely $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$ and $v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$. We check whether they are stable models. To investigate whether $v_1$ is a stable model, first we evaluate the stb-reduct of $D$ under $v_1$, namely $D^{v_1} = (A^{v_1}, L^{v_1}, C^{v_1})$. Here $A^{v_1} = \{a, c\}$, $L^{v_1} = \{(a, c)\}$, and $\varphi_a : \neg\bot \equiv \top$ and $\varphi_c : \neg a \vee \neg\bot \equiv \top$.*

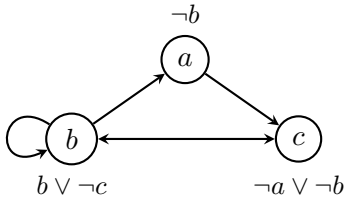


**Figure 2.8:** The ADF of Example 2.51.

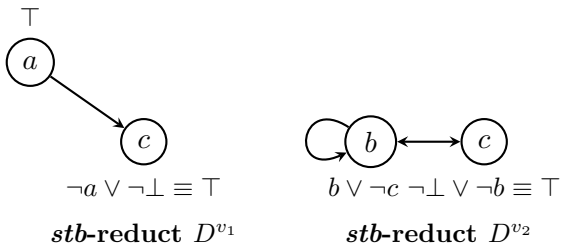


**Figure 2.9:** The reduct of ADF $D$ of Example 2.51.

*The reduct $D^{v_1}$ is depicted in Figure 2.9 (on the left). Since the unique grounded interpretation of $D^{v_1}$ is $w = \{a \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$, i.e., $w^{\mathbf{t}} = v_1^{\mathbf{t}}$, two-valued model $v_1$ is a stable model of $D$.*

*However, we show that $v_2$ is not a stable model of $D$. To this end, we first evaluate $D^{v_2} = (A^{v_2}, L^{v_2}, C^{v_2})$, where $A^{v_2} = \{b, c\}$, $L^{v_2} = \{(b, b), (b, c), (c, b)\}$, and $\varphi_b : b \vee \neg c$ and $\varphi_c : \neg\bot \vee \neg b \equiv \top$, depicted in Figure 2.9 (on the right). Since the unique grounded interpretation of $D^{v_2}$ is $w = \{b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\}$, i.e., $w^{\mathbf{t}} \neq v_2^{\mathbf{t}}$, it holds that two-valued model $v_2$ is not a stable model of $D$. Intuitively, model $v_2$ is not a stable model of $D$, since $v_2$ contains a support cycle over $b$, i.e., the acceptance of $b$ in $v_2$ depends on $b$ itself, that is, there is a cyclic justification. Thus, $v_2$ violates the main condition of stable semantics that a stable model should have no self-justifying cycles of atoms. Thus, $stb(D) = \{v_1\}$.*

An ADF may have no stable model. Example 2.52 presents an ADF that has a two-valued model, but no stable model.

**Example 2.52** *Let $D = (\{a, b, c\}, \{\varphi_a : c \vee b, \varphi_b : c, \varphi_c : a \leftrightarrow b\})$, depicted in Figure 2.10. The only two-valued model of $D$ is $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$. We investigate whether $v$ is a stable model of $D$. Since all arguments in $v$ are assigned to $\mathbf{t}$, all of them stay in the stb-reduct $D^v$. There is no argument in $v$ that is assigned to $\mathbf{f}$, thus no argument is replaced by*

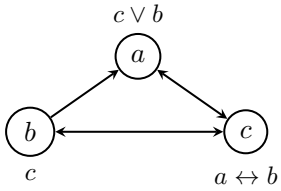


**Figure 2.10:** The ADF of Example 2.52

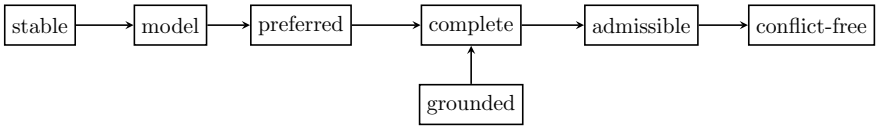


**Figure 2.11:** Relation among semantics of ADFs

*$\bot$ in the acceptance conditions of $C^v$. Hence, $D^v = D$. Now we have to check whether the set of arguments that are assigned to true in $v$ and the grounded interpretation of $D^v$ are equivalent. The unique grounded interpretation of $D^v$ is $w = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}$. That is, $w^{\mathbf{t}} = \{\}$. Thus, $w^{\mathbf{t}} \neq v^{\mathbf{t}}$. Hence, $v$ is not a stable model of $D$.*

Semantics of ADFs generalize semantics for AFs. Note that stable semantics of AFs have two generalisations in ADFs, namely two-valued model semantics and stable semantics. Two-valued model semantics reflect the 'zero-and-one' character of classical logic in ADFs, where in each two-valued model each argument is either acceptable or deniable without loosing of admissibility. Furthermore, stable semantics of ADFs indicate support cycle in a model. Akin to AFs, an ADF may have neither a two-valued model nor a stable model. Since in AFs there is no direct support, both notions of two-valued semantics and stable semantics coincide. In an ADF $D$ the following inclusions holds:

- $stb(D) \subseteq prf(D) \subseteq mod(D) \subseteq com(D) \subseteq adm(D) \subseteq cf(D)$;
- $grd(D) \subseteq com(D)$.

The relation between ADF semantics are shown in Figure 2.11. An arrow from $\sigma$ to $\gamma$, where $\sigma$ and $\gamma$ are semantics of ADFs, denotes that for any ADF $D$ it holds that $\sigma(D) \subseteq \gamma(D)$. It is shown in (Brewka et al., 2013), that semantics of ADFs directly generalize semantic of AFs. Definition 2.53 presents the associated ADF for a given AF.

**Definition 2.53** *For an AF $F = (A, R)$, define the ADF associated with $F$ as $D_F = (A, R, C)$ with $C = \{\varphi_a\}_{a \in A}$ such that for each $a \in A$ the acceptance condition is as follows:*

$$\varphi_a = \bigwedge_{(b,a) \in R} \neg b$$

In (Brewka et al., 2013) it is shown that the semantics of ADFs generalize the corresponding notions defined for AFs. In (Brewka et al., 2013, Theorem 2), it is presented that: an extension is admissible, complete, preferred, grounded for $F$ iff it is admissible, complete, preferred, grounded for $D_F$. To investigate the correspondence between semantics of an AF $F$ and its associated ADF $D_F$, we show how the extension-based semantics and labelled-based semantics of AFs relate to the interpretation-based semantics of ADFs.

Given an AF $F = (A, R)$ and its corresponding ADF $D_F = (A, R, C)$ (see Definition 2.53), the set of all possible conflict-free extensions of $F$ is denoted by $\mathscr{E}$ and the set of all possible conflict-free interpretations of $D_F$ is denoted by $\mathscr{V}$. The functions $Ext2Int_F$ and $Int2Ext_{D_F}$ in Definitions 2.54–2.56, are modifications of the labelling functions as given in (Baroni et al., 2018a), which we recalled in Definitions 2.34–2.35. Function $Ext2Int_F(S)$ represents the interpretation associated with a given extension $S$ in $F$, and function $Int2Ext_{D_F}(v)$ indicates the extension associated with a given interpretation $v$ of $D_F$.

**Definition 2.54** *Let $F = (A, R)$ be an AF, and let $S$ be an extension of $F$. The truth value assigned to each argument $a \in A$ by the three-valued interpretation $v_S$ associated with $S$ is given by $Ext2Int_F : \mathscr{E} \to \mathscr{V}$ as follows.*

$$Ext2Int_F(S)(a) = \begin{cases} \mathbf{t} & \text{if } a \in S, \\ \mathbf{f} & \text{if } \exists b \in A \text{ such that } (b, a) \in R \text{ and } b \in S, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

**Proposition 2.55** *Let $F = (A, R)$ be an AF, let $D_F$ be its associated ADF, and let $S$ be a conflict-free extension of $F$. Then $Ext2Int_F(S)$ is well-defined.*

**Proof**

1. Assume that $a \in S$. We show that $a$ is only assigned to $\mathbf{t}$ in $\mathit{Ext2Int}_F(S)$. By Definition 2.54, definitely it holds that $a \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(S)$, thus $a \mapsto \mathbf{u} \notin \mathit{Ext2Int}_F(S)$. We show that $a$ cannot be assigned the value $\mathbf{f}$ in $\mathit{Ext2Int}_F(S)$. Toward a contradiction, assume that $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(S)$. That is, by Definition 2.54, there exists a parent of $a$, namely $p_a$ such that $p_a \in S$. However, this means that $S$ contains conflict arguments, i.e., $a$ and $p_a$ with $(p_a, a) \in R$. Thus, $S$ is not a conflict-free extension. This contradicts the assumption that $S$ is a conflict-free extension of $F$. Thus, the assumption that $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(S)$ is wrong.
2. Assume that $a \notin S$. We show that either $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(S)$ or $a \mapsto \mathbf{u} \in \mathit{Ext2Int}_F(S)$, but not both of them. Either at least one parent of $a$ belongs to $S$ or none of them belong to $S$. By Definition 2.54, it is straightforward that if $a \notin S$ and a parent of $a$ belongs to $S$, then $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(S)$. In other words, if $a \notin S$ and none of the parents of $a$ belong to $S$, then $a \mapsto \mathbf{u} \in \mathit{Ext2Int}_F(S)$. That is, if $a \notin S$, then either $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F$ or $a \mapsto \mathbf{u} \in \mathit{Ext2Int}_F(S)$ but not both of them together.

Thus, if $S$ is a conflict-free extension, then $\mathit{Ext2Int}_F(S)$ is well-defined. □

Note that in Definition 2.54, the basic condition that $S$ has to be a conflict-free extension is a necessary condition for $\mathit{Ext2Int}_F(S)$ being well-defined. For instance, let $F = (\{a, b\}, \{(a, b)\})$. Set $S = \{a, b\}$ is an extension of $F$. However, $S$ does not satisfy the conflict-free property. On the other hand, $\mathit{Ext2Int}_F(S) = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, b \mapsto \mathbf{f}\}$. In other words, the correspondence between extensions and interpretations via $\mathit{Ext2Int}_F(.)$ is well-defined for conflict-free sets of arguments. This is the reason why we restrict $\mathscr{E}$ and $\mathscr{V}$ to the set of all conflict-free extensions of $F$ and conflict-free interpretations of $D_F$, respectively. By (Caminada and Dunne, 2019, Theorem 4), every strongly admissible extension of an AF is a conflict-free extension. Thus, if $S$ is a strongly admissible extension of AF $F$, then, by Proposition 2.55, it holds that $\mathit{Ext2Int}_F(S)$ is well-defined. So extensions of $F$ can be presented as interpretations of $D_F$. Also an interpretation of $D_F$ can be represented as an extension via the following function.

**Definition 2.56** *Let $D_F = (A, R, C)$ be the ADF associated with AF $F$, and let $v$ be an interpretation of $D_F$, that is, $v \in \mathscr{V}$. The associated extension $S_v$ of $v$ is obtained via application of $\mathit{Int2Ext}_{D_F} : \mathscr{V} \to \mathscr{E}$ on $v$,*

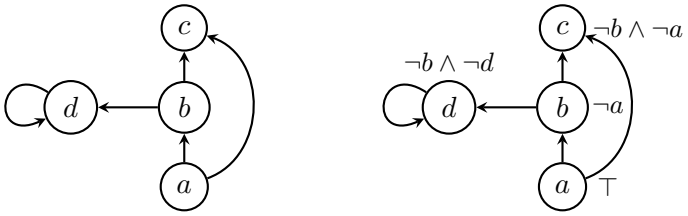


**Figure 2.12:** AF conversion to ADF of Example 2.59

*as follows:*

$$Int2Ext_{D_F}(v) = \{a \in A \mid a \mapsto \mathbf{t} \in v\}$$

Furthermore, the relation between labelled-based semantics of AFs and interpretation-based semantics of ADFs is presented in Definitions 2.57 and 2.58. Note that $\mathscr{L}$ denotes the set of labellings of AF $F$.

**Definition 2.57** *The function* $Lab2Int(\cdot) : \mathscr{L} \mapsto \mathscr{V}$ *maps three-valued labellings to three-valued interpretations such that*

- $Lab2Int(\lambda)(a) = \mathbf{t}$ *iff* $\lambda(a) = \texttt{in}$,
- $Lab2Int(\lambda)(a) = \mathbf{f}$ *iff* $\lambda(a) = \texttt{out}$, *and*
- $Lab2Int(\lambda)(a) = \mathbf{u}$ *iff* $\lambda(a) = \texttt{undec}$.

**Definition 2.58** *The function* $Int2Lab(\cdot) : \mathscr{V} \mapsto \mathscr{L}$ *maps three-valued interpretations to three-valued labellings such that*

- $Int2Lab(v)(a) = \texttt{in}$ *iff* $v(a) = \mathbf{t}$;
- $Int2Lab(v)(a) = \texttt{out}$ *iff* $v(a) = \mathbf{f}$;
- $Int2Lab(v)(a) = \texttt{undec}$ *iff* $v(a) = \mathbf{u}$.

Example 2.59 presents an instance of AF and its associated ADF.

**Example 2.59** *Let* $F = (\{a, b, c, d\}, \{(a, b), (b, c), (a, c), (b, d), (d, d)\})$ *be an AF. The associated ADF* $D_F$ *to* $F$ *is* $D_F = (A, R, \{\varphi_a : \top, \varphi_b : \neg a, \varphi_c : \neg b \wedge \neg a, \varphi_d : \neg b \wedge \neg d\})$, *depicted in Figure 2.12. For instance,* $S = \{a\}$ *is an admissible interpretation of* $F$. *The associated interpretation of* $S$ *is* $Ext2Int_F(S) = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}, d \mapsto \mathbf{u}\}$ *which is also an admissible interpretation of* $D_F$.

Links in ADFs are abstract in the sense that their meaning are determined solely by the acceptance conditions of the arguments. However, in ADFs, relations between arguments can be classified into four types, reflecting the relationship of attack and/or support that exists between the arguments. These are listed in Definition 2.61. Further, we present the update of an interpretation with a truth value of a given argument in Definition 2.60.

**Definition 2.60** *Let $D = (S, L, C)$ be an ADF, and let $v$ be an interpretation of $D$. The update of $v$ with a truth value $x \in \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ for an argument $b$ is denoted by $v|_x^b$, where,*

$$v|_x^b(a) = \begin{cases} x & \text{for } a = b, \\ v(a) & \text{for } a \neq b. \end{cases}$$

**Definition 2.61** *Let $D = (S, L, C)$ be an ADF. A relation $(b, a) \in L$ is called*

- supporting *(in $D$) if for every two-valued interpretation $v$, $v(\varphi_a) = \mathbf{t}$ implies $v|_{\mathbf{t}}^b(\varphi_a) = \mathbf{t}$;*
- attacking *(in $D$) if for every two-valued interpretation $v$, $v(\varphi_a) = \mathbf{f}$ implies $v|_{\mathbf{t}}^b(\varphi_a) = \mathbf{f}$;*
- redundant *(in $D$) if it is both attacking and supporting;*
- dependent *(in $D$) if it is neither attacking nor supporting.*

**Example 2.62** *Consider the acceptance condition of $\varphi_a : b \vee \neg c$ for argument $a$. By $\varphi_a$ the set of parents of $a$ is $\{b, c\}$. Thus, we clarify the type of $(b, a)$ and $(c, a)$. There are three satisfying two-valued interpretations, i.e., $v_1 = \{b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$, $v_2 = \{b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}$ and $v_3 = \{b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$, and one that does not satisfy the formula, i.e., $v_4 = \{b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$. By the definition of supporting links we have to check that whether $v_i(\varphi_a) = \mathbf{t}$ for $i$ with $1 \leq i \leq 3$ implies $v_i|_{\mathbf{t}}^b(\varphi_a) = \mathbf{t}$. Since for $i$ with $1 \leq i \leq 3$, $v_i(\varphi_a) = \mathbf{t}$ implies $v_i|_{\mathbf{t}}^b(\varphi_a) = \mathbf{t}$, it holds that $(b, a)$ is a supporting link. Furthermore, since $v_4(\varphi_a) = \mathbf{f}$ but $v_4|_{\mathbf{t}}^b(\varphi_a) = \mathbf{t}$, link $(b, a)$ is not an attack link.*

*Moreover, since $v_3(a) = \mathbf{t}$ but $v_3|_{\mathbf{t}}^c(\varphi_a) = \mathbf{f}$, it holds that $(c, a)$ is not a support link. However, it holds that $v_4(\varphi_a) = \mathbf{f}$ and $v_4|_{\mathbf{t}}^c(\varphi_a) = \mathbf{f}$. Thus, $(c, a)$ is only an attacking link.*

*As an example for a link that is both attacking and supporting, consider $\varphi_a : b \vee \neg b$. There are two satisfying two-valued interpretations for the formula, i.e., $v_1 = \{b \mapsto \mathbf{t}\}$ and $v_2 = \{b \mapsto \mathbf{f}\}$. Since for $i$ with $1 \leq i \leq 2$ it*

*holds that* $v_i(\varphi_a) = \mathbf{t}$ *implies* $v_i|^b_{\mathbf{t}}(\varphi_a) = \mathbf{t}$*, it holds that* $(b, a)$ *is a supporting link. Furthermore, since there is no two-valued interpretation that does not satisfy the formula, the link* $(b, a)$ *is also an attacking link. Thus,* $(b, a)$ *is a redundant link in* $\varphi_a : b \vee \neg b$*.*

*As an example for a link that is neither an attacking nor supporting, consider* $\varphi_a : (\neg c \vee b) \wedge (c \vee \neg b)$*. Let* $v = \{b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$ *be a two-valued interpretation that satisfies the formula. However,* $v|^b_{\mathbf{t}}(\varphi_a) = \mathbf{f}$*. Thus,* $(b, a)$ *is not a support link. Further, let* $v = \{b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$ *be a two-valued interpretation that does not satisfy the formula. However, it holds that* $v|^b_{\mathbf{t}}(\varphi_a) = \mathbf{t}$*. Thus,* $(b, a)$ *is not attacking. Hence* $(b, a)$ *is a dependent link.*

### 2.5.2 Subclasses of ADFs

In this section we restrict the syntactic structure of the acceptance conditions of ADFs to define a subclass of ADFs, called bipolar ADFs, introduced in (Brewka and Woltran, 2010). Furthermore, we show how a generalization of AFs, namely SETAFs (Nielsen and Parsons, 2006), presented in Section 2.4, can be embedded in ADFs.

#### Bipolar ADFs

Bipolar ADFs are now defined as ADFs which contain only supporting and attacking links.

**Definition 2.63** *Let* $D = (A, L, C)$ *be an ADF and* $L^+$ *be the set of all support links of* $L$ *and* $L^-$ *be the set of all attack links of* $L$*. ADF* $D$ *is named a bipolar ADF (BADF for short) iff* $L = L^+ \cup L^-$*.*

Since all links in the associated ADF $D_F$ of AF $F$ are attacking, it is clear that $D_F$ is a BADF, presented in Corollary 2.64.

**Corollary 2.64** *Let* $F$ *be an AF and let* $D_F$ *be its associated ADF. It holds that* $D_F$ *is a BADF.*

Example 2.65 is an instance of ADF which is also a BADF.

**Example 2.65** *Let* $D = (\{a, b, c\}, \{\varphi_a : \neg c \vee b, \varphi_b : \neg a \vee a, \varphi_c : a \wedge c\})$*, depicted in Figure 2.13. It holds that* $L^+ = \{(a, b), (b, a), (a, c), (c, c)\}$ *and* $L^- = \{(c, a), (a, b)\}$*, where* $(a, b)$ *is a redundant link of* $D$*, since* $(a, b) \in L^+ \cap L^-$*. Since* $L = L^+ \cup L^-$*, it holds that* $D$ *is a BADF.*

In contrast with Example 2.65, Example 2.66 presents an instance of ADF which is not a BADF. This shows that the class of BADFs is a strict subclass of ADFs.

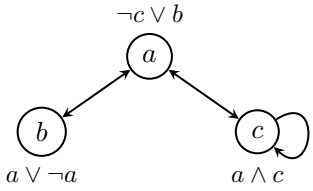


**Figure 2.13:** The ADF/BADF of Example 2.65

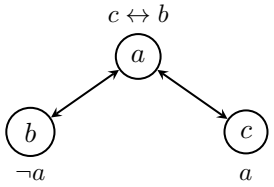


**Figure 2.14:** The ADF which is not a BADF of Example 2.66

**Example 2.66** *Let* $D = (\{a, b, c\}, \{\varphi_a : c \leftrightarrow b, \varphi_b : \neg a, \varphi_c : a\})$*, depicted in Figure 2.14. It holds that* $L^+ = \{(a, c)\}$ *and* $L^- = \{(a, b)\}$*. Similar to Example 2.62 one can check that* $(b, a), (c, a) \notin L^+ \cup L^-$*. That is,* $(b, a)$ *and* $(c, a)$ *are dependent links of* $D$*. Thus,* $D$ *is not a BADF.*

### Embedding SETAFs in ADFs

Translations between SETAFs and other abstract argumentation formalisms are studied in (Polberg, 2017). Furthermore, as observed by Polberg (2016) and Linsbichler et.al (2016), the notion of collective attacks, introduced in (Nielsen and Parsons, 2006), can also be represented in ADFs by using the right acceptance conditions. That is, SETAFs can be seen as a certain subclass of ADFs where sets of attacking arguments are captured in the acceptance conditions of these ADFs as conjunctions of disjunctions of negated atoms. Definition 2.67 presents the associated ADF for a given SETAF.

**Definition 2.67** *Let* $F = (A, R)$ *be a SETAF. The ADF associated to* $F$ *is a tuple* $D_F = (A, L, C)$ *in which* $L = \{(a, b) \mid (B, b) \in R, a \in B\}$ *and* $C = \{\varphi_a\}_{a \in S}$ *is the collection of acceptance conditions defined, for each* $a \in A$*, as*

$$\varphi_a = \bigwedge_{(B,a) \in R} \bigvee_{a' \in B} \neg a'.$$

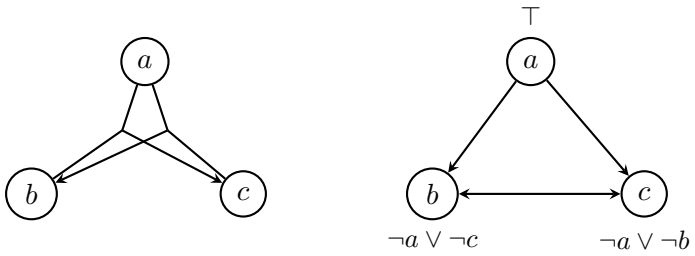


**Figure 2.15:** SETAF conversion to ADF of Example 2.68

Example 2.68 considers SETAF of Example 2.39 and presents its associated ADF.

**Example 2.68** *Let* $F = (\{a, b, c\}, \{(\{a, b\}, c), (\{a, c\}, b)\})$ *be a SETAF, presented in Example 2.39. The associated ADF* $D_F$ *to* $F$ *is* $D_F = (\{a, b, c\}, \{(a, c), (b, c), (a, b), (c, b)\}, \{\varphi_a : \top, \varphi_b : \neg a \vee \neg c, \varphi_c : \neg a \vee \neg b\})$. *The conversion of* $F$ *to* $D_F$ *is depicted in Figure 2.15.*

The relation between different sub-classes of ADFs, i.e. AFs, SETAFs and bipolar ADFs, has been formally studied in (Linsbichler et al., 2016). Note that the ADFs associated to AFs and SETAFs, respectively, only contain attacking links, and are therefore bipolar ADFs. As discussed in (Polberg, 2017), in general, SETAFs translate to bipolar ADFs that contain attacking and redundant links. Furthermore, it is proven that an extension is admissible, complete, preferred, grounded, stable for SETAF $F$ iff it is admissible, complete, preferred, grounded, stable for the associated ADF $D_F$.

## 2.6 Computational Complexity

In the field of computational complexity one is eager to investigate the question 'how difficult is a computational problem?' or in other words, to answer: 'How much does it cost to solve a computational problem?' The word 'cost' may relate to the amount of time or space (i.e., memory) taken by an algorithm to solve a problem. Thus, we mostly consider two types of complexity, namely, time complexity and space complexity to investigate how much time and space an algorithm (or a formalized machine) requires to solve a problem in the worst case. Complexity classes contain those computational problems that have a formally comparable

complexity. Also the notion of a Turing machine, as a representation of a formalized computational machine, is considered in complexity theory.

This section briefly explains some complexity classes and presents reasoning tasks related to AFs and ADFs. For comprehensive introductions to complexity theory we refer to (Papadimitriou, 2007; Arora and Barak, 2009) and for an overview of the complexity of reasoning tasks for AFs and ADFs we refer to (Dvořák and Dunne, 2018).

### 2.6.1 Basics

A decision problem $L$ consists of a set of instances (possibly infinite) and a query which can be answered *yes* or *no*. That is for an instance $l$ an algorithm decides whether $l \in L$; if actually $l$ belongs to $L$, then the answer to the decision problem would be *yes*, and if $l \notin L$, then the answer would be *no*. An instance of a specific decision problem would be $\mathsf{SAT}$ which has all propositional formula as its instances and the query to answer is that whether a given instance is satisfiable.

Informally speaking, a Turing machine (TM for short) is a description of an algorithm. If in a state of an algorithm (a machine) we have several choices, then the Turing machine is non-deterministic, otherwise it is deterministic. Then, the class of decision problems decidable with a deterministic TM in polynomial time is called $\mathsf{P}$. Using a non-deterministic TM leads to the definition of another important complexity class. The class of decision problems decidable with a non-deterministic TM in polynomial time is called $\mathsf{NP}$. Typically problems of $\mathsf{P}$ are considered 'tractable' and the ones outside of this class are considered 'intractable'. A useful notion is the concept of 'co-problem' of a problem. $L'$ is a co-problem of $L$ if for each $l$ the following property holds: $l$ is a *yes* instance of $L$ if and only if it is a *no* instance of $L'$. For $\mathsf{NP}$ the class containing the co-problems is called $\mathsf{coNP}$.

Given the complexity class $\mathscr{C}$, an oracle machine for a class $\mathscr{C}$ decides a given problem from $\mathscr{C}$ in one computational step. The class $\mathsf{P}^{\mathscr{C}}$ contains the problems that can be decided in polynomial time by a deterministic TM with access to a $\mathscr{C}$-oracle. The class $\mathsf{NP}^{\mathscr{C}}$ contains the problems that can be decided in polynomial time by a non-deterministic TM with access to a $\mathscr{C}$-oracle. Finally, the class $\mathsf{coNP}^{\mathscr{C}}$ contains the problems whose complementary problems can be decided in polynomial time by a non-deterministic TM with an access to a $\mathscr{C}$-oracle. We can now define the complexity classes of the polynomial hierarchy as follow.

**Definition 2.69** *Let* $\Sigma_0^{\mathsf{P}} = \Pi_0^{\mathsf{P}} = \Delta_0^{\mathsf{P}} = \mathsf{P}$. *Define for* $i > 0$ *the following classes.*

- $\Delta_i^{\mathsf{P}} = \mathsf{P}^{\Sigma_{i-1}^{\mathsf{P}}}$;
- $\Sigma_i^{\mathsf{P}} = \mathsf{NP}^{\Sigma_{i-1}^{\mathsf{P}}}$;
- $\Pi_i^{\mathsf{P}} = \mathsf{coNP}^{\Sigma_{i-1}^{\mathsf{P}}}$.

Furthermore, L is the class of problems that can be decided by a TM that only uses a logarithmic amount of memory with respect to the size of input. Moreover, the class DP 'difference class' contains an instance of $x =< x_1, x_2 >$ where $x_1$ is accepted by a problem $L_1$ (with $L_1 \in \mathsf{NP}$) and $x_2$ is accepted by a problem $L_2$ (with $L_2 \in \mathsf{coNP}$). For instance, the decision problem $x =< \varphi_1, \varphi_2 >$ in which $\varphi_1$ is satisfiable and $\varphi_2$ is unsatisfiable is an instance of DP problem, since $x$ is the intersection of positive instance of $L_1$ and $L_2$ where $L_1 = \{< \varphi_1, \varphi_2 > \mid \varphi_1 \text{ is satisfiable } \}$ and $L_2 = \{< \varphi_1, \varphi_2 > \mid \varphi_2 \text{ is unsatisfiable } \}$. The last class we consider is $\Theta_i^{\mathsf{P}}$. This class contains problems which are decidable by a deterministic TM in polynomial time, where the number of oracle calls is bounded by $\mathcal{O}(\log(n))$ such that $n$ is the input size. To further study the polynomial hierarchy we refer the reader to (Stockmeyer, 1976).

A useful method to investigate the complexity class of a given problem is the concept of *reduction*, presented in Definition 2.70.

**Definition 2.70** *Let A and B be two decision problems. Decision problem A is* P*-reducible to B if there exists a function f that satisfies the following conditions. For each instance x,*

1. *x is a* yes *instance of A if and only if f(x) is a* yes *instance of B;*
2. *f(x) is computable by a deterministic TM in polynomial time.*

We wrap up this section by rephrasing the notions of $\mathscr{C}$-hard and $\mathscr{C}$-complete, for a complexity class $\mathscr{C}$, in Definition 2.71.

**Definition 2.71** *Let* $\mathscr{C}$ *be a complexity class and let L be a decision problem.*

- *L is* $\mathscr{C}$*-hard when every problem* $L'$ *in* $\mathscr{C}$ *can be reduced in polynomial time to L.*
- *L is* $\mathscr{C}$*-complete, if it is* $\mathscr{C}$*-hard and* $L \in \mathscr{C}$.

### 2.6.2 Decision Problems and Complexity of AFs

Decision problems have been defined for the semantics of AFs (Dvořák and Dunne, 2018). Central reasoning tasks within argumentation formalisms are the credulous and sceptical acceptance problems as well as the verification problem. The credulous/skeptical decision problems are presented in Definition 2.72 and the verification problem is presented in Definition 2.73.

**Definition 2.72** *Let $F = (A, L)$ be an AF, let $a$ be an argument of $F$, and let $\sigma$ be semantics i.e., $\sigma \in \{adm, prf, grd, cf, com, idl, semi\text{-}stb, sadm, stb\}$.*

- *$a$ is credulously acceptable under $\sigma$ if there exists a $\sigma$-extension $S$ of $F$, i.e., $S \in \sigma(F)$ such that $a \in S$, denoted by $Cred_\sigma$. The credulous decision problem is presented formally in the following.*

$$Cred_\sigma(a, F) = \begin{cases} yes & \text{if } \exists S \in \sigma(F) \text{ such that } a \in S, \\ no & \text{otherwise} \end{cases}$$

- *$a$ is skeptically acceptable under $\sigma$ if for all $\sigma$-extension $S$ of $F$, i.e., $\forall S \in \sigma(F)$ it holds that $a \in S$, denoted by $Skept_\sigma$. The skeptical decision problem is presented formally in the following.*

$$Skept_\sigma(a, F) = \begin{cases} yes & \text{if } \forall S \in \sigma(F) \text{ it holds that } a \in S, \\ no & \text{otherwise} \end{cases}$$

**Definition 2.73** *Let $F = (A, L)$ be an AF, let $S$ a set of arguments of $F$, and let $\sigma$ be semantics of AFs. The verification problem indicates whether a given set $S$ is a $\sigma$-extension of $F$, i.e., whether $S \in \sigma(F)$, denoted by $Ver_\sigma$. The verification problem is presented formally in the following.*

$$Ver_\sigma(S, F) = \begin{cases} yes & \text{if } S \in \sigma(F), \\ no & \text{otherwise} \end{cases}$$

For strong admissibility semantics of AFs one more decision problem is proposed in (Dvořák and Wallner, 2020), called the minimum size strongly admissible set problem. Since we introduce this reasoning task for strong admissibility semantics of ADFs in Chapter 4, here we rephrase this notion for AFs as it is presented in (Dvořák and Wallner, 2020). Considering strong admissibility semantics we are interested in investigating whether a queried argument belongs to at least one strongly admissible extension $S$ of size of at most $k$, i.e., $|S| = k$, this decision problem is denoted by

| $\sigma$ | $Cred_\sigma$ | $Skept_\sigma$ | $Ver_\sigma$ |
|---|---|---|---|
| *cf* | in L | trivial | in L |
| *adm* | NP-c | trivial | in L |
| *prf* | NP-c | $\Pi_2^P$-c | coNP-c |
| *com* | NP-c | P-c | in P |
| *grd* | P-c | P-c | P-c |
| *stb* | NP-c | coNP-c | in P |
| *semi-stb* | $\Sigma_2^P$-c | $\Pi_2^P$-c | coNP-c |
| *sadm* | P | trivial | P |
| *idl* | in $\Theta_2^P$ | in $\Theta_2^P$ | in $\Theta_2^P$ |

**Table 2.2:** Complexity of reasoning with AFs.

*k-Witness*$_{sadm}$. Note that $k$ is a part of the input of this problem and specifies the size constraint. In (Dvořák and Wallner, 2020) it is shown that *k-Witness*$_{sadm}$ is NP-complete.

The complexity of reasoning problems for AFs has been studied in (Dunne and Caminada, 2008; Dunne and Bench-Capon, 2002; Dvořák, 2012; Dvořák and Dunne, 2018; Dvořák and Wallner, 2020) and is summarized in Table 2.2 for the semantics considered in this work. In Table 2.2, for reasons of brevity, instead of writing $\mathscr{C}$-complete we have $\mathscr{C}$-c, for a class $\mathscr{C}$.

### 2.6.3 Decision Problems and Complexity of ADFs

Key reasoning problems in ADFs are the credulous and skeptical decision problems of arguments, and the problem of verification of an interpretation. In ADFs beside an argument being acceptable in an interpretation, there is a symmetric notion of an argument being deniable. Thus, first in Definition 2.74 we present the notions of an argument being accepted or denied in an interpretation. Then, we present the associated decision problems for ADFs.

**Definition 2.74** *Let $D = (A, L, C)$ be an ADF and let $v$ be an interpretation of $D$.*

- *An argument $a \in A$ is called acceptable with respect to $v$ if $\varphi_a^v$ is irrefutable.*
- *An argument $a \in A$ is called deniable with respect to $v$ if $\varphi_a^v$ is unsatisfiable.*

*We say that an argument $a$ is* justifiable *with respect to $v$ if it is either acceptable or deniable with respect to $v$.*

For instance, let $D = (\{a, b\}, \{\varphi_a : a, \varphi_b : \neg a\})$ be an ADF. It holds that $a$ is acceptable with respect to $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}\}$, and $b$ is deniable with respect to $v$. Furthermore, both $a$ and $b$ are justifiable with respect to $v$. Dung's Fundamental Lemma of AFs is reformulated in (Diller et al., 2018) for ADFs, represented in Lemma 2.75.

**Lemma 2.75** *Let $D$ be an ADF. Assuming $v$ is an admissible interpretation of $D$, and $a$ and $a'$ are arguments which are justifiable with respect to $v$. Then,*

- $v' = v|^a_{\mathbf{t/f}}$ *is an admissible interpretation of $D$,*
- $a'$ *is justifiable with respect to $v'$.*

The credulous and skeptical decision problems of ADFs are presented in Definition 2.76.

**Definition 2.76** *Let $D = (A, L, C)$ be an ADF, let $\sigma$ be semantics of ADFs, let $a$ be an argument of $A$, and let $x$ be a two-valued truth value, i.e., $x \in \{\mathbf{t}, \mathbf{f}\}$.*

- *$a$ is credulously justifiable under $\sigma$ if there exists a $\sigma$-interpretation $v$ of $D$ in which $v(a) = x$, denoted by $Cred_\sigma$ and presented formally in the following.*

$$Cred_\sigma(a \mapsto x, D) = \begin{cases} yes & \text{if } \exists v \in \sigma(D) \text{ such that } v(a) = x; \\ no & \text{otherwise} \end{cases}$$

  *Note that in the above definition if $x = \mathbf{t}$, then it is called that $a$ is credulously acceptable under $\sigma$ (in $v$), and if $x = \mathbf{f}$, then it is called that $a$ is credulously deniable under $\sigma$ (in $v$).*

- *$a$ is skeptically justifiable under $\sigma$ if for each $v$ with $v \in \sigma(D)$ it holds that $v(a) = x$, denoted by $Skept_\sigma$ and presented formally in the following.*

$$Skept_\sigma(a \mapsto x, D) = \begin{cases} yes & \text{if } \forall v \in \sigma(D) \text{ it holds that } v(a) = x; \\ no & \text{otherwise} \end{cases}$$

*If $x = \mathbf{t}$, then it is called that argument $a$ is skeptically acceptable under $\sigma$ (in $v$) , and if $x = \mathbf{f}$, then it is called that $a$ is skeptically deniable under $\sigma$ (in $v$).*

**Example 2.77** *Let us consider again ADF $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$ from Example 2.43. Since there is an admissible interpretation assigning $a$ to true ($\mathbf{t}$) (e.g. $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, d \mapsto \mathbf{f}\}$, see Table 2.1) it holds that $a$ is credulously acceptable under admissible semantics in the ADF $D$. However, argument $d$ is not credulously acceptable under admissible semantics in $D$. We explain why $d$ cannot be credulously accepted under admissible semantics of $D$. Note that the only preferred interpretation of $D$ is $v_3 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$, as in Table 2.1, in which $d$ is assigned to $\mathbf{f}$. That is, $d$ is credulously and skeptically deniable under preferred semantics of $D$. Since any admissible interpretation must be equally or less informative than at least one preferred interpretation, $d$ cannot be credulously accepted under admissible semantics of ADFs.*

*Furthermore, since the unique preferred interpretation $v_3$ is also a unique complete interpretation, stable model, two-valued model, and the grounded interpretation of $D$, it holds that $b$ is skeptically acceptable, and $c$ is skeptically deniable under $\sigma$ semantics, for $\sigma \in \{prf, mod, com, stb, grd\}$.*

Due to the relations between different semantics one can find relations between reasoning tasks. For instance, by the definition of preferred semantics; each preferred interpretation is $\leq_i$-maximal admissible interpretation and each preferred interpretation is a complete interpretation. Thus, it holds that the credulous decision problems under preferred, admissible, and complete semantics coincide.

Furthermore, since there is only one grounded interpretation in any ADF, credulous and skeptical decision problems under grounded semantics coincide. Also, since the grounded interpretation is the $\leq_i$-minimal complete interpretation, skeptical decision problems under complete and grounded semantics coincides.

Another important reasoning task of ADFs is the verification problem, presented in Definition 2.78.

**Definition 2.78** *Let $D$ be an ADF, let $\sigma$ be semantics of ADFs, and let $v$ be an interpretation of $D$. The verification problem decides whether $v$ is a $\sigma$-interpretation, i.e., if $v \in \sigma(D)$, denoted by $Ver_{sadm}(v, D)$, and is presented formally as follows:*

$$Ver_\sigma(v, D) = \begin{cases} yes & \text{if } v \in \sigma(D), \\ no & \text{otherwise} \end{cases}$$

| $\sigma$ | $Cred_\sigma$ | $Skept_\sigma$ | $Ver_\sigma$ |
|---|---|---|---|
| *cf* | NP-c | trivial | NP-c |
| *adm* | $\Sigma_2^P$-c | trivial | coNP-c |
| *prf* | $\Sigma_2^P$-c | $\Pi_3^P$-c | $\Pi_2^P$-c |
| *com* | $\Sigma_2^P$-c | coNP-c | DP-c |
| *grd* | coNP-c | coNP-c | DP-c |
| *stb* | $\Sigma_2^P$-c | $\Pi_2^P$-c | coNP-c |
| *mod* | NP-c | coNP-c | in P |

**Table 2.3:** Complexity of reasoning with ADFs.

**Example 2.79** *Considering ADF* $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$ *from Example 2.43. Let* $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$. *Since* $v$ *is an admissible interpretation of* $D$ *but it is not a preferred interpretation of* $D$, *it holds that* $Ver_{adm}(v, D) =$ yes *and* $Ver_{prf}(v, D) =$ no.

The complexity of nearly all ADF reasoning tasks has been analyzed in (Strass and Wallner, 2015; Dvořák and Dunne, 2018; Gaggl et al., 2021). The complexity landscape of reasoning in ADFs, under semantics which are presented in this work, is shown in Table 2.3. Except for the trivial tasks, computational complexity of nearly all reasoning tasks in ADFs when compared to AFs increases by one step in the polynomial hierarchy.

# Part II

# Semantics

# Chapter 3

# Strong Admissibility

Different criteria for settling the acceptance of arguments are called semantics. Semantics of ADFs have so far mainly been defined based on the concept of admissibility. A new type of admissibility-based semantics is usually proposed by introducing further restrictions on the set of accepted arguments. However, the notion of strongly admissible semantics studied for abstract argumentation frameworks has not yet been introduced for ADFs. Strong acceptability of AFs is not only a new point of view on the acceptability of arguments, but it is also a way of providing reasons why an argument belongs to the grounded extension. In this part of the thesis, we present the concept of strong admissibility of interpretations for ADFs. Furthermore, we show that strongly admissible interpretations of ADFs form a lattice with the grounded interpretation as the maximal element. We also present algorithms to answer the following decision problems:

1. whether a given interpretation is a strongly admissible interpretation of a given ADF, and
2. whether a given argument is strongly acceptable/deniable in a given interpretation of a given ADF.

In addition, we show that the strongly admissible semantics of ADFs forms a proper generalization of the strongly admissible semantics of AFs.

## 3.1 Introduction

Interest and attention in argumentation theory from artificial intelligence-related researchers has been increasing, as witnessed by the wide variety of formalisms to model argumentation and by the variety of semantics that

clarify the acceptance of arguments (Baroni et al., 2018b; van Eemeren et al., 2014). Abstract argumentation frameworks (AFs for short) as introduced in the landmark paper by Dung (1995) have gained more and more significance in the AI domain. First of all, it has been shown in (Dung, 1995) that AFs are useful to capture the essence of different non-monotonic formalisms. In addition, compared to other non-monotonic formalisms (which are built on top of classical logical syntax), AFs are a much simpler formalism; indeed, they are just directed graphs in which nodes present arguments and directed edges indicate attack relations among arguments. Moreover, AFs are nowadays an integral concept in several advanced argumentation-based formalisms in the sense that their semantics are defined based on a translation (typically called an instantiation) to Dung's AFs. Finally, the simplicity of the syntax of AFs has made them an attractive modeling tool in several areas, such as multi-agent systems (McBurney et al., 2012), multi-agent negotiation (Amgoud et al., 2007), and legal reasoning (Bench-Capon and Dunne, 2005).

Despite the popularity and simplicity of AFs, these frameworks are used to model argumentation with simple attack relations among arguments. Thus, there exist a number of generalizations of AFs, for instance, modeling group attacks among arguments (Nielsen and Parsons, 2006) or modeling preference over the arguments (Bench-Capon, 2003; Bench-Capon and Atkinson, 2009). Among all generalizations of AFs, abstract dialectical frameworks (ADFs) were first introduced in (Brewka and Woltran, 2010) and further refined in (Brewka et al., 2013, 2018a). They are expressive generalizations of AFs in which the logical relations among arguments can be represented. This allows researchers to express notions of support, collective attacks, and even more complicated relations. Thanks to their flexibility in formalizing relations between arguments, ADFs have recently been used in several applications; in legal reasoning (Al-Abdulkarim et al., 2016, 2014; Collenette et al., 2020), online dialog systems (Neugebauer, 2017, 2019), and text exploration (Cabrio and Villata, 2016).

A key question in formal argumentation is 'How is it possible to evaluate arguments in a given formalism?'. Answering this question leads to the introduction of several types of semantics. In AFs, different semantics single out coherent subsets of arguments that "fit" together, according to specific criteria (Baroni et al., 2011). More formally, an AF semantics takes an argumentation framework as input and produces as output a collection of sets of arguments, called extensions. Thus, different semantics reflect different points of view about the acceptance or denial of arguments. Most

of the semantics of AFs/ADFs are based on the concept of admissibility.

In AFs, admissibility plays an important role with respect to rationality postulates (Caminada and Amgoud, 2007). Often a new semantics is an improvement of an already existing one by introducing further restrictions on the set of accepted arguments (that are chosen together) or possible attackers. One of the main admissibility-based semantics of AFs is the grounded semantics. First, each AF has a unique grounded extension. Second, the elements of the grounded extension usually belong to other semantics of AFs. The grounded extension collects all unattacked (undoubted) arguments and each argument that can be iteratively supported by these unattacked arguments. Informally, the grounded extension accepts those arguments that no one can avoid to accept; it rejects all the arguments that no one avoid to reject; and it does not have any idea about all other arguments. Thus, no one has any doubt on the acceptance of the arguments that are in the grounded extension. Thus, answering the credulous decision problem under grounded semantics of AFs (i.e., investigating whether a queried argument is part of the grounded extension of a given AF) has a considerable importance.

It has been shown that to answer the credulous decision problem under grounded semantics, not all arguments within the grounded extension are necessary. As a remedy, another set of semantics, namely *strong admissibility semantics* has been introduced for AFs (Baroni and Giacomin, 2007; Caminada, 2014; Caminada and Dunne, 2019). While the grounded extension collects all the arguments of a given AF that can be accepted without any doubt, strongly admissible extensions are subsets of the grounded extension that satisfy the same condition. Actually a strongly admissible extension explains why its arguments can be accepted without any doubt, without presenting further information of all arguments in the grounded extension. In AFs, the concept of strong admissibility semantics has first been defined in the work of Baroni and Giacomin (2007), based on the notion of strong defence. Later in (Caminada, 2014) this concept was introduced without referring to strong defence. Furthermore, in (Caminada and Dunne, 2019), Caminada and Dunne presented a labelling account of strong admissibility to answer the decision problems of AFs under grounded semantics.

The role and the relevance of strong admissibility semantics and grounded discussion games for AFs has been studied in (Caminada, 2018, 2014; Caminada and Dunne, 2019). That is, it has been shown that strongly admissible extensions/labellings make a lattice with the maximum element

being the grounded extension of a given AF. Therefore, the concept of strong admissibility semantics of AFs relates to grounded semantics of AFs in a similar way as the relation between admissible semantics of AFs and preferred semantics of AFs. That is, to answer the credulous decision problem of AFs under grounded semantics, it is sufficient to solve the decision problem for AFs under strongly admissible semantics, i.e., it is enough to indicate whether there exists a strongly admissible extension that contains a queried argument. Furthermore, it has been shown that the strong admissibility semantics and the corresponding discussion games may be the basis for an algorithm that can be used not only for answering the decision problem but also for human-machine interaction (cf. (Caminada and Uebis, 2020; Booth et al., 2018)). In addition, the computational complexity of strong admissibility of AFs has been analyzed (Caminada and Dunne, 2020; Dvořák and Wallner, 2020).

It has been shown in (Brewka et al., 2018a) that each AF can be represented as an ADF; furthermore, it has been shown that ADFs provide all of Dung's standard semantics, proposed in (Dung, 1995) for AFs, so there is no loss in semantic richness. By the use of general propositional formulas as argument acceptance conditions, ADFs allow for richer relations between arguments than AFs, which only allow attack. Because ADFs are at least as expressive as AFs, they can represent all important problem aspects that AFs can represent. However, some of the semantics of AFs have not yet been introduced for ADFs, namely, *strongly admissible semantics*. Because of the special structure of ADFs, the definition of strong admissibility semantics of AFs cannot be directly reused in ADFs. Because of the importance of the notion of strongly admissible semantics to investigate the acceptance of a queried argument under the grounded semantics, in the current work we introduce the concept of strongly admissible semantics of ADFs that satisfies/follows the same set of properties as this concept has in AFs, presented in Section 3.1.1. We do so not only because ADFs are generalisations of AFs, but also because ADFs are expressive enough to model a wide range of non-monotonic knowledge representation languages, and the role of strongly admissible semantics to answer the credulous decision problem. ADFs have been very actively researched (Brewka et al., 2011; Brewka and Gordon, 2010; Ellmauthaler, 2012; Strass and Wallner, 2015; Strass, 2013a, 2018; Wallner, 2020).

A main characteristic of strongly admissible semantics of ADFs is that they can be used to explain the answer to the question: 'Why is an argument justified under grounded semantics of ADFs?'. For instance,

suppose that an ADF is used to formalize a knowledge-base that presents methods to cure a disease or to make a decision in the legal domain. It is not enough to tell a patient or client that we pick a certain argument since it is presented in a semantics, but they to be convinced why this is the case. Moreover, having automated argumentation systems that can help people to make better choices is a goal of human-machine interaction (Hunter, 2018; Chalaguine and Hunter, 2020). To persuade agents to perform (or not to perform) an certain action, a user needs to have further explanation about the acceptance of arguments. To address this open problem, we previously considered grounded semantics of ADFs, since no one has any doubt on the evaluation of arguments in the grounded interpretation. Then, as a first remedy in (Keshavarzi Zafarghandi et al., 2020), we introduced a discussion game to answer the credulous decision problem of ADFs under grounded semantics without constructing the full grounded interpretation of the given ADF. Subsequently, in the current work we propose the notion of strong admissibility semantics of ADFs. Both methods can be used to explain the truth values of arguments in the grounded interpretation. We think that grounded discussion games of ADFs and strong admissibility semantics of ADFs are two sides of the same coin. However, studying the relation between grounded discussion games, presented in (Keshavarzi Zafarghandi et al., 2020), and the strong admissibility semantics of ADFs is beyond the topic of this work and is left for future research.

### 3.1.1 Requirements of strong admissibility semantics

As mentioned before, it is important to investigate whether a queried argument is in the grounded extension of an AF. This is mainly because each AF has a unique grounded extension and no one has any doubt on the acceptance of the arguments of the grounded extension. Furthermore, in applications it is significant not only to answer whether a queried argument is in the grounded extension but also to explain why it is so.

On the one hand, some discussion games have been presented to answer this decision problem under the grounded semantics (Caminada and Podlaszewski, 2012a; Caminada, 2015; Modgil and Caminada, 2009; Caminada, 2018). The idea is that these discussion games can be used as proof procedures for the grounded semantics. That is, a queried argument belongs to the grounded extension of a given AF iff it is possible to win the associated discussion game. This makes it possible to use the discussion games for the purpose of explanation "why is an argument in the grounded extension?". That is, instead of simply mentioning that an argument is in

the grounded extension, a discussion game explains why no one has any doubt to accept the argument in question. A web-based implementation of grounded discussion games is presented in (Booth et al., 2018).

On the other hand, the notion of strongly admissible semantics of AFs has been presented in (Baroni and Giacomin, 2007; Caminada, 2014; Caminada and Dunne, 2019). to deal with the same issue. That is, the notion of strongly semantics of AFs explains "Why does a queried argument belong to the grounded extension of an AF?" without presenting the whole grounded extension, that is, without presenting any further explanation about the irrelevant arguments to the argument in question. Furthermore, the role and the relevance of strong admissibility semantics of AFs in the Standard Grounded Game (Modgil and Caminada, 2009; Prakken and Sartor, 1997) and the Grounded Persuasion Game (Caminada and Podlaszewski, 2012a,b) has been studied in (Caminada, 2014; Caminada and Dunne, 2019).

AFs have the property that each AF has a strongly admissible extension. In addition, the set of strongly admissible extensions of a given AF forms a lattice with the least element being the empty set and the maximum element being the grounded extension. Moreover, it has been shown that the notion of strongly admissible semantics differs from all other existing semantics of AFs, namely conflict-free, admissible, preferred, complete, grounded and ideal semantics (Caminada and Dunne, 2019; Baroni and Giacomin, 2007). That is, indeed the notion of strongly admissible semantics is a new point of view on the acceptance of arguments of a given AF.

Similar to AFs, in ADFs the concept of grounded semantics is an important point of view on the acceptance of arguments. Each ADF has a unique grounded interpretation that presents the truth values of arguments about which no one has any doubt. Thus, it is crucial to investigate the truth value of a queried argument in the grounded interpretation of an ADF. Furthermore, it is required to explain why a queried argument has a specific truth value in the grounded interpretation. To investigate this issue in (Keshavarzi Zafarghandi et al., 2020), the notion of discussion game for grounded semantics of ADFs has been presented. This game works locally by considering those ancestors of a certain argument that can affect the evaluation of the argument in the grounded interpretation. In this way, the grounded decision problem can be answered without constructing the full grounded interpretation. However, the notion of strongly admissible semantics has not yet been presented for ADFs to explain the reason for a truth value of a queried argument in the grounded interpretation. We

show that the notion of strongly admissible semantics of ADFs presented in this work will satisfy the following conditions, which are akin to the properties of the notion of strongly admissible semantics of AFs.

- Strong admissibility is defined in terms of strongly justified arguments.
- Strongly justified arguments are recursively reconstructed from their strongly justified parents.
- Each ADF has at least one strongly admissible interpretation.
- The set of strongly admissible semantics of ADFs forms a lattice with the least element being the trivial interpretation and the maximum element being the grounded interpretation.
- Strongly admissible semantics is used to answer whether an argument is justified in the grounded interpretation of a given ADF. This is because this notion has a close relation to the grounded semantics, in the formally precise sense that the maximal element of the lattice of strongly admissible interpretations is the grounded interpretation.
- The notion of strongly admissible semantics of ADFs differs from the notions of admissible, conflict-free, complete and grounded semantics of ADFs.
- The notion of strongly admissible semantics for ADFs is a proper generalization of strongly admissible semantics for AFs.

Our result leads to the presentation of an algorithm to answer the decision problem whether a given interpretation is a strongly admissible interpretation in a given ADF. In addition, since some generalizations of Dung's AFs can be seen as special cases of ADFs, for instance, SETAFs (Nielsen and Parsons, 2006), as shown in (Polberg, 2017, 2016), and bipolar AFs (Cayrol and Lagasquie-Schiex, 2005; Oren et al., 2010; Nouioua, 2013), the notion of strongly admissible semantics presented for ADFs may carry over to these special cases. However, the focus of this work is to present a formal proof of clarifying the relation between strongly admissible semantics of AFs and ADFs.

This chapter is structured as follows. In Section 3.2, the main contribution of our work is to introduce the concept of strongly admissible semantics for ADFs. Subsequently, we show that in each ADF, the set of

strongly admissible interpretations forms a lattice with the trivial interpretation as the unique minimal element and the grounded interpretation as the unique maximal element. In Section 3.3 we show that the concept of strongly admissible semantics of ADFs is a generalization of the notion of strongly admissible semantics of AFs.

Section 3.4 presents an alternative definition for strongly admissible semantics of ADFs that is presented without referring to strongly justified arguments, when compared to the definition in Section 3.2. This definition also leads to a straightforward algorithm to answer the verification problem of ADFs under strongly admissible semantics. We also present an alternative definition for investigating whether a given argument is strongly justified in a given interpretation, which does not have the difficulties of the definition of strongly justified of arguments in an interpretation that is presented in Section 3.2. This method also leads to an algorithm to answer whether a given argument is a strongly justified argument in a given interpretation.

In Section 3.5, we present a finer relation between the sequence of strongly admissible extensions of a given AF and the sequence of strongly admissible interpretations constructed in the associated ADF. Finally, in Section 3.6, we present the conclusion of our work and we present some future research questions arising from this work.

A preliminary version of the material included in this chapter appeared as (Keshavarzi Zafarghandi et al., 2021b) As the first addition to the previous work, we prove that strongly admissible semantics of ADFs form a generalization of strongly admissible semantics of AFs, presented in Section 3.3. This extended version contains new technical results including answering the verification problem under strong admissibility semantics of ADFs without considering whether all of the arguments that are presented in the given interpretation are strongly justified, reported in Section 3.4. In addition, in Section 3.4, we present a new method to investigate whether a given argument is strongly justified in a given interpretation. Further, in Section 3.5, we study finer relations between the sequence of strongly admissible extensions constructed based on a given extension of an AF and the sequence of strongly admissible interpretations of the associated ADF.

## 3.2 The Strongly Admissible Semantics for ADFs

In the following, we present the concept of strong admissibility semantics for ADFs. As we discussed in the introduction, we are aiming to generalize

the notion of strong admissibility semantics of AFs, so that the concept of strong admissibility semantics of ADFs relates to grounded semantics of ADFs in a similar way as the concept of admissible semantics of ADFs relates to preferred semantics of ADFs. As we mentioned in the introduction, following the definition by Baroni and Giacomin (2007), Caminada showed in (2018; 2014) that strong admissibility plays a critical role in discussion games for AFs under grounded semantics (Caminada, 2014; Caminada and Dunne, 2019). In (Keshavarzi Zafarghandi et al., 2020), we introduced a discussion game to answer the credulous decision problem of ADFs under grounded semantics without constructing the full grounded interpretation of the given ADF. However, the concept of strong admissibility semantics of ADFs has not been introduced in the literature so far. This was a motivation for us to present the notion of strong admissibility semantics for ADFs and study the characteristics of this concept. Similarly to the AF case, a strongly admissible interpretation of an ADF may be used not only to answer the credulous decision problem, but also to explain why an argument is justified in the grounded interpretation.

In ADFs, beside an argument being acceptable in an interpretation, there is a symmetric notion of an argument being deniable. In contrast with Definition 2.23, in which the concept of strong admissibility semantics of AFs is defined based on the concept of strongly defended arguments, in ADFs we define the concept of strong admissibility semantics based on the concept of strongly acceptable/deniable arguments. To this end, in Definition 3.1 we introduce the notion of strong justification (i.e., strongly accepted/denied) of an argument in an ADF in a given interpretation.

Note that in the following, $v_{|_P}$ is equal to $v(p)$ for any $p \in P$; however, it assigns all other arguments that do not belong to $P$ to $\mathbf{u}$, i.e., $v_{|_P} = v_{\mathbf{u}}|^{p \in P}_{v(p)}$.

**Definition 3.1** *Let $D = (A, L, C)$ be an ADF and let $v$ be an interpretation of $D$. Argument $a$ is a* strongly justified *argument in interpretation $v$ with respect to set $E$ if one of the following two conditions hold:*

- *$v(a) = \mathbf{t}$ and there exists a subset $P$ of parents of $a$ excluding $E$, namely $P \subseteq par(a) \setminus E$, such that (a) $a$ is acceptable with respect to $v_{|_P}$ and (b) all $p \in P$ are strongly justified in $v$ w.r.t. set $E \cup \{p\}$.*
- *$v(a) = \mathbf{f}$ and there exists a subset $P$ of parents of $a$ excluding $E$, namely $P \subseteq par(a) \setminus E$, such that (a) $a$ is deniable with respect to $v_{|_P}$ and (b) all $p \in P$ are strongly justified in $v$ w.r.t. set $E \cup \{p\}$.*

*An argument $a$ is* strongly acceptable*, respectively* strongly deniable*, in $v$ if $v(a) = \mathbf{t}$, respectively $v(a) = \mathbf{f}$, and $a$ is strongly justified in $v$ with*

*respect to set* $\{a\}$. *We say that an argument is* strongly justified *in* $v$ *if it is either strongly acceptable or deniable in* $v$.

Note that in Definition 3.1, if $par(a) = \emptyset$, then $\varphi_a$ must be either $\top$ or $\bot$. In case $\varphi_a = \top$ and $v(a) = \mathbf{t}$, then it follows that $a$ is strongly acceptable in $v$ by taking $P = \emptyset$. Similarly, in case $\varphi_a = \bot$ and $v(a) = \mathbf{f}$, then it follows that $a$ is strongly deniable in $v$. In all other cases, $a$ is not strongly justified in $v$. Furthermore, we say that $a$ *is not strongly justified in an interpretation* $v$ if there is no subset $P$ of parents of $a$ that satisfies the condition of Definition 3.1 for $a$.

Definition 3.1 is well-defined. That is, checking whether $a$ is a strongly justified argument in interpretation $v$ is decidable. Because a given ADF is finite, $a$ has a finite number of ancestors. Furthermore, in each step, to check whether the conditions of Definition 3.1 are satisfied for $a$, we exclude the argument in question from the set of parents of $a$. Hence, checking whether there exists a set of arguments that satisfies the conditions of Definition 3.1 for a given argument will stop after finitely many steps, since at some point $par(a) \setminus E$ must be the empty set, and hence the only possible $P$ is the empty set, which means that we can decide whether $a$ is strongly justified in $v$ without inspecting any further parents of $a$.

Since the class of ADFs is a generalization of the class of AFs (see Definition 2.53), in the following, we informally discuss why Definition 3.1 can be viewed as a generalization of Definition 2.22 for AFs. In Section 3.3, we formally show that the strongly admissible semantics of ADFs is a proper generalization of strongly admissible semantics of AFs.

In the two items of Definition 3.1, the set $P$ contains exactly those parents of $a$, excluding $a$, that satisfy $v(a)$ and of which the truth value is presented in $v$. Definition 3.1 presents the same idea as Definition 2.22: that an argument $a$ is strongly defended (accepted) if it can be defended by some arguments other than itself. Furthermore, each defender of $a$ has to be strongly defended. Akin to AFs, in ADFs an argument $a$ is strongly justifiable if its truth value is justified by some arguments other than itself, where each of those other arguments is strongly justified.

The notion of strong acceptability/deniability of arguments in a given interpretation is illustrated in Example 3.2.

**Example 3.2** *Consider the ADF presented in Example 2.43, i.e.,* $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$*, depicted in Figure 2.7. Let* $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$. *We show that* $c$ *and* $d$ *are strongly justified in* $v$ *and* $b$ *is not strongly justified in* $v$. *Since* $v(c) = v(d) = \mathbf{f}$,

*we show that c and d are strongly deniable in v. First, since* $\varphi_d^{v_{\mathbf{u}}} \equiv \bot$ *and* $v(d) = \mathbf{f}$*, it holds that d is strongly deniable in v.*

*Furthermore, to show that c is strongly deniable in v, we show that c is strongly deniable in v with respect to* $E = \{c\}$*. We choose a subset of parents of c excluding c, namely,* $P = \{d\}$*. It is easy to check that* $\varphi_c^{v_{|P}}$ *is unsatisfiable, i.e.,* $\varphi_c^{v_{|P}} \equiv \varphi_c^{v_{|d}} \equiv \bot$*. Now we have to show that each* $p \in P$ *is strongly justified in v. The only parent of c in P is d. Since* $d \in P$*,* $v(d) = \mathbf{f}$ *and d is also strongly deniable in v, it holds that c is strongly deniable in v.*

*To show that b is not strongly justified in v, since* $v(b) = \mathbf{t}$*, we show that b is not strongly acceptable in v. Toward a contradiction, assume that b is strongly acceptable in v. Thus, we have to choose a set P of parents of b that satisfies* $\varphi_b^{v_{|P}} \equiv \top$*. Let* $P = par(b)$*. Since* $\varphi_b^{v_{|P}} \not\equiv \top$*, there is no subset of* $par(b)$ *that satisfies the conditions of Definition 3.1 for b. Therefore, b is not strongly acceptable in v.*

Note that in Example 3.2, if we choose a subset of parents of $c$ equal to $P = \{b\}$, then we cannot show that $c$ is strongly deniable in interpretation $v$. While the first condition of strong deniability holds for $c$, i.e., $\varphi_c^{v_{|b}} \equiv \bot$, the second condition does not hold, i.e., $b$ is not strongly acceptable in $v$, as is shown in Example 3.2.This shows the importance of choosing a right set of parents that satisfies the conditions of Definition 3.1 for a queried argument. Furthermore, if we choose a subset of parents of $c$ equal to $\{b, d\}$, then we face with the sane issue since $b$ is not strongly acceptable in $v$. This way you illustrate that it is necessary to allow $P$ to be a subset of $par(a) \setminus E$ in Definition 3.1.

However, there exists an alternative definition for strongly justified arguments, which we present in Definition 3.53, in which there is no need of indicating a set of parents of a queried argument. In order to prove the main result of this section, which is that the set of strongly admissible interpretations forms a lattice, we need some auxiliary results that are proven based on the current definition of an argument being strongly justified in an interpretation, i.e., Definition 3.1.

In Definition 3.3, similar to Definition 2.23, we introduce the concept of strong admissibility of interpretation $v$ of a given ADF, using the notion of strong justifiability of arguments presented in $v$.

**Definition 3.3** *Let* $D = (A, L, C)$ *be an ADF and let v be an interpretation of D. An interpretation v is a* strongly admissible *interpretation if*

*and only if for each $a$ such that $v(a) \in \{\mathbf{t}, \mathbf{f}\}$, we have that $a$ is a strongly justified argument in $v$.*

*The set of all strongly admissible interpretations of ADF $D$ is denoted by $sadm(D)$.*

To clarify the notion of strongly admissible interpretations of ADFs, we continue Example 3.2 in Example 3.4.

**Example 3.4** *Consider ADF of Example 3.2, i.e., $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$, depicted in Figure 2.7. Let $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$. As was shown in Example 3.2, $c$ and $d$ are strongly deniable in $v$. However, $b$ is not strongly justified in $v$. Thus, $v$ is not a strongly admissible interpretation of $D$. However, $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}\}$, $v_2 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$, $v_3 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$ are strongly admissible interpretations of $D$. We show that $b$ is strongly acceptable in $v_3$. To this end, let $P = \{a, c\}$ be a set of parents of $b$. First, it holds that $\varphi_b^{v_3|_P} \equiv \top$. Thus, the first condition is satisfied for $b$. We also have to check whether each parent of $b$ in $P$ is also strongly justified in $v_3$. To this end, we show that $a$ is strongly acceptable in $v_3$ and $c$ is strongly deniable in $v_3$. The latter is obvious by the method that was presented in Example 3.2 to show that $c$ is strongly deniable in $v$. In addition, $\varphi_a^{v_\mathbf{u}} \equiv \top$, thus, $a$ is strongly acceptable in $v_3$. Hence, $b$ and $a$ are strongly justified in $v_3$. Thus, $v_3$ is a strongly admissible interpretation of $D$. Furthermore, $v_3$ is a unique grounded interpretation of $D$.*

In Example 3.4, $c$ is strongly deniable both in $v_2$ and $v_3$, however, $v_2$ presents less information than $v_3$. Interpretation $v_2$ explains that $c$ is strongly deniable in a strongly admissible interpretation, in other words, $c$ is credulously deniable in the grounded interpretation of $D$, since its parent $d$ is strongly deniable in $v_2$. Based on the acceptance condition of $c$, namely $\varphi_c : \neg b \wedge d$, this piece of information about parents of $c$ is enough to convince a user about the truth value of $c$ in a strongly admissible interpretation and the grounded interpretation as well. That is, to convince a user about the truth value of $c$ in the grounded interpretation of $D$, there is no need of further information about the truth values of $a$ and $b$ in the grounded interpretation. We are interested in finding a strongly admissible interpretation with the least amount of information in which the truth value of a queried argument is satisfied.

**Definition 3.5** *Let $D$ be an ADF, let $a$ be an argument and let $v$ be a strongly admissible interpretation of $D$. Interpretation $v$ is called a* witness *of strong justifiability of $a$ in $D$ if $v \in sadm(D)$ and $v(a) \in \{\mathbf{t}, \mathbf{f}\}$. Interpretation $v$ is called a* least witness *of strong justifiability of $a$ if the following conditions hold.*

- *$v$ is a witness of strong justifiability of $a$; and*
- *there is no strongly admissible interpretation $v'$ such that $v'(a) \in \{\mathbf{t}, \mathbf{f}\}$ and $v' <_i v$.*

*The set of all least witnesses of strong justifiability of $a$ in $D$ is denoted by $LWSJ_a$.*

Note that if argument $a$ is strongly justified in an interpretation, then there exists a strongly admissible interpretation that is a least witness of strong justifiability of $a$. Intuitively, the reason is that the number of arguments of a given ADF is finite, so one can guess an interpretation $v$ and check whether it is the least witness of strong justifiability of $a$. More formally, to find a least witness of strong justifiability of $a$, follow the following steps:

1. Guess an interpretation $v$;
2. check whether $v$ is a strongly admissible interpretation. If the answer is *yes*, then go to item 3, else go to item 1.
3. check whether $v(a) \in \{\mathbf{t}, \mathbf{f}\}$. If the answer is *yes*, then go to item 4, else go to item 1.
4. Pick $v'$ where $v' <_i v$ and check if $v'(a) \in \{\mathbf{t}, \mathbf{f}\}$ and $v$ is a strongly admissible interpretation. If the answer is *yes*, then replace $v$ with $v'$ and repeat this item, else pick another $v'$ and repeat this item.
5. If there is no $v'$ that satisfies item 4, then $v$ is a least witness of strong justifiability of $a$.

For instance, in Example 3.4, interpretation $v_2$ is a least witness of strong justifiability (or deniability) of $c$. This is because $v_2$ contains the truth values of $c$ and $d$, and the truth value of $d$ is the necessary and sufficient piece of information needed for denying $c$ in the grounded interpretation.

Note that an argument may have more than one least witness of strong justifiability. For instance, let $D = (\{a, b, c\}, \{\varphi_a : \top, \varphi_b : a \vee c, \varphi_c : \top\})$. Argument $b$ is strongly acceptable in both $v_1 = \{a, b\}$ and $v_2 = \{b, c\}$; furthermore, by Definition 3.5, both $v_1$ and $v_2$ are least witnesses of strong justifiability of $b$ in $D$, i.e., $\text{LWSJ}_b = \{v_1, v_2\}$.

In the following, let $v^* = v^{\mathbf{t}} \cup v^{\mathbf{f}}$. The sets $v^{\mathbf{t}}$, $v^{\mathbf{f}}$, and $v^{\mathbf{u}}$ contain those arguments that $v$ maps to true, false and undecided, respectively, as it is presented in Definition 2.2. Note that the update of an interpretation $v$ with a truth value $x \in \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ for an argument $b$, as it is presented in Definition 2.60, is denoted by $v|_x^b$, where,

$$v|_x^b(a) = \begin{cases} x & \text{for } a = b, \\ v(a) & \text{for } a \neq b. \end{cases}$$

**Definition 3.6** *Let $D$ be an ADF, $a$ be an argument, let $v \in LWSJ_a$ and let $p \in (par(a) \setminus \{a\}) \cap v^*$. Furthermore, let $F = ((anc(p) \setminus \{a\}) \cap v^*) \cup \{p\}$ and let $F' = A \setminus F$. We define $v_{(a,v,p)} = v|_{\mathbf{u}}^{F'}$, that is, $v_{(a,v,p)}(e) = v(e)$ for $e \in F$ and $v_{(a,v,p)}(e) = \mathbf{u}$ for $e \in F'$.*

The notation $v_{(a,v,p)}$, presented in Definition 3.6, stands for a witness of strong justifiability of $p$ constructed based on $a$ and $v$.

**Proposition 3.7** *Let $D$ be an ADF, let $a$ be an argument, let $v \in LWSJ_a$ and let $p \in (par(a) \setminus \{a\}) \cap (v^*)$.*

- $v_{(a,v,p)} <_i v$;
- $v_{(a,v,p)} \in sadm(D)$.

*In particular, $v_{(a,v,p)}$ is a witness of $p$'s strong justifiability, with strictly less information than $v$.*

**Proof**

- By the definition of $v_{(a,v,p)}$, it holds that $v_{(a,v,p)}(e) = v(e)$ for any $e \in F$, and $v_{(a,v,p)}(e) = \mathbf{u}$ for any $e \in F'$. Since $F = ((anc(p) \setminus \{a\}) \cap v^*) \cup \{p\}$, that is $F$ does not contain $a$, it holds that $F \subset v^*$. Thus, $v(a) \in \{\mathbf{t}, \mathbf{f}\}$, while $v_{(a,v,p)}(a) = \mathbf{u}$. Hence, $v_{(a,v,p)} <_i v$.
- In order to show that $v_{(a,v,p)} \in sadm(D)$, we show that if $v_{(a,v,p)}(e) \in \{\mathbf{t}, \mathbf{f}\}$, then $e$ is strongly justified in $v_{(a,v,p)}$. Toward a contradiction assume that there exists $e$ such that $v_{(a,v,p)}(e) \in \{\mathbf{t}, \mathbf{f}\}$, but $e$ is not strongly justified in $v_{(a,v,p)}$. If $v_{(a,v,p)}(e) \in \{\mathbf{t}, \mathbf{f}\}$, then $e \in F$. Thus, $v_{(a,v,p)}(e) = v(e)$. Since $v$ is a strongly admissible interpretation of $D$, by Definition 3.3, for each $e$ with $v(e) \in \{\mathbf{t}, \mathbf{f}\}$, it holds that $e$ is strongly justified in $v$. Thus, for each $e$ with $v(e) \in \{\mathbf{t}, \mathbf{f}\}$ there exists a subset $P_e \subset par(e)$ that satisfies the condition of Definition 3.1 for $e$.

If $v_{(a,v,p)}(e) \in \{\mathbf{t}, \mathbf{f}\}$, then $e \in F = ((anc(p) \setminus \{a\}) \cap v^*) \cup \{p\}$, that is, $e \in anc(p) \cup p$. Thus, $anc(e) \subseteq anc(p) \cup p$. Hence, $P_e \subseteq F$. That is, for each $e \in F$, it holds that $e$ is strongly justified in $v_{(a,v,p)}$. Thus, the assumption that there exists $e$ such that $v_{(a,v,p)}(e) \in \{\mathbf{t}, \mathbf{f}\}$, but $e$ is not strongly justified in $v_{(a,v,p)}$ was wrong. Hence, $v_{(a,v,p)} \in sadm(D)$.

□

Let $v \in \text{LWSJ}_a$ and $p \in (par(a) \setminus \{a\}) \cap v^*$. Since $v_{(a,v,p)}(p) = v(p)$ and by Proposition 3.7, $v_{(a,v,p)} \in sadm(D)$, it holds that $v_{(a,v,p)}$ is a witness of strong justifiability of $p$. Example 3.8 shows how one can construct a witness of strong justifiability of $p$, with $p \in (par(a) \setminus \{a\}) \cap v^*$, based on $v$ and $a$. Note that in Example 3.8 each $v_{(a,v,p)}$, for $p \in (par(a) \setminus \{a\}) \cap v^*$, is a least witness of strong justifiability of $p$, i.e., $v_{(a,v,p)} \in \text{LWSJ}_p$.

In Example 3.8 for reasons of brevity, we use the shortened notion of three-valued interpretations. That is, instead of $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}, d \mapsto \mathbf{f}\}$ we write $v = \{c, \neg d\}$.

**Example 3.8** *Let $D = (\{a, b, c, d\}, \{\varphi_a : b, \varphi_b : e \vee c, \varphi_c : \neg d, \varphi_d : \bot, \varphi_e : b \vee \neg b\})$ be an ADF, depicted in Figure 3.1. We show the least witnesses of strong justifiability of each argument of $D$: $LWSJ_d = \{\{\neg d\}\}$, $LWSJ_c : \{\{c, \neg d\}\}$, $LWSJ_e = \{\{e\}\}$, $LWSJ_b = \{\{e, b\}, \{b, c, \neg d\}\}$, and $LWSJ_a = \{\{e, b, a\}, \{b, c, \neg d, a\}\}$. In this example arguments $a$ and $b$ have more than one least witness of strong justifiability.*

*Since $v_a = \{e, b, a\} \in LWSJ_a$ and $b \in (par(a) \setminus \{a\}) \cap v_a^*$, one can construct $v_{(a,v_a,b)}$ as it is presented in Definition 3.6. By this definition, $F = ((anc(b) \setminus \{a\}) \cap v_a^*) \cup \{b\} = ((\{e, c, d\} \setminus \{a\}) \cap v_a^*) \cup \{b\} = \{e, b\}$ and $F' = \{a, c, d\}$. Thus, $v_{(a,v_a,b)} = \{e, b\}$. As it is shown in Proposition 3.7, $v_{(a,v_a,b)}$ is a witness of strong justifiability of $b$ such that $v_{(a,v_a,b)} <_i v_a$. Furthermore, here $v_{(a,v_a,b)} \in LWSJ_b$.*

*Furthermore, $v'_a = \{b, c, \neg d, a\}$ is another member of $LWSJ_a$ in which $b \in (par(a) \setminus \{a\}) \cap v'^*_a$. Thus, we construct $v_{(a,v'_a,b)}$ as it is presented in Definition 3.6. By this definition it holds that $F = ((anc(b) \setminus \{a\}) \cap v'^*_a) \cup \{b\} = ((\{e, c, d\} \setminus \{a\}) \cap v'^*_a) \cup \{b\} = \{d, c, b\}$ and $F' = \{a, e\}$. Thus, $v_{(a,v'_a,b)} = \{\neg d, b, c\}$. It holds that $v_{(a,v'_a,b)}$ is a member of $LWSJ_b$ such that $v_{(a,v'_a,b)} <_i v'_a$. As we see $v_{(a,v_a,b)} \in LWSJ_b$ and $v_{(a,v'_a,b)} \in LWSJ_b$ are different least witnesses of strong justifiability of $b$ in $D$.*

*Moreover, we show how one can construct a least witness of strong justifiability of $e$ based on $b$ and $v_b = \{b, e\} \in LWSJ_b$. By Definition 3.6, it holds that $F = ((anc(e) \setminus \{b\}) \cap v_b{}^*) \cup \{e\} = ((\{b\} \setminus \{b\}) \cap v_b{}^*) \cup \{e\} = \{e\}$*

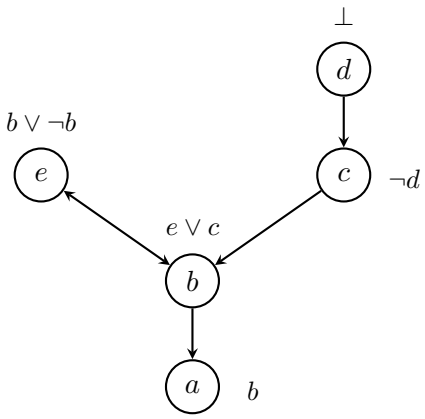


**Figure 3.1:** ADF of Examples 3.8 and 3.12

*and* $F' = \{a, b, c, d\}$. *Thus,* $v_{(b,v_b,e)} = \{e\}$. *It holds that* $v_{(b,v_b,e)}$ *is the unique member of LWSJ*$_e$ *for which* $v_{(b,v_b,e)} <_i v_b$.

Corollary 3.9 is a direct result of Proposition 3.7. This corollary states that if $v \in \text{LWSJ}_a$, then for each $p \in (par(a) \setminus \{a\}) \cap v^*$, there exists at least one $v_p$ such that $v_p \in \text{LWSJ}_p$ and $v_p \leq_i v_{(a,v,p)} <_i v$, i.e., $v_p$ is a least witness of $p$'s strong justifiability, with strictly less information than $v$.

**Corollary 3.9** *Let* $D$ *be an ADF, let* $a$ *be an argument, let* $v \in LWSJ_a$, *let* $p \in (par(a) \setminus \{a\}) \cap (v^*)$ *and let* $v_{(a,v,p)}$ *be a witness of strong justifiability of* $p$ *constructed based on* $a$ *and* $v$. *Then there exists a* $v_p$ *such that* $v_p \in LWSJ_p$ *and* $v_p \leq_i v_{(a,v,p)} <_i v$.

**Proof** Let $v \in \text{LWSJ}_a$, let $p \in (par(a) \setminus \{a\}) \cap (v^*)$. By Definition 3.5, it holds that $v_{(a,v,p)}(p) = v(p)$. Furthermore, by Proposition 3.7, it holds that $v_{(a,v,p)} \in sadm(D)$. That is, $v_{(a,v,p)}$ is a witness of strong justifiability of $p$ in $D$. Since $v_{(a,v,p)}$ is a witness of strong justifiability of $p$ in $D$, there exists a least witness of strong justifiability of $p$ in $D$, namely $v_p$, such that $v_p \leq_i v_{(a,v,p)}$. By Proposition 3.7, it holds that $v_{(a,v,p)} <_i v$. Thus, there exists a least witness of strong justifiability of $p$ in $D$, namely $v_p$ such that $v_p \leq_i v_{(a,v,p)} <_i v$. □

Example 3.10 presents an instance of ADF in which $v_{(a,v,b)}$ is a witness of strong justifiability of $b$, constructed based on $v$ and $a$, but $v_{(a,v,b)}$ is not a least one, that is, $v_{(a,v,b)} \notin \text{LWSJ}_b$.

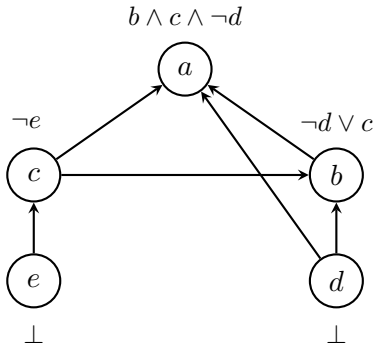


**Figure 3.2:** ADF of Example 3.10

**Example 3.10** *Let* $D = (\{a, b, c, d, e\}, \{\varphi_a : c \wedge b \wedge \neg d, \varphi_b : \neg d \vee c, \varphi_c : \neg e, \varphi_d : \bot, \varphi_e : \bot\})$ *be an ADF, depicted in Figure 3.2. We show the least witnesses of strong justifiability of each argument of* $D$*:* $LWSJ_d = \{\{\neg d\}\}$*,* $LWSJ_e = \{\{\neg e\}\}$*,* $LWSJ_c = \{\{\neg e, c\}\}$*,* $LWSJ_b = \{\{\neg d, b\}, \{\neg e, c, b\}\}$*, and* $LWSJ_a = \{\{\neg d, \neg e, b, c, a\}\}$*. In this example, since* $v_a = \{\neg d, \neg e, b, c, a\} \in LWSJ_a$ *and* $b \in (par(a) \setminus \{a\}) \cap v_a^*$*, one can construct* $v_{(a,v_a,b)}$ *as it is presented in Definition 3.6. By this definition,* $F = ((anc(b) \setminus \{a\}) \cap v_a^*) \cup \{b\} = ((\{d, c, e\} \setminus \{a\}) \cap v_a^*) \cup \{b\} = \{d, c, e, b\}$ *and* $F' = \{a\}$*. Thus,* $v_{(a,v_a,b)} = \{\neg d, \neg e, b, c\}$*. As we see, it holds that* $v_{(a,v_a,b)}$ *is a witness of strong justifiability of* $b$ *in* $D$ *such that* $v_{(a,v_a,b)} <_i v_a$*, which is also shown in Proposition 3.7. However,* $v_{(a,v_a,b)} \notin LWSJ_b$*, i.e.,* $v_{(a,v_a,b)}$ *is not a least witness of strong justifiability of* $b$ *in* $D$*. However, it holds that* $v_b = \{\neg d, b\} \in LWSJ_b$ *and* $v'_b = \{\neg e, c, b\} \in LWSJ_b$ *such that* $v_b <_i v_{(a,v_a,b)} <_i v_a$ *and* $v'_b <_i v_{(a,v_a,b)} <_i v_a$*, as it is presented in Corollary 3.9.*

We now define the *level* of $a$ in a least witness of strong justifiability of $a$; see Definition 3.11.

**Definition 3.11** *Let* $D$ *be an ADF, let* $a$ *be an argument, and let* $v \in LWSJ_a$*. We define* $level_v(a)$ *as follows:*

- *if* $v^* = \{a\}$*, then* $level_v(a) = 1$*;*
- *if* $v^* \neq \{a\}$*, then* $level_v(a) = max\{level_{v_p}(p) \mid p \in (par(a) \setminus \{a\}) \cap v^* \text{ and } v_p \in LWSJ_p \text{ and } v_p <_i v\} + 1$

*We call* $level_v(a)$ *the level of* $a$ *with respect to* $v$*.*

Note that Corollary 3.9 indicates that Definition 3.11 is well-defined. This is because if $v^* \neq \{a\}$, then $(par(a) \setminus \{a\}) \cap v^* \neq \emptyset$. If $v \in \text{LWSJ}_a$

and $p \in (par(a) \setminus \{a\}) \cap v^*$, then there exists at least one $v_p$ such that $v_p \in \text{LWSJ}_p$ and $v_p <_i v$. To clarify the notion of level function of ADFs, we continue Example 3.8 in Example 3.12.

**Example 3.12** *Consider the ADF of Example 3.8, shown in Figure 3.1, i.e.,* $D = (\{a, b, c, d\}, \{\varphi_a : b, \varphi_b : e \vee c, \varphi_c : \neg e, \varphi_d : \bot, \varphi_e : b \vee \neg b\})$. *Since* $v_d = \{\neg d\} \in LWSJ_d$, *it holds that* $v_d^* = \{d\}$. *Thus, by the first item of Definition 3.11, it holds that* $level_{v_d}(d) = 1$.

*Let* $v_c = \{\neg d, c\}$. *It holds that* $v_c \in LWSJ_c$. *Since* $v_c^* = \{c, d\}$, *by the second item of Definition 3.11, it holds that* $level_{v_c}(c) = max\{level_{v_p}(p) \mid p \in (par(c) \setminus \{c\}) \cap v_c^*$ *and* $v_p \in LWSJ_p, v_p <_i v_c\} + 1$. *Since* $par(c) \cap v_c^* = \{d\}$, *it holds that* $level_{v_c}(c) = max\{level_{v_d}(d) \mid d \in par(c) \setminus \{c\} \cap v_c^*, v_d \in LWSJ_d, v_d <_i v_c\} + 1$. *Since* $level_{v_d}(d) = 1$, *it holds that* $level_{v_c}(c) = 2$.

*Let* $v_e = \{e\}$. *By the first item of Definition 3.11, it holds that* $level_{v_e}(e) = 1$.

*Let* $v_b = \{b, c, \neg d\}$. *By the second item of Definition 3.11, it holds that* $level_{v_b}(b) = max\{level_{v_p}(p) \mid p \in (par(b) \setminus \{b\}) \cap v_b^*, v_p \in LWSJ_p, v_p <_i v_b\} + 1$. *Since* $par(b) \cap v_b^* = \{c\}$, *it holds that* $level_{v_b}(b) = max\{level_{v_c}(c) \mid c \in (par(b) \setminus \{b\}) \cap v_b^*$ *and* $v_c \in LWSJ_c, v_c <_i v_b\} + 1$. *Since* $level_{v_c}(c) = 2$, *it holds that* $level_{v_b}(b) = 3$.

*Let* $v_b' = \{b, e\}$. *Since* $level_{v_b'}(b) = max\{level_{v_p}(p) \mid p \in (par(b) \setminus \{b\}) \cap v_b'^*$ *and* $v_p \in LWSJ_p, v_p <_i v_b'\} + 1$ *and* $par(b) \cap v_b'^* = \{e\}$, *it holds that* $level_{v_b'}(b) = max\{level_{v_e}(e) \mid e \in (par(b) \setminus \{b\}) \cap v_b'^*, v_e \in LWSJ_e, v_e <_I v_b'\} + 1$. *Since* $level_{v_e}(e) = 1$, *it holds that* $level_{v_b'}(b) = 2$.

*Let* $v_a = \{a, b, c, \neg d\}$. *Since* $par(a) \cap v_a{}^* = \{b\}$, $v_b <_i v_a$, *and* $level_{v_b}(b) = 3$, *it holds that* $level_{v_a}(a) = 4$. *That is, the level of* $a$ *with respect to* $v_a$ *is* 4.

*Let* $v_a' = \{a, b, e\}$. *Since* $par(a) \cap v_a'{}^* = \{b\}$, $v_b' <_i v_a'$, *and* $level_{v_b'}(b) = 2$, *it holds that* $level_{v_a'}(a) = 3$. *That is, the level of* $a$ *with respect to* $v_a'$ *is* 3.

*This example shows that the level of an argument depends on the given least witness of strong justifiability of the argument in question.*

Intuitively, a least witness $v$ of strong justifiability of $a$ collects exactly the truth values of those ancestors of $a$ that suffice to determine the truth value of $a$. For instance, in Example 3.10, interpretation $v_{(a,v_a,b)} = \{\neg d, \neg e, b, c\}$ is a witness of strong justifiability of $b$ in $D$, however, this interpretation contains extra information to show that $b$ is strongly acceptable in a strongly admissible interpretation of $D$. The information of either $\{\neg e, c, b\}$ or $\{\neg d, b\}$ is enough to show that $b$ is strongly acceptable in a strongly admissible interpretation of $D$. The level of $a$ in a least witness $v$ of strong

justifiability presents the largest distance of $a$ from an initial ancestor of $a$ in $v$ that may have an effect on the truth value of $a$ in a given ADF.

We continue Example 3.10 in Example 3.13 to clarify the notion of level when an argument has more than one least witness of strong justifiability, each of which has less information than a given interpretation. For instance, in Example 3.10, we see that $\text{LWSJ}_b = \{\{\neg d, b\}, \{\neg e, c, b\}\}$ such that $\{\neg d, b\} <_i v_a$ and $\{\neg e, c, b\} <_i v_a$ where $v_a = \{\neg d, \neg e, c, b, a\}$.

**Example 3.13** *Consider the ADF of Example 3.10, shown in Figure 3.2, i.e.,* $D = (\{a, b, c, d, e\}, \{\varphi_a : c \wedge b \wedge \neg d, \varphi_b : \neg d \vee c, \varphi_c : \neg e, \varphi_d : \bot, \varphi_e : \bot\})$. *Since* $v_d = \{\neg d\} \in \mathit{LWSJ}_d$ *and* $v_e = \{\neg e\} \in \mathit{LWSJ}_e$, *it holds that* $\mathit{level}_{v_d}(d) = 1$ *and* $\mathit{level}_{v_e}(e) = 1$.

*Since* $v_c = \{\neg e, c\} \in \mathit{LWSJ}_c$, $e \in \mathit{par}(c) \cap (v_c^*)$, $v_e \in \mathit{LWSJ}_e$, *and* $\mathit{level}_{v_e}(e) = 1$, *it holds that* $\mathit{level}_{v_c}(c) = 2$.

*Since* $v_b = \{\neg d, b\} \in \mathit{LWSJ}_b$, $d \in \mathit{par}(b) \cap (v_b^*)$, $v_d \in \mathit{LWSJ}_d$, *and* $\mathit{level}_{v_d}(d) = 1$, *it holds that* $\mathit{level}_{v_b}(b) = 2$.

*Since* $v_b' = \{\neg e, c, b\} \in \mathit{LWSJ}_b$, $c \in \mathit{par}(b) \cap (v_b^*)$, $v_c \in \mathit{LWSJ}_c$, *and* $\mathit{level}_{v_c}(c) = 2$, *it holds that* $\mathit{level}_{v_b'}(b) = 3$.

*Since* $v_a = \{\neg d, \neg e, c, b, a\} \in \mathit{LWSJ}_a$, $(\mathit{par}(a) \setminus \{a\}) \cap (v_a^*) = \{b, c, d\}$, $v_b, v_b' \in \mathit{LWSJ}_b$, $v_b <_i v_a$ *and* $v_b' <_i v_a$, *and* $\mathit{level}_{v_b}(b) = 2$, $\mathit{level}_{v_b'}(b) = 3$, *it holds that* $\mathit{level}_{v_a}(a) = \mathit{max}\{\mathit{level}_{v_b}(b), \mathit{level}_{v_b'}(b'), \mathit{level}_{v_c}(c), \mathit{level}_{v_d}(d)\} + 1 = 4$.

In Lemma 3.14 we show that if $v \in \text{LWSJ}_a$, then $\text{level}_v(a)$ is at most $|v^*|$.

**Lemma 3.14** *Let* $D$ *be an ADF, let* $a$ *be an argument, and let* $v \in \mathit{LWSJ}_a$. *It holds that* $\mathit{level}_v(a) \leq |v^*|$.

**Proof** Assume that $v \in \text{LWSJ}_a$. Since $v^* \neq \emptyset$, it holds that $|v^*| > 0$. We prove the lemma by induction on $|v^*|$. That is, we show that if $|v^*| = n$, then $\text{level}_v(a) \leq n$.

Base case: Let $|v^*| = 1$. We show that $\text{level}_v(a) = 1$. If $|v^*| = 1$ and $v \in \text{LWSJ}_a$, then $v^* = \{a\}$. Thus, by Definition 3.11, $\text{level}_v(a) = 1$.

Induction hypothesis: Let $n > 0$ and assume that if $|v^*| = n$, then $\text{level}_v(a) \leq n$.

Inductive step: We show that this property also holds for $|v^*| = n + 1$. That is, we show that if $|v^*| = n+1$, then $\text{level}_v(a) \leq n+1$. Since $|v^*| > 1$, it holds that $v^* \neq \{a\}$. Thus, by the second item of Definition 3.11, $\text{level}_v(a) = \mathit{max}\{\text{level}_{v_p}(p) \mid p \in \mathit{par}(a) \cap v^* \text{ and } v_p \in \text{LWSJ}_p, v_p <_i v\} + 1$. By the first item of Proposition 3.7, for each $p \in \mathit{par}(a) \cap v^*$ it holds that $v_{(a,v,p)} <_i v$. By Corollary 3.9, for each $p \in \mathit{par}(a) \cap v^*$ and $v_{(a,v,p)}$, i.e., a

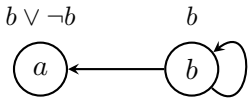


**Figure 3.3:** ADF of Example 3.16

witness of strong justifiability of $p$, constructed based on $a$ and $v$, there exists a $v_p$ such that $v_p \in \text{LWSJ}_p$ and $v_p <_i v$. Thus, it holds that $v_p^* \subset v^*$, that is, $|v_p^*| < |v^*|$. By the induction hypothesis, for each $p \in par(a) \cap v^*$, it holds that $\text{level}_{v_p}(p) \leq |v_p^*| < |v^*|$. That is, for each $p \in par(a) \cap v^*$, it holds that $\text{level}_{v_p}(p) \leq |v_p^*| < n + 1$. Thus, $\text{level}_v(a) \leq n + 1$.

□

Corollary 3.15 is a direct result of Lemma 3.14.

**Corollary 3.15** *Let $D$ be an ADF, let $a$ be an argument, and let $v_a \in LWSJ_a$. Then $level_{v_a}(a)$ is finite.*

Example 3.16 is an instance of an ADF with a redundant link.

**Example 3.16** *Let $D = (\{a, b\}, \{\varphi_a : b \vee \neg b, \varphi_b : b\})$ be an ADF, depicted in Figure 3.3. We show that $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}\}$ is a strongly admissible interpretation of $D$. To this end, we show that $a$ is strongly acceptable in $v$ with respect to $E = \{a\}$. Now let a subset of parents of $P = \emptyset$. Thus, $v_{|_p} = v_{\mathbf{u}}$, i.e., the trivial interpretation. It is clear that $\varphi_a^{v_{\mathbf{u}}} \equiv \top$, i.e., the evaluation of the acceptance condition of $a$ under the trivial interpretation, is irrefutable. Thus, $a$ is strongly acceptable in $v$. Furthermore, $v$ is a least witness of strong justifiability (acceptability) of $a$ and by Definition 3.11, the level of $a$ in $v$ is 1, since $v^* = \{a\}$. Note that $v' = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}\}$ is an admissible interpretation of $D$. However, it is not a strongly admissible interpretation of $D$, since $b$ is not strongly acceptable in $v'$.*

Example 3.17 presents the associated ADF $D_F$ to the AF $F$ presented in Example 2.33.

**Example 3.17** *Consider AF $F = (\{a, b, c, d\}, \{(a, b), (c, d), (d, c)\})$, presented in Example 2.33, and the associated ADF $D_F = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : \neg a, \varphi_c : \neg d, \varphi_d : \neg c\})$. The set of all strongly admissible interpretations of $D_F$ is as follows: $sadm(D_F) = \{\{\}, \{a\}, \{a, \neg b\}\}$. We show that, interpretation $v = \{a, \neg b\}$ is a strongly admissible interpretation of $D_F$.*

*To substantiate our claim, we show that $a$ is strongly acceptable in $v$ and $b$ is strongly deniable in $v$. The former one is clear, since $\varphi_a^{v_{\mathbf{u}}} \equiv \top$. For the latter one, let $P = \{a\}$ be the set of parents of $b$. First, it holds that $\varphi_b^{v_{|a}} \equiv \bot$; second, $a$ is strongly acceptable in $v$. Thus, $b$ is strongly deniable in $v$. Hence, $v$ is a strongly admissible interpretation of $D$. Furthermore, $v$ is a least witness of strong justifiability (deniability) of $b$ and the level of $b$ in $v$ is 2, since $a \in par(b)$ and the level of $a$ in $v_a = \{a\}$ is 1.*

Lemma 3.18 presents the monotonic characteristic of strongly justified arguments, i.e., if $a$ is strongly justified in $v$ and $v \leq_i v'$, then $a$ is strongly justified in $v'$.

**Lemma 3.18** *Let $D$ be an ADF. If $a \in A$ is strongly justified in interpretation $v$ of $D$ and $v \leq_i v'$, then $a$ is also strongly justified in $v'$.*

**Proof** Assume that $a$ is strongly justified in $v$, thus, either $a$ is strongly acceptable in $v$ or it is strongly deniable in $v$. We show the lemma for the case that $a$ is strongly acceptable in $v$; the proof method for the case that $a$ is strongly deniable in $v$ is similar. Assume that $v$ is also a least witness of strong justifiability of $a$. We complete the proof by induction on the level of argument $a$ in $v$.

Base case: let $a$ be an argument of the level 1 that is strongly acceptable with respect to $v$. Therefore, $\varphi_a^{v_{\mathbf{u}}} \equiv \top$. Thus, $a$ is clearly strongly acceptable with respect to $v'$.

Induction hypothesis: Assume that the property holds for each argument of the level $j$ with $1 \leq j < i$ in $v$, i.e., if $a$ is an argument with the level $j$ in $v$ and $a$ is strongly acceptable in $v$, then $a$ is strongly acceptable in $v'$.

Inductive step: We show that this property also holds for arguments of level $i$. That is, if $a$ is an argument with the level $i$ in $v$ and $a$ is strongly acceptable in $v$, then $a$ is strongly acceptable in $v'$. Let $a$ be an argument of the level $i$. Since $a$ is strongly acceptable in $v$, by Lemma 3.14, the level $i$ of $a$ is a finite number. Since $a$ is strongly acceptable in $v$, there exists a set of parents $P$ of $a$ excluding $a$ where $a$ is acceptable with respect to $v_{|_P}$ and all $p \in P$ are strongly justified in $v$. Since $v \leq_i v'$, it holds that $P \subseteq v'^{\mathbf{t}} \cup v'^{\mathbf{f}}$. Thus, it holds that $a$ is acceptable with respect to $v'_{|_P}$. We have to show that each $p \in P$ is strongly justified in $v'$. By Corollary 3.15, the level of each $p \in P$ is at most $i-1$ in $v$. Thus, by induction hypothesis, $p$ is strongly justified in $v'$. Therefore, the second condition of strong acceptability of $a$ in $v'$ also holds. Thus, $a$ is strongly acceptable in $v'$.

□

A sequence of interpretations for a given ADF $D$, each member of which is strongly admissible, is presented in Lemma 3.19. In Proposition 3.20, it is shown that the maximum element of this sequence is the grounded interpretation of $D$.

**Lemma 3.19** *Let $D$ be a finite ADF, let $v_0 = v_{\mathbf{u}}$ and let $v_i = \Gamma_D(v_{i-1})$ for $i > 0$. For each $0 \leq i$ it holds that:*

- $v_i \leq_i v_{i+1}$*;*
- $v_i$ *is a strongly admissible interpretation of $D$.*

**Proof**

- The first item holds because the characteristic operator is a monotonic function.
- We show by induction on $i$ that each $v_i$ is a strongly admissible interpretation.

  Base case: For $i = 0$, it is clear that $v_0 = v_{\mathbf{u}}$ is a strongly admissible interpretation.

  Induction hypothesis: Assume that each $v_j$ for $j$ with $0 \leq j < i$ is a strongly admissible interpretation.

  Inductive step: We show that $v_i$ is a strongly admissible interpretation. Let $a$ be an argument that is assigned to either $\mathbf{t}$ or $\mathbf{f}$ in $v_i$. We show that $a$ is strongly justifiable in $v_i$. If $v_{i-1}(a) \in \{\mathbf{t}, \mathbf{f}\}$, then there is nothing to prove, since by the induction hypothesis $v_{i-1}$ is a strongly admissible interpretation. Thus, $a$ is strongly justified in $v_{i-1}$, and since $v_{i-1} \leq_i v_i$, by Lemma 3.18, $a$ is strongly justifiable in $v_i$.

  Assume that $a \mapsto \mathbf{t} \in v_i$ and $a \mapsto \mathbf{u} \in v_{i-1}$. We show that $a$ is strongly acceptable in $v_i$. For the case that $a \mapsto \mathbf{f} \in v_i$, the proof follows a similar method. Since $v_i(a) = \mathbf{t}$, we can conclude that $\varphi_a^{v_{i-1}}$ is irrefutable. Let $P$ be a subset of parents of $a$ the truth value of which appears in $v_{i-1}$ and $\varphi_a^{v_{i-1}|P} \equiv \top$. Otherwise, $\varphi_a^{v_{i-1}}$ cannot be irrefutable. Assume that $P \neq \{\}$, otherwise, there is nothing to prove. Let $p \in P$. Since $v_{i-1}(p) \in \{\mathbf{t}, \mathbf{f}\}$ for each $p \in P$, $p$ is strongly justifiable in $v_i$ by the induction hypothesis. Thus, by Lemma 3.18, $p$ is strongly justifiable in $v_i$. Thus, both conditions of

Definition 3.1 hold for $a$ in $v_i$. Therefore, for an arbitrary argument $a$, if $v_i(a) \in \{\mathbf{t}, \mathbf{f}\}$, then it holds that $a$ is strongly justifiable in $v_i$. Thus, $v_i$ is a strongly admissible interpretation. Hence, every interpretation in the sequence $v_{\mathbf{u}}, \Gamma_D(v_{\mathbf{u}}), \ldots$ is a strongly admissible interpretation.

□

**Proposition 3.20** *Let $D$ be an ADF.*

- *$D$ has at least one strongly admissible interpretation.*
- *The least strong admissible interpretation of $D$, with respect to the $\leq_i$ ordering, is the trivial interpretation.*
- *The maximal strongly admissible interpretation in the sequence of interpretations as in Lemma 3.19, with respect to the $\leq_i$ ordering, is the unique grounded interpretation of $D$.*

**Proof**

- The first and the second item of the lemma are clear by Lemma 3.19, which says that $v_{\mathbf{u}}$ is a strongly admissible interpretation.
- By definition, the grounded interpretation of $D$ is the least fixed-point of the characteristic operator. By Lemma 3.19, each element of the sequence $\Gamma_D^n(v_{\mathbf{u}})$, for $n$ with $n > 0$, is a strongly admissible interpretation. Since the number of arguments is finite, this sequence has a limit; that is, there exists an $m$ with $m > 0$ where $\Gamma_D^m(v_{\mathbf{u}}) = \Gamma_D^{m+1}(v_{\mathbf{u}})$. Therefore, the limit of this sequence, namely $\Gamma_D^m(v_{\mathbf{u}})$, which is the grounded interpretation of $D$, is also a strongly admissible interpretation. Note that the $n$th power of $\Gamma_D$ is defined inductively, that is, $\Gamma_D^{n+1} = \Gamma_D(\Gamma_D^n)$.

□

In Theorem 3.21, we show that each strongly admissible interpretation is an admissible interpretation as well as conflict-free. However, the other direction of the following theorem does not work.

**Theorem 3.21** *Let $D = (A, L, C)$ be an ADF and let $v$ be a strongly admissible interpretation of $D$. Then the following hold:*

- *$v$ is an admissible interpretation of $D$.*
- *$v$ is a conflict-free interpretation of $D$.*

**Proof**

- Let $v$ be a strongly admissible interpretation of $D$. We show that $v$ is an admissible interpretation. Toward a contradiction, assume that $v$ is not an admissible interpretation, that is, $v \not\leq_i \Gamma_(D)(v)$. Therefore, there exists an $a$ such that either $v(a) = \mathbf{t}$ but $\Gamma_D(v)(a) \neq \mathbf{t}$, or $v(a) = \mathbf{f}$ but $\Gamma_D(v)(a) \neq \mathbf{f}$. By the assumption, $v$ is a strongly admissible interpretation. That is, if $v(a) \in \{\mathbf{t}, \mathbf{f}\}$, then $a$ is strongly acceptable/deniable in $v$. Thus, there exists a subset of parents $P$ of $a$ such that $a$ is acceptable with respect to $v_{|_P}$ if $v(a) = \mathbf{t}$, and $a$ is deniable with respect to $v_{|_P}$ if $v(a) = \mathbf{f}$. However, $\varphi_a^{v_{|_P}} \equiv \top$ implies that $\varphi_a^v$ is irrefutable and $\varphi_a^{v_{|_P}} \equiv \bot$ implies that $\varphi_a^v$ is unsatisfiable. The former implies that if $v(a) = \mathbf{t}$, then $\Gamma_D(v)(a) = \mathbf{t}$; the latter one implies that if $v(a) = \mathbf{f}$, then $\Gamma_D(v)(a) = \mathbf{f}$. This contradicts the assumption that there exists an $a$ such that either $v(a) = \mathbf{t}$ and $\Gamma_D(v)(a) \neq \mathbf{t}$, or $v(a) = \mathbf{f}$ and $\Gamma_D(v)(a) \neq \mathbf{f}$. Thus, the assumption that $v$ is not an admissible interpretation is wrong. Hence, if $v$ is a strongly admissible interpretation, then it is also an admissible interpretation.
- If $v$ is a strongly admissible interpretation, then by the first item of this theorem, it is an admissible interpretation. By the fact that in ADFs every admissible interpretation is a conflict-free interpretation, based on the definition of conflict-freeness (presented in Definition 2.47), we conclude that $v$ is a conflict-free interpretation, as well.

□

Example 3.22 indicates the distinction between the notion of strong admissibility semantics of ADFs and the notions of admissible and conflict-free semantics of ADFs.

**Example 3.22** *let $D = (\{a, b\}, \{\varphi_a : \neg b \vee a, \varphi_b : \neg a\})$ be a given ADF. The interpretation $v = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}\}$ is an admissible interpretation of $D$. However, $a$ is not strongly deniable, nor is $b$ strongly acceptable in $v$. Thus, $v$ is not a strongly admissible interpretation of $D$. Furthermore, $v' = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}\}$ is a conflict-free interpretation of $D$ that is neither*

*an admissible nor a strongly admissible interpretation. The only strongly admissible interpretation of $D$, which is also the grounded interpretation of $D$, is the trivial interpretation.*

### 3.2.1 Lattice Structure

Although the sequence of interpretations presented in Lemma 3.19 produces a sequence of strongly admissible interpretations of a given ADF $D$, this sequence does not contain all of the strongly admissible interpretations of $D$. For instance, in ADF $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$, presented in Example 3.2, it holds that $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$ is a strongly admissible interpretation of $D$. However, $v$ is not equal to any of the elements of the sequence $v_{\mathbf{u}}, \Gamma_D(v_{\mathbf{u}}), \ldots$ (for $D$), as in Lemma 3.19. In this section we show that the strongly admissible interpretations of an ADF form a lattice.

Theorem 3.23 indicates that any strongly admissible interpretation of an ADF $D$ is bounded by the grounded interpretation of $D$.

**Theorem 3.23** *Let $D$ be an ADF, let $w$ be a strongly admissible interpretation of $D$, let $v_i$ for $0 \leq i$ be the sequence of interpretations presented in Lemma 3.19, and let $g$ be the grounded interpretation of $D$. Then, $w \leq_i g$.*

**Proof** Let $w^* = w^{\mathbf{t}} \cup w^{\mathbf{f}} = \{a_1, \ldots, a_n\}$. For each $i$ with $1 \leq i \leq n$, since $a_i$ is strongly justified in $w$, we have that there exists $w_i \in \mathrm{LWSJ}_{a_i}$. Let $l = max\{\mathrm{level}_{w_i}(a_i) \mid a_i \in w^*, w_i \in \mathrm{LWSJ}_{a_i}, w_i \leq_i w\}$. We show that for all $a_i \in w^*$, if $\mathrm{level}_{w_i}(a_i) = k$ for $k$ with $1 \leq k \leq n$, then $g(a_i) \in \{\mathbf{t}, \mathbf{f}\}$, that is, $a_i$ is strongly justified in $g$. We do so by induction on $\mathrm{level}_{w_i}(a_i)$.

Base case: Let $\mathrm{level}_{w_i}(a_i) = 1$. We show that $g(a_i) \in \{\mathbf{t}, \mathbf{f}\}$. Since $\mathrm{level}_{w_i}(a_i) = 1$, the first item of Definition 3.11 says that $w_i^* = \{a_i\}$. Thus, $\varphi_{a_i}^{v_{\mathbf{u}}} \equiv \top$ or $\varphi_{a_i}^{v_{\mathbf{u}}} \equiv \bot$, that is, $\Gamma_D(v_0)(a) \in \{\mathbf{t}, \mathbf{f}\}$. Hence, $g(a) \in \{\mathbf{t}, \mathbf{f}\}$.

Induction hypothesis: For each $a_i \in w^*$, if $\mathrm{level}_{w_i}(a_i) = j$ with $1 \leq j \leq k < l$, it holds that $g(a_i) \in \{\mathbf{t}, \mathbf{f}\}$.

Induction step: Let $a_i \in w^*$ such that $\mathrm{level}_{w_i}(a_i) = k+1$. We show that $g(a_i) \in \{\mathbf{t}, \mathbf{f}\}$. Since $a_i$ is strongly justified in $w_i$, there exists a non-empty subset $P \subseteq par(a_i) \setminus \{a_i\}$ such that $\varphi_{a_i}^{w_{i|_P}} \equiv \top$ or $\varphi_{a_i}^{w_{i|_P}} \equiv \bot$. Since $p$, with $p \in P$, is a parent of $a_i$, by Definition 3.1, $p$ is also a strongly justified argument in $w_i$. By Corollary 3.15 the level of $a_i$ and the level of each $p$, for $p \in P$, is finite. Furthermore, $p \neq a_i$ for each $p$, with $p \in P$. Thus, by Definition 3.11 the level of each $p$ in its associated $v_p \in \mathrm{LWSJ}_p$ is strictly less than the level of $a_i$ in $w_i$, i.e. $\mathrm{level}_{v_p}(p) \leq k$. Then, by the induction

hypothesis, $g(p) \in \{\mathbf{t}, \mathbf{f}\}$, for each $p \in P$. Therefore, $\varphi_{a_i}^{w_{i_{|P}}} \equiv \varphi_{a_i}^{g_{|P}}$. Since $g$ is the grounded interpretation of $D$, it holds that if $\varphi_{a_i}^{g_{|P}} \equiv \top$, then $g(a_i) = \mathbf{t}$, and if $\varphi_{a_i}^{g_{|P}} \equiv \bot$, then $g(a_i) = \mathbf{f}$.

That is, for all $a_i \in w^*$, it holds that $a_i$ is strongly justified in $g$. Thus, $w \leq_i g$.

□

Corollary 3.24 is a direct result of Theorem 3.23.

**Corollary 3.24** *Let $D$ be an ADF, let $w$ be a strongly admissible interpretation of $D$, and let $v_i$ for $0 \leq i$ be the sequence of interpretations presented in Lemma 3.19. Then there exists a least $m \geq 0$ such that $w \leq_i v_m$.*

**Proof** Let $w$ be a a strongly admissible interpretation of $D$, and let $g$ be the grounded interpretation of $D$. We show that set $X = v_j | w \leq v_j$ is non-empty, where $v_j$ is an interpretation in the sequence of interpretations as presented in Lemma 3.19. Based on Theorem 3.23, it holds that $w \leq_i g$. Furthermore, based on Proposition 3.20 grounded interpretation $g$ is the maximal strong admissible interpretation in the sequence of interpretations as in Lemma 3.19. Thus, $X \neq \emptyset$. By the fact that every non-empty subset of the natural numbers has a minimum and the fact that elements of $X$ form a chain, we get a least element of $X$. Thus, there exists a least $m \geq 0$ such that $w \leq_i v_m$. □

To show that the set of strongly admissible interpretations of a given ADF form a lattice, first, in Theorem 3.28 we show that every two strongly admissible interpretations of $D$ have a unique supremum. To this end, we first introduce the notion of *join* of two strongly admissible interpretations; see Definition 3.25.

**Definition 3.25** *Let $D$ be an ADF and let $v$ and $w$ be two strongly admissible interpretations of $D$. The* join *$v \sqcup_i w$ is defined as*

$$v \sqcup_i w(a) = \begin{cases} v(a) & \text{if } v(a) \in \{\mathbf{t}, \mathbf{f}\}, \\ w(a) & \text{if } w(a) \in \{\mathbf{t}, \mathbf{f}\}, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

**Proposition 3.26** *The join of two strongly admissible interpretations of $D$ is a well-defined function.*

**Proof** Let $D$ be an ADF and let $v$ and $w$ be two strongly admissible interpretations of $D$. We show that the join operator is a well-defined function. That is, we have to show that there is no $a$ that has two different values via $(v \sqcup_i w)(a)$. Toward a contradiction, assume that there is an $a$ that has two different outputs via $(v \sqcup_i w)(a)$. That is, $a$ is assigned to $\mathbf{t}$ in one of the interpretations and to $\mathbf{f}$ in another one. For instance, $v(a) = \mathbf{t}$ and $w(a) = \mathbf{f}$. By Corollary 3.24, there exist the least natural numbers $k$ and $m$ such that $v \leq_i v_k$ and $w \leq_i v_m$, respectively. Since $v \leq_i v_k$ and $v(a) = \mathbf{t}$, $a \mapsto \mathbf{t} \in v_k$. Furthermore, since $w \leq_i v_m$ and $w(a) = \mathbf{f}$, $a \mapsto \mathbf{f} \in v_m$. That is, $v_k \not\leq_i v_m$ and $v_m \not\leq_i v_k$. This is a contradiction by Lemma 3.19, which says that either $v_k \leq_i v_m$ or $v_m \leq_i v_k$, because $v_k$ and $v_m$ are elements of the sequence of interpretations presented in Lemma 3.19. Thus, the assumption that there exists $a$ that is acceptable in a strongly admissible interpretation of $D$ but that is deniable in another strongly admissible of $D$ is wrong. Thus, $v \sqcup_i w$ is a well-defined function. □

Lemma 3.27 presents that the join of two strongly admissible interpretations of a given ADF is also a strongly admissible interpretation of that ADF.

**Lemma 3.27** *Let $D$ be an ADF and let $v$ and $w$ be strongly admissible interpretations of $D$. Then $v \sqcup_i w$ is also a strongly admissible interpretation of $D$.*

**Proof** Toward a contradiction, assume that $v \sqcup_i w$ is not a strongly admissible interpretation of $D$. Thus, there exists an $a$ such that $v \sqcup_i w(a) \in \{\mathbf{t}, \mathbf{f}\}$ but it is not strongly justifiable in $v \sqcup_i w$. By Definition 3.25, either $v(a) \in \{\mathbf{t}, \mathbf{f}\}$ or $w(a) \in \{\mathbf{t}, \mathbf{f}\}$. Since $v$ and $w$ are strongly admissible interpretations, $a$ is strongly justifiable in $v$ or $w$. Since $v \leq_i v \sqcup_i w$ and $w \leq_i v \sqcup_i w$, by Lemma 3.18, $a$ is strongly justifiable in $v \sqcup_i w$. This contradicts the assumption that $a$ is not strongly justifiable in $v \sqcup_i w$. Thus, the assumption that $v \sqcup_i w$ is not a strongly admissible interpretation was wrong. That is, the join of two strongly admissible interpretations of $D$ is a strongly admissible interpretation of $D$. □

**Theorem 3.28** *Let $D$ be an ADF. Every two strongly admissible interpretations of $D$ have a unique supremum.*

**Proof** Let $D$ be an ADF and let $v$ and $w$ be two strongly admissible interpretations of $D$. We show that $v \sqcup_i w$ is a supremum of $v$ and $w$. By

Definition 3.25, $v \sqcup_i w$ is an upper bound of $v$ and $w$. By Lemma 3.27, $v \sqcup_i w$ is a strongly admissible interpretation of $D$. It remains to show that $v \sqcup_i w$ is a least upper bound of $v$ and $w$. Toward a contradiction, assume that $v \sqcup_i w$ is not the least upper bound of $v$ and $w$. That is, there exists a strongly admissible interpretation $w'$ of $D$ such that $v \leq_i w'$, $w \leq_i w'$ and $w' <_i v \sqcup_i w$. Thus there exists $a$ with $a \mapsto \mathbf{u} \in w'$ and $(v \sqcup_i w)(a) \in \{\mathbf{t}, \mathbf{f}\}$. However, $(v \sqcup_i w)(a) \in \{\mathbf{t}, \mathbf{f}\}$ implies that either $v(a) \in \{\mathbf{t}, \mathbf{f}\}$ or $w(a) \in \{\mathbf{t}, \mathbf{f}\}$. That is, either $v \not\leq_i w'$ or $w \not\leq_i w'$. This contradicts the assumption that $w'$ is the least upper bound of $v$ and $w$. Thus, the assumption that $v \sqcup_i w$ is not the least upper bound of $v$ and $w$ was wrong. □

Subsequently, to show that the set of strongly admissible interpretations of ADF $D$ form a lattice, in Theorem 3.31 we show that every two strongly admissible interpretations of $D$ have an infimum. To this end, in Definition 3.29, we present the concept of the maximum strongly admissible interpretation contained in an interpretation of $D$.

**Definition 3.29** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Interpretation $w$ is called a* maximum strongly admissible interpretation contained in *$v$ with respect to $\leq_i$ ordering if the following conditions hold:*

1. *$w$ is a strongly admissible interpretation of $D$ such that $w \leq_i v$;*
2. *If $w <_i v$, then there is no strongly admissible interpretation $w'$ of $D$ such that $w <_i w' \leq_i v$.*

**Lemma 3.30** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Then there exists a unique maximum strongly admissible interpretation contained in $v$, with respect to the $\leq_i$ ordering.*

**Proof** Each interpretation of $D$ has at least as much information as the trivial interpretation. Thus, each $v$ of $D$ has at least as much information as $v_{\mathbf{u}}$, which is a strongly admissible interpretation. Since the number of arguments of $D$ is finite, there exists at least one maximal strongly admissible interpretation of $D$ (with respect to the $\leq_i$ ordering), let us call it $w$, contained in a given interpretation $v$. We show that this $w$ is unique. Toward a contradiction, assume that there are two maximal strongly admissible interpretations that satisfy the condition of the lemma, namely $w$ and $w'$. By Lemma 3.27, $w \sqcup_i w'$ is a strongly admissible interpretation of $D$. Since $w \leq_i v$ and $w' \leq_i v$, it also holds that $w \sqcup_i w' \leq_i v$. However, $w \leq_i w \sqcup_i w'$, $w' \leq_i w \sqcup_i w'$ and $w \sqcup_i w' \leq_i v$ together with the assumption

that $w$ and $w'$ are maximal strongly admissible interpretations contained in $v$ lead to $w = w \sqcup_i w'$ and $w' = w \sqcup_i w'$. That is, $w = w'$. Thus, the maximum strongly admissible interpretation which is contained in $v$ is unique. □

**Theorem 3.31** *Let $D$ be an ADF. Every two strongly admissible interpretations of $D$ have a unique infimum.*

**Proof** Let $D$ be an ADF and let $v$ and $v'$ be two strongly admissible interpretations of $D$. Let $w = v \sqcap_i v'$. By Lemma 3.30, there exists a unique maximum strongly admissible interpretation $w'$ with $w' \leq_i w$. That is, $w'$ is a lower bound of $v$ and $v'$. It remains to show that $w'$ is the greatest lower bound of $v$ and $v'$. Toward a contradiction, assume that there exists a $w''$ that is a lower bound of $v$ and $v'$. That is, $w'' \leq_i v$ and $w'' \leq_i v'$. Then by the definition, $w'' \leq_i (v \sqcap_i v' = w)$. By the assumption, $w'$ is the maximum strongly admissible interpretation that is less or equal to $w$, thus, $w'' \leq_i w'$. Thus, $w'$ is an infimum of $v$ and $v'$. □

**Theorem 3.32** *Let $D$ be an ADF. The strongly admissible interpretations of $D$ form a lattice with respect to the $\leq_i$-ordering, with the least element $v_{\mathbf{u}}$ and the top element grd($D$).*

**Proof** We have to show that each pair of strongly admissible interpretations of $D$ has a supremum and an infimum. Theorem 3.28 shows the former one and Theorem 3.31 indicates the latter one. Thus, the strongly admissible interpretations of $D$ form a lattice with respect to the $\leq_i$-ordering. In Proposition3.20, it is shown that $v_{\mathbf{u}}$ is the least strongly admissible interpretation and the grounded interpretation of $D$ is the largest strongly admissible interpretation of the sequence of the interpretations presented in Lemma 3.19. This fact, together with Theorem 3.23, shows that the grounded interpretation $D$ is the greatest element of this lattice. It is trivial that $v_{\mathbf{u}}$ is the least element of this lattice. □

**Corollary 3.33** *The maximal strongly admissible interpretation of ADF $D$, with respect to the $\leq_i$ ordering, is the unique grounded interpretation of $D$.*

The set of strongly admissible interpretations of ADF $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$ given in Example 3.2 form a lattice, depicted in Figure 3.4. The top element of this lattice is the grounded interpretation of $D$ $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\} = \{a, b, \neg c, \neg d\}$.

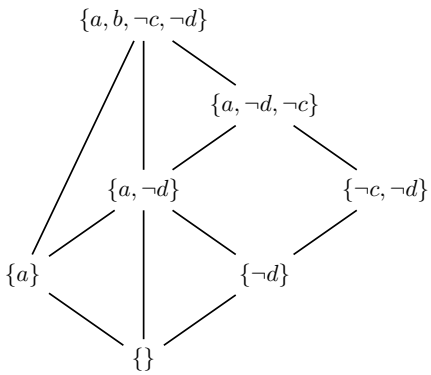


**Figure 3.4:** Complete lattice of the strongly admissible interpretations of the ADF of Example 3.2

## 3.3 Strong Admissibility for ADFs Generalizes Strong Admissibility for AFs

In this section we show that the concept of strong admissibility semantics for ADFs is a proper generalization of the concept of strong admissiblity semantics for AFs (Caminada and Dunne, 2019).

Given an AF $F = (A, R)$ and its corresponding ADF $D_F = (A, R, C)$ (see Definition 2.53), the set of all possible conflict-free extensions of $F$ is denoted by $\mathscr{E}$ and the set of all possible conflict-free interpretations of $D_F$ is denoted by $\mathscr{V}$. The functions $Ext2Int_F$ and $Int2Ext_{D_F}$ in Definitions 3.34–3.36, are modifications of the labelling functions as given in (Baroni et al., 2018a), which we recalled in Definitions 2.34–2.35. Function $Ext2Int_F(S)$ represents the interpretation associated with a given extension $S$ in $F$, and function $Int2Ext_{D_F}(v)$ indicates the extension associated with a given interpretation $v$ of $D_F$.

**Definition 3.34** *Let $F = (A, R)$ be an AF, and let $S$ be an extension of $F$. The truth value assigned to each argument $a \in A$ by the three-valued interpretation $v_S$ associated with $S$ is given by $Ext2Int_F : \mathscr{E} \to \mathscr{V}$ as follows.*

$$Ext2Int_F(S)(a) = \begin{cases} \mathbf{t} & \text{if } a \in S, \\ \mathbf{f} & \text{if } \exists b \in A \text{ such that } (b, a) \in R \text{ and } b \in S, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

**Proposition 3.35** *Let $F = (A, R)$ be an AF, let $D_F$ be its associated ADF, and let $S$ be a conflict-free extension of $F$. Then $Ext2Int_F(S)$ is well-defined.*

**Proof**

1. Assume that $a \in S$. We show that $a$ is only assigned to $\mathbf{t}$ in $Ext2Int_F(S)$. By Definition 3.34, it definitely holds that $a \mapsto \mathbf{t} \in Ext2Int_F(S)$, thus $a \mapsto \mathbf{u} \notin Ext2Int_F(S)$. We show that $a$ cannot assign to $\mathbf{f}$ in $Ext2Int_F(S)$. Toward a contradiction, assume that $a \mapsto \mathbf{f} \in Ext2Int_F(S)$. That is, by Definition 3.34, there exists a parent $p_a$ of $a$ such that $p_a \in S$. However, this means that $S$ contains conflicting arguments, i.e., $a$ and $p_a$ with $(p_a, a) \in R$. Thus, $S$ is not a conflict-free extension. This contradicts the assumption that $S$ is a conflict-free extension of $F$. Thus, the assumption that $a \mapsto \mathbf{f} \in Ext2Int_F(S)$ is wrong.
2. Assume that $a \notin S$. We show that either $a \mapsto \mathbf{f} \in Ext2Int_F(S)$ or $a \mapsto \mathbf{u} \in Ext2Int_F(S)$, but not both. Either at least one parent of $a$ belongs to $S$ or none of them belong to $S$. By Definition 3.34, it is straightforward that if $a \notin S$ and a parent of $a$ belongs to $S$, then $a \mapsto \mathbf{f} \in Ext2Int_F(S)$. In other words, if $a \notin S$ and none of the parents of $a$ belong to $S$, then $a \mapsto \mathbf{u} \in Ext2Int_F(S)$. That is, if $a \notin S$, then either $a \mapsto \mathbf{f} \in Ext2Int_F$ or $a \mapsto \mathbf{u} \in Ext2Int_F(S)$ but not both.

Thus, if $S$ is a conflict-free extension, then $Ext2Int_F(S)$ is well-defined. □

Note that in Definition 3.34, the basic condition that $S$ has to be a conflict-free extension is a necessary condition for $Ext2Int_F(S)$ being well-defined. For instance, let $F = (\{a, b\}, \{(a, b)\})$. Set $S = \{a, b\}$ is an extension of $F$. However, $S$ does not satisfy the conflict-free property. On the other hand, $Ext2Int_F(S) = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, b \mapsto \mathbf{f}\}$. In other words, the correspondence between extensions and interpretations via $Ext2Int_F(.)$ is well-defined for conflict-free sets of arguments. This is the reason why we restrict $\mathscr{E}$ and $\mathscr{V}$ to the set of all conflict-free extensions of $F$ and the set of all conflict-free interpretations of $D_F$, respectively. By Theorem 4 in (Caminada and Dunne, 2019), every strongly admissible extension of an AF is a conflict-free extension. Thus, if $S$ is a strongly admissible extension of AF $F$, then, by Proposition 3.35, $Ext2Int_F(S)$ is well-defined.

So extensions of $F$ can be represented as interpretations of $D_F$. Also an interpretation of $D_F$ can be represented as an extension via the following function.

**Definition 3.36** *Let $D_F = (A, R, C)$ be the ADF associated with AF $F$, and let $v$ be an interpretation of $D_F$, that is, $v \in \mathscr{V}$. The associated extension $S_v$ of $v$ is obtained via application of $Int2Ext_{D_F} : \mathscr{V} \to \mathscr{E}$ on $v$, as follows:*

$$Int2Ext_{D_F}(v) = \{a \in A \mid a \mapsto \mathbf{t} \in v\}$$

To present the correspondence between strongly admissible extensions of $F$ and strongly admissible interpretations of the associated $D_F$, we first present the correspondence between strongly admissible labellings of $F$ and strongly admissible interpretations of the associated $D_F$ in Lemma 3.40. To this end, we define two functions, *Lab2Int* and *Int2Lab* to indicate the correspondence between labellings and interpretations in Definitions 3.37 and 3.38. Note that $\mathscr{L}$ denotes the set of labellings of AF $F$.

**Definition 3.37** *The function $Lab2Int(\cdot) : \mathscr{L} \mapsto \mathscr{V}$ maps three-valued labellings to three-valued interpretations such that*

- $Lab2Int(\lambda)(a) = \mathbf{t}$ *iff* $\lambda(a) = \texttt{in}$,
- $Lab2Int(\lambda)(a) = \mathbf{f}$ *iff* $\lambda(a) = \texttt{out}$, *and*
- $Lab2Int(\lambda)(a) = \mathbf{u}$ *iff* $\lambda(a) = \texttt{undec}$.

**Definition 3.38** *The function $Int2Lab(\cdot) : \mathscr{V} \mapsto \mathscr{L}$ maps three-valued interpretations to three-valued labellings such that*

- $Int2Lab(v)(a) = \texttt{in}$ *iff* $v(a) = \mathbf{t}$;
- $Int2Lab(v)(a) = \texttt{out}$ *iff* $v(a) = \mathbf{f}$;
- $Int2Lab(v)(a) = \texttt{undec}$ *iff* $v(a) = \mathbf{u}$.

In Proposition 3.39 we show that *Int2Lab* is the inverse of *Lab2Int*.

**Proposition 3.39** *$Lab2Int(Int2Lab(.)) = id_{\mathscr{V}}$ and $Int2Lab(Lab2Int(.)) = id_{\mathscr{L}}$.*

**Proof** We show that *Lab2Int* and *Int2Lab* are bijective functions. To this end, we show that *Lab2Int* is a surjective and injective function. Let $\lambda$ be a labelling. We define interpretation $v_\lambda$ as follows:

$$v_\lambda(a) = \begin{cases} \mathbf{t} & \text{if } \lambda(a) = \texttt{in}; \\ \mathbf{f} & \text{if } \lambda(a) = \texttt{out}; \\ \mathbf{u} & \text{if } \lambda(a) = \texttt{undec} \end{cases}$$

By Definition 3.37, it holds that $\mathit{Lab2Int}(\lambda) = v_\lambda$. Thus, $\mathit{Lab2Int}(.)$ is a surjective function.

Toward a contradiction, assume that $\mathit{Lab2Int}(.)$ is not an injective function. That is, there are $\lambda_1, \lambda_2 \in \mathscr{L}$ such that $\mathit{Lab2Int}(\lambda_1) = \mathit{Lab2Int}(\lambda_2)$, and $\lambda_1 \neq \lambda_2$. That is, there exists $a$ such that $\lambda_1(a) \neq \lambda_2(a)$. Thus, by Definition 3.37, $\mathit{Lab2Int}(\lambda_1)(a) \neq \mathit{Lab2Int}(\lambda_2)(a)$, i.e., $\mathit{Lab2Int}(\lambda_1) \neq \mathit{Lab2Int}(\lambda_2)$. This contradicts our assumption. Thus, the assumption that $\mathit{Lab2Int}(.)$ is not an injective function was wrong.

Thus, *Lab2Int* is a bijective function. With a similar method we have that $\mathit{Int2Lab}(.)$ is a bijective function. Thus, *Lab2Int* and *Int2Lab* have inverse functions.

Let $v$ be an interpretation. We show that $\mathit{Lab2Int}(\mathit{Int2Lab}(v)) = v$. 1. If $v(a) = \mathbf{t}$, then by Definition 3.38, $\mathit{Int2Lab}(v)(a) = \texttt{in}$. Then, by definition 3.37, $\mathit{Lab2Int}(\mathit{Int2Lab}(a)) = \mathbf{t}$. 2. If $v(a) = \mathbf{f}$, then by Definition 3.38, $\mathit{Int2Lab}(v)(a) = \texttt{out}$. Then, by Definition 3.37, $\mathit{Lab2Int}(\mathit{Int2Lab}(a)) = \mathbf{f}$. 3. If $v(a) = \mathbf{u}$, then by Definition 3.38, $\mathit{Int2Lab}(v)(a) = \texttt{undec}$. Then, by Definition 3.37, $\mathit{Lab2Int}(\mathit{Int2Lab}(a)) = \mathbf{u}$. Thus, $\mathit{Lab2Int}(\mathit{Int2Lab}(v)) = v$. Hence, $\mathit{Lab2Int}(\mathit{Int2Lab}(.)) = id_{\mathscr{V}}$. With the same method it is easy to show that, $\mathit{Int2Lab}(\mathit{Lab2Int}(.)) = id_{\mathscr{L}}$. Thus, $\mathit{Lab2Int}(.)$ is the inverse of $\mathit{Int2Lab}(.)$. □

**Lemma 3.40** *For any argument framework $F = (A, R)$ and its associated ADF $D_F$, the following properties hold:*

- *if $\lambda$ is a strongly admissible labelling of $F$, then $\mathit{Lab2Int}(\lambda)$ is a strongly admissible interpretation of $D_F$;*
- *if $v$ is a strongly admissible interpretation of $D_F$, then $\mathit{Int2Lab}(v)$ is a strongly admissible labelling of $F$.*

**Proof**

- Assume that $\lambda$ is a strongly admissible labelling of $F$. Let $v = \mathit{Lab2Int}(\lambda)$. By Definition 2.32, a strongly admissible labelling $\lambda$ of $F$ is an admissible labelling whose min-max numbering yields natural numbers only. We show that $v$ is a strongly admissible interpretation of $D_F$. To this end, we show that if $\lambda(a) \in \{\mathtt{in}, \mathtt{out}\}$, then $a$ is a strongly justified argument in $v$. We show this by induction on the min-max numbering of arguments.

  Base case: If $\lambda(a) \in \{\mathtt{in}, \mathtt{out}\}$ and the min-max numbering of $a$ is 1, then we show that $a$ is a strongly justified in $v$. If the min-max numbering $a$ is 1, then $a$ is an initial argument of $F$ and $D_F$. That is, $\varphi_a^{v_{\mathbf{u}}} \equiv \top$ or $\varphi_a^{v_{\mathbf{u}}} \equiv \bot$. Thus, $a$ is strongly acceptable in $v$.

  Induction hypothesis: For all $j$ with $0 \leq j < i$, if $\lambda(a) \in \{\mathtt{in}, \mathtt{out}\}$ and the min-max numbering $a$ is $j$, then $a$ is a strongly justified argument in $v$.

  Inductive step: We show that if $\lambda(a) \in \{\mathtt{in}, \mathtt{out}\}$ and the min-max numbering $a$ is $i$, then $a$ is a strongly justified argument in $v$.

  1. Assume that $\lambda(a) = \mathtt{in}$ and min-max numbering $a$ is $i$. Since $\lambda$ is a strongly admissible labelling of $F$, it is also an admissible labelling of $F$. Thus, $\lambda(a) = \mathtt{in}$ implies that any attacker of $a$, namely $p_a$ is labelled $\mathtt{out}$ in $\lambda$. That is, by Definition 3.37, each attacker of $a$ is assigned to $\mathbf{f}$ in $v$, i.e., $v(p_a) = \mathbf{f}$. Thus, $\varphi_a^v = (\bigwedge_{(b,a)\in R} \neg b)^v \equiv \top$. By Definition 2.30 (min-max numbering) and since $\lambda$ is a strongly admissible labelling of $F$, for each attacker $p_a$, it holds that $\mathscr{MM}_{\mathscr{L}}(p_a) < \mathscr{MM}_{\mathscr{L}}(a)$ (otherwise, $\mathscr{MM}_{\mathscr{L}}$ does not yield natural numbers). Since $\mathscr{MM}_{\mathscr{L}}(p_a) < i$ and $\lambda(p_a) = \mathtt{out}$, by the induction hypothesis, each $p_a$ is strongly deniable in $v$. This implies that the conditions of Definition 3.1 are satisfied for $a$. Thus, $a$ is strongly acceptable in $v$.
  2. By the similar proof method one can check when $\lambda(a) = \mathit{out}$, then $a$ is strongly deniable in $v$.

  Thus, $v$ is a strongly admissible interpretation of $D_F$.

- Let $v$ be a strongly admissible interpretation of $D_F$. Let $\lambda = \mathit{Int2Lab}(v)$. We show that $\lambda$ is a strongly admissible labelling of

$F$. To this end, we show that $\lambda$ is an admissible labelling whose min-max numbering yields natural numbers only.

1. Since by Theorem 3.21 each strongly admissible interpretation is an admissible interpretation, $v$ is an admissible interpretation. Thus, $\lambda = \mathit{Int2Lab}(v)$ is an admissible labelling of $F$.
2. To complete the proof, we show that min-max numbering of $\lambda$ leads to natural numbers only. For each $a$ with $a \in \{\mathbf{t}, \mathbf{f}\}$, let $v_a$ be a least witness of strong admissibility of $a$ (as Definition 3.5), where $v_a \leq_i v$. We show that the level of $a$ in a least $v_a$ is equal to the min-max numbering of $a$ in $\lambda$.

   (a) Assume that $v(a) = \mathbf{t}$ (i.e., $\lambda(a) = \texttt{in}$). Thus, by the acceptance condition of $a$ and since each AF does not have any redundant links, for each parent $p_a$ of $a$ it holds that $v(p_a) = \mathbf{f}$ (i.e., $\lambda(p_a) = \texttt{out}$). Thus, all parents of $a$ are assigned to $\mathbf{f}$ in $v_a$. By Definition 3.11, the level of $a$ in $v_a$, i.e., $\mathrm{level}_{v_a}(a)$ is the level of $p_a$ plus 1 such that $p_a$ has the maximum level among the parents of $a$ in its associated $v_p$ where $v_p \in \mathrm{LWSJ}_p$. This is exactly equal to $\mathscr{MM}_{\mathscr{L}}(a)$ in $\lambda$.

   (b) Assume that $v(a) = \mathbf{f}$ (i.e., $\lambda(a) = \texttt{out}$). Thus, by the acceptance condition of $a$ and Definition 3.1 and since AF does not contain any redundant or dependent links, there exists a parent $p_a$ of $a$ such that $v_a(p_a) = \mathbf{t}$. Thus, the level of $a$ in $v_a$ is $max\{\mathrm{level}_{v_p}(p_a) \mid p_a \in par(a) \cap v_a^{\mathbf{t}}, v_p <_i v_a, v_p \in \mathrm{LWSJ}_p\} + 1$. Let us fix a $p_a$ such that $\mathrm{level}_{v_a}(a) = \mathrm{level}_{v_p}(p_a) + 1$. Since $v(p_a) = \mathbf{t}$, by the previous item, it holds that $\mathrm{level}_{v_{p_a}}(p_a) = \mathscr{MM}_{\mathscr{L}}(p_a)$. Thus, $\mathrm{level}_{v_{p_a}}(p_a) + 1 = \mathscr{MM}_{\mathscr{L}}(p_a) + 1$. That is, $\mathrm{level}_{v_a}(a) = \mathscr{MM}_{\mathscr{L}}(a)$

   Thus, for each $a$ with $v(a) \in \{\mathbf{t}, \mathbf{f}\}$ the level of $a$ in $v_a$ is equal to $\mathscr{MM}_{\mathscr{L}}(a)$ in $\lambda$. Furthermore, by Lemma 3.14, each $a$ has a finite level in $v_a$ since $\mathrm{level}_{v_a}(a) \leq |{v_a}^*|$. Thus, $\lambda$ is a strongly admissible labelling of $F$.

□

Theorem 3.41 is a direct result of Propositions 2.36, 3.39, and Lemma 3.40. It presents the correspondence between strongly admissible extensions of $F$ and strongly admissible interpretations of the associated $D_F$.

**Theorem 3.41** *For any argument framework $F = (A, R)$ and its associated ADF $D_F$, the following properties hold:*

- *If $S$ is a strongly admissible extension of $F$, then $Ext2Int_F(S)$ is a strongly admissible interpretation of $D_F$;*

- *If $v$ is a strongly admissible interpretation of $D_F$, then $Int2Ext_{D_F}(v)$ is a strongly admissible extension of $F$.*

**Proof**

- It is enough to show that $Ext2Int_F(S) = Lab2Int(Ext2Lab(S))$.

  1. Let $a$ be an argument such that $a \in S$. By Definition 3.34, $a \in S$ if and only if $Ext2Int_F(S)(a) = \mathbf{t}$. In other words, $a \in S$ if and only if $Ext2Lab(S)(a) = \texttt{in}$ if and only if $Lab2Int(Ext2Lab(S)(a)) = \mathbf{t}$.

  2. Let $a$ be an argument such that $a \notin S$ and there exists a parent of $a$, namely $p_a$, with $p_a \in S$. By Definition 3.34, $a \notin S$ and $p_a \in S$ if and only if $Ext2Int_F(S)(a) = \mathbf{f}$. In other words, $a \notin S$ and $p_a \in S$ if and only if $Ext2Lab(S)(a) = \texttt{out}$ if and only if $Lab2Int(Ext2Lab(S)(a)) = \mathbf{f}$.

  Thus, $Ext2Int_F(S) = Lab2Int(Ext2Lab(S))$. Hence, by Proposition 2.36, if $S$ is a strongly admissible extension of $F$, then $\lambda = Ext2Lab(S)$ is a strongly admissible labelling of $F$. Furthermore, by Lemma 3.40, if $\lambda$ is a strongly admissible labelling of $F$, then $Lab2Int(\lambda)$ is a strongly admissible interpretation of $D_F$. That is, $Ext2Int_F(S)$ is a strongly admissible interpretation of $D_F$.

- By the similar method as the one presented in the proof of the previous item, we have $Int2Ext_{D_F}(v) = Lab2Ext(Int2Lab(v))$. By Lemma 3.40, if $v$ is a strongly admissible interpretation of $D_F$, then $\lambda = Int2Lab(v)$ is a strongly admissible labelling of $F$. By Proposition 2.36, if $\lambda$ is a strongly admissible labelling of $F$, then $Lab2Ext(\lambda)$ is a strongly admissible extension of $F$. Hence, $Int2Ext_{D_F}(v)$ is a strongly admissible extension of $F$.

□

We have already shown that the projection of a strongly admissible extension/labelling of AF $F$ via $Ext2Int_F(.)/Lab2Int(.)$ is a strongly admissible interpretation of the associated $D_F$. The commutative diagrams in Figure 3.5 show the relation between strong admissibility semantics of AF $F$ and strong admissibility semantics of its associated ADF $D_F$. In the following, the set of strongly admissible extensions of $F$, the set of strongly admissible

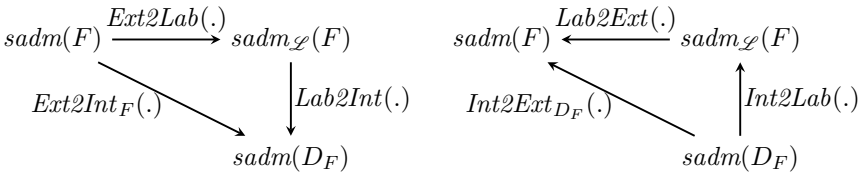


**Figure 3.5:** The left diagram shows that strong admissibility semantics of $F$ project to strong admissibility semantics of $D_F$, via $Ext2Int_F$. The right diagram shows that strong admissibility semantics of $D_F$ project to strong admissibility semantics of $F$, via $Int2Ext_{D_F}$.

labellings of $F$, and the set of strongly admissible interpretations of $D_F$ are denoted by $sadm(F)$, $sadm_{\mathscr{L}}(F)$, and $sadm(D_F)$, respectively.
The direct result of Theorem 3.41 is that the strong admissibility semantics of ADFs form a proper generalization of strong admissibility semantics of AFs, as presented in Corollary 3.42.

**Corollary 3.42** *Let $F$ be an AF and let $D_F$ be its associated ADF. An extension $S$ is a strongly admissible semantics of $F$ if and only if $v = Ext2Int_F(S)$ is a strongly admissible interpretation of $D_F$.*

However, Corollary 3.42 does not claim that there is a one-to-one relation between the set of strong admissibility extensions of $F$ and the set of strong admissibility interpretations of $D_F$. In other words, for the strong admissibility semantics, neither $Ext2Int_F(.)$ nor $Int2Ext_{D_F}(.)$ is a bijective function. The reason is that for the strong admissibility semantics, the function $Ext2Int_F$ is not a surjective function and function $Int2Ext_{D_F}$ is not an injective function, as clarified in Example 3.43.[1]

**Example 3.43** *Let $F = (\{a, b, c, d\}, \{(a, b), (b, c), (c, d)\})$. The associated ADF of $F$ is $D_F = (\{a, b, c, d\}, \{(a, b), (b, c), (c, d)\}, \{\varphi_a : \top, \varphi_b : \neg a, \varphi_c : \neg b, \varphi_d : \neg c\})$. The interpretation $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}, d \mapsto \mathbf{u}\}$ is a strongly admissible interpretation of $D_F$, but it is not a projection of any strongly admissible extension of $F$ via $Ext2Int_F$. Thus, $Ext2Int_F$ is not a surjective function. Furthermore, both of the strongly admissible interpretations of $D_F$, namely, $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}, d \mapsto \mathbf{u}\}$ and $v_2 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}, d \mapsto \mathbf{f}\}$, are mapped to the strongly admissible extension $\{a, c\}$ via $Int2Ext_{D_F}$. Thus, $D_F$ is not an injective*

[1] In (Baroni et al., 2018a; Caminada and Dunne, 2019), it is shown that *Ext2Lab* is not a a surjective function and function *Lab2Ext* is not an injective function, for strong admissibility semantics of a given AF.

*function. In other words, we did not claim that* $Ext2Int_F$ *is an inverse of* $Int2Ext_{D_F}$ *for strongly admissible semantics. For instance, in this example* $Ext2Int_F(Int2Ext_{D_F}(v_1)) = v_2$.

Although *Int2Ext*(.) is not an injective function, by the second item of Theorem 3.41, function *Int2Ext*(.) maps any strongly admissible interpretation of $D_F$ to a strongly admissible extension of $F$. That is, *Int2Ext*(.) may map an element of $sadm(D_F)$ to an element of $sadm(F)$. On the other hand, it is possible that there exists an element of $sadm(D_F)$ that is not an image of any element of $sadm(F)$ by *Ext2Int*(.), since *Ext2Int*(.) is not a surjective function. However, the image of any element of $sadm(F)$ by *Ext2Int*(.) is a strongly admissible interpretation of $D_F$, by the first item of Theorem 3.41. These results together lead to Corollary 3.44.

**Corollary 3.44** *The concept of strong admissibility semantics of ADFs is a generalization of the concept of strong admissibility semantics of AFs.*

## 3.4 Algorithm for Strong Admissibility Semantics of ADFs

Definition 3.3 says that an interpretation $v$ is strongly admissible in a given ADF $D$ if and only if for each argument $a$ that is presented in $v$, i.e., $v(a) \in \{\mathbf{t}, \mathbf{f}\}$, we have that $a$ is a strongly justified argument in $v$. In the following, we present an alternative method to answer the verification problem under strong admissibility semantics of ADFs. In this method of investigating whether the given interpretation $v$ is a strongly admissible interpretation of $D$, there is no need to check whether each argument is strongly justified in $v$. The results of this section lead to algorithms to answer the following decision problems.

1. The verification problem: whether a given interpretation is a strongly admissible interpretation of a given ADF.
2. The strong justification problem: whether a given argument $a$ is a strongly justified argument in a given interpretation.

To this end, we introduce $\Gamma_{D,v}$ a variant of the characteristic operator restricted to a given interpretation $v$, presented in Definition 3.45.

**Definition 3.45** *Let* $D = (A, L, C)$ *be a given ADF and let* $v, w$ *be interpretations of* $D$. *Let* $\Gamma_{D,v}(w) = \Gamma_D(w) \sqcap_i v$ *where* $\Gamma^n_{D,v}(w) = \Gamma_{D,v}(\Gamma^{n-1}_{D,v}(w))$

*for $n$ with $n \geq 1$. Note that $\Gamma^0_{D,v}(w) = w$. We call the collection of the interpretations of $\Gamma^n_{D,v}(v_{\mathbf{u}})$ for $n \geq 1$, the set of interpretations constructed based on $v$ in $D$.*

In Lemma 3.46, we show that each interpretation in the set of interpretations constructed based on $v$ is a strongly admissible interpretation.

**Lemma 3.46** *Let $D = (A, L, C)$ be a given ADF and let $v$ be an interpretation of $D$. Let $\Gamma^n_{D,v}(v_{\mathbf{u}})$ be the sequence of interpretations constructed based on $v$, as in Definition 3.45. For each $i$ with $i \geq 0$ it holds that;*

- $\Gamma^i_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^{i+1}_{D,v}(v_{\mathbf{u}})$*;*
- $\Gamma^i_{D,v}(v_{\mathbf{u}})$ *is a strongly admissible interpretation of $D$;*
- *if $\Gamma^i_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$, then $a$ is strongly justifiable in $\Gamma^i_{D,v}(v_{\mathbf{u}})$.*

**Proof**

- We show the first item by induction on $i$.

  Base case: By Definition 3.45, for $i = 0$ it holds that $\Gamma^0_{D,v}(v_{\mathbf{u}}) = v_{\mathbf{u}}$ and it is clear that $v_{\mathbf{u}} \leq_i \Gamma^1_{D,v}(v_{\mathbf{u}})$.

  Induction hypothesis: Suppose that $\Gamma^j_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^{j+1}_{D,v}(v_{\mathbf{u}})$ for each $j$ with $0 \leq j < i$.

  Inductive step: We show that this property holds for $j = i$, i.e., $\Gamma^i_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^{i+1}_{D,v}(v_{\mathbf{u}})$. From the fact that the characteristic operator is monotonic together with the induction hypothesis, it follows that $\Gamma_D(\Gamma^j_{D,v}(v_{\mathbf{u}})) \leq_i \Gamma_D(\Gamma^{j+1}_{D,v}(v_{\mathbf{u}}))$, for $j$ with $0 \leq j < i$. Thus, $\Gamma_D(\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})) \leq_i \Gamma_D(\Gamma^i_{D,v}(v_{\mathbf{u}}))$ and further, $\Gamma_D(\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})) \sqcap_i v \leq_i \Gamma_D(\Gamma^i_{D,v}(v_{\mathbf{u}})) \sqcap_i v$. That is, $\Gamma^i_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^{i+1}_{D,v}(v_{\mathbf{u}})$.

- We show the second item by induction on $i$.

  Base case: For $i = 0$ it is clear that $\Gamma^0_{D,v}(v_{\mathbf{u}}) = v_{\mathbf{u}}$ is a strongly admissible interpretation of $D$.

  Induction hypothesis: Suppose that for each $j$ with $0 \leq j < i$ it holds that $\Gamma^j_{D,v}(v_{\mathbf{u}})$ is a strongly admissible interpretation of $D$.

  Inductive step: We prove that $\Gamma^i_{D,v}(v_{\mathbf{u}})$ is also a strongly admissible interpretation of $D$. To this end, we show that if the truth value of $a$ is presented in $\Gamma^i_{D,v}(v_{\mathbf{u}})$, then $a$ is a strongly justified argument in $\Gamma^i_{D,v}(v_{\mathbf{u}})$. We assume that $\Gamma^i_{D,v}(v_{\mathbf{u}})(a) = \mathbf{t}$ and $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})(a) = \mathbf{u}$,

otherwise if the truth value of $a$ is presented in $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$, then $a$ is strongly acceptable in $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$ and there is nothing to prove (because by induction hypothesis $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$ is a strongly admissible interpretation of $D$). We show that $a$ is strongly acceptable in $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$. For the case that $\Gamma^{i}_{D,v}(v_{\mathbf{u}})(a) = \mathbf{f}$ the proof follows a similar method. Since $\Gamma^{i}_{D,v}(v_{\mathbf{u}})(a) = \mathbf{t}$, we can conclude that $\Gamma_D(\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})) = \mathbf{t}$, that is, the evaluation of $\varphi_a$ under $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$ is irrefutable. Thus, there exists a non-empty subset of parents of $a$, namely $P$ such that the truth value of each $p \in P$ is presented in $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$. Since by induction hypothesis $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$ is strongly admissible if the truth value of an argument is presented in $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$, then that argument is strongly justified in $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$. That is, each $p \in P$ is also strongly justified in $\Gamma^{i-1}_{D,v}(v_{\mathbf{u}})$. This satisfies the conditions of Definition 3.1. Thus, every interpretation in the sequence $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ for $i$ with $i \geq 0$ is a strongly admissible interpretation.

- Toward a contradiction, assume that there exists an $i$ and an argument $a$ such that $\Gamma^{i}_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$ but $a$ is not a strongly justified argument in $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$. Thus, by Definition 3.3, $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ is not a strongly admissible interpretation of $D$. This contradicts the second item of the current lemma, in which we showed that for each $i$ with $i \geq 0$, it holds that $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ is a strongly admissible interpretation of $D$. Thus the assumption that there exists an argument $a$ such that $\Gamma^{i}_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$ but $a$ is not a strongly justified argument in $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ was wrong. Thus, if $\Gamma^{i}_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$, then $a$ is a strongly justified argument in $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$.

□

**Remark 3.46.1** *The sequence of interpretations $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ as in Definition 3.45, is named the sequence of strongly admissible interpretations constructed based on $v$ in $D$.*

Proposition 3.47 presents that for each interpretation $v$ the sequence of interpretations constructed based on $v$ has a limit.

**Proposition 3.47** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Let $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) be the sequence of strongly admissible interpretations constructed based on $v$ in $D$. Then there is an $m$ with $m \geq 0$ such that $\Gamma^{m}_{D,v}(v_{\mathbf{u}}) = \Gamma^{m+1}_{D,v}(v_{\mathbf{u}})$.*

**Proof** Let $v_{\mathbf{u}}, \Gamma^1_{D,v}(v_{\mathbf{u}}), \ldots$ be the sequence of strongly admissible interpretations constructed based on $v$ in $D$. Since $\Gamma^i_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^{i+1}_{D,v}(v_{\mathbf{u}})$ for $i \geq 0$, by the first item of Lemma 3.46, and the number of arguments of $D$ is finite, the sequence $v_{\mathbf{u}}, \Gamma^1_{D,v}(v_{\mathbf{u}}), \ldots$ has to stop. That is, there exists $m \geq 0$ such that $\Gamma^m_{D,v}(v_{\mathbf{u}}) = \Gamma^{m+1}_{D,v}(v_{\mathbf{u}})$. □

**Definition 3.48** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Let $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) be the sequence of strongly admissible interpretations constructed based on $v$ in $D$. Consider an $m$ with $m \geq 0$ such that $\Gamma^m_{D,v}(v_{\mathbf{u}}) = \Gamma^{m+1}_{D,v}(v_{\mathbf{u}})$. Then, $w = \Gamma^m_{D,v}(v_{\mathbf{u}})$ is called the limit of the sequence of $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) which is the least fixed-point of $\Gamma_{D,v}$.*

Theorem 3.49 proposes the necessary and sufficient condition for an interpretation being a strongly admissible interpretation.

**Theorem 3.49** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Let $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) be the sequence of strongly admissible interpretations constructed based on $v$ in $D$. Interpretation $v$ is a strongly admissible interpretation if and only if $v$ is the limit of the sequence of $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$), (i.e., there exists an $m$ such that $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$).*

**Proof** '$\rightarrow$' Assume that $v$ is a strongly admissible interpretation of $D$. By Proposition 3.47, there exists an $m$ ($m \geq 0$) such that $\Gamma^m_{D,v}(v_{\mathbf{u}}) = \Gamma^{m+1}_{D,v}(v_{\mathbf{u}})$. We show that $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$.

- By the definition of the constructed sequence of interpretations based on $v$ in Definition 3.45 it is clear that $\Gamma^i_{D,v}(v_{\mathbf{u}}) \leq_i v$ for $i \geq 0$. Therefore, $\Gamma^m_{D,v}(v_{\mathbf{u}}) \leq_i v$.

- It remains to show that $v \leq_i \Gamma^m_{D,v}(v_{\mathbf{u}})$. Toward a contradiction assume that $v \not\leq_i \Gamma^m_{D,v}(v_{\mathbf{u}})$. This means that there exists $a$ such that either $v(a) = \mathbf{t}$, but $\Gamma^m_{D,v}(v_{\mathbf{u}})(a) \neq \mathbf{t}$ or $v(a) = \mathbf{f}$, but $\Gamma^m_{D,v}(v_{\mathbf{u}})(a) \neq \mathbf{f}$. Since $v$ is a strongly admissible interpretation, $a$ is a strongly justified argument in $v$. Thus, by Definition 3.1, there exists a non-empty subset of parents of $a$, namely $P$ such that the truth value of each $p \in P$ is presented in $v$, such that $P$ satisfies the condition of Definition 3.1 for $a$. This means that each $p \in P$ is also a strongly justified argument in $v$. Note that if $P = \emptyset$ the fact that $a$ is strongly justified in $v$ implies that $\Gamma^1_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$, i.e., $\Gamma^m_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$. Thus, $P$ has to be a non-empty set to satisfy the assumption that $\Gamma^m_{D,v}(v_{\mathbf{u}})(a) \notin \{\mathbf{t}, \mathbf{f}\}$.

If the truth value of arguments of $P$ are presented in $\Gamma^m_{D,v}(v_{\mathbf{u}})$, then there exists a $j$ with $0 \leq j \leq m$ such that the truth value of arguments of $P$ are also presented in $\Gamma^j_{D,v}(v_{\mathbf{u}})$. If so, it holds that $\Gamma^{j+1}_{D,v}(v_{\mathbf{u}})(a) \in \{\mathbf{t}, \mathbf{f}\}$. This contradicts the assumption that $\Gamma^m_{D,v}(v_{\mathbf{u}})(a) \notin \{\mathbf{t}, \mathbf{f}\}$.

Hence, there exists $p \in P$ such that the truth value of $p$ is not presented in $\Gamma^m_{D,v}(v_{\mathbf{u}})$. The fact that $p$ is a strongly justified argument in $v$ implies that there exists a non-empty subset of parents of $p$, namely $P_1$ such that the truth value of elements of $P_1$ are presented in $v$, such that $P_1$ satisfies the condition of Definition 3.1 for $p$. Using the same method of reasoning for $p$, we conclude that there exists a parent of $p$, namely $p_1$ such that the truth value of $p_1$ is not presented in $\Gamma^m_{D,v}(v_{\mathbf{u}})$.

Following the same method of reasoning, we find that there exists a sequence of ancestors of $a$, namely $p, p_1, \ldots$ such that the truth value of none of them is presented in $\Gamma^m_{D,v}(v_{\mathbf{u}})$. Since the number of arguments of $A$ is finite, the sequence $p, p_1, \ldots$ cannot be an infinite sequence. If the sequence $p, p_1, \ldots$ is finite, then for some $i$, $P_i = \emptyset$. If $P_i = \emptyset$, then by Definition 3.1, $\varphi^{v_{\mathbf{u}}}_{p_{i-1}}$ is irrefutable/unsatisfiable. This means that $\Gamma^1_{D,v}(v_{\mathbf{u}})(p_{i-1}) \in \{\mathbf{t}, \mathbf{f}\}$. This contradicts the assumption that the truth values of arguments of sequence $p, p_1, \ldots$ are not presented in $\Gamma^m_{D,v}(v_{\mathbf{u}})$. Thus, the assumption that $\Gamma^m_{D,v}(v_{\mathbf{u}})(a) \notin \{\mathbf{t}, \mathbf{f}\}$ is wrong. Hence, $v \leq_i \Gamma^m_{D,v}(v_{\mathbf{u}})$.

- '$\leftarrow$' Assume that $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$. We show that $v$ is a strongly admissible interpretation. Lemma 3.46 says that each $\Gamma^i_{D,v}(v_{\mathbf{u}})$, for $i \geq 0$, is a strongly admissible interpretation of $D$. Thus, $\Gamma^m_{D,v}(v_{\mathbf{u}})$ is a strongly admissible interpretation of $D$. As $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$ it follows that $v$ is a strongly admissible interpretation of $D$.

□

Based on the above observations, one can characterise a strongly admissible interpretation $v$ as the least fixed point of the corresponding operator $\Gamma_{D,v}$. That is, we can verify an interpretation by computing this sequence of strongly admissible interpretations. By Theorem 3.49, to investigate whether interpretation $v$ is a strongly admissible there is no need of indicating whether each argument which is presented in $v$ is a strongly justified argument in $v$. That is, there is no need of following Definition 3.1 to answer the verification problem for strong admissibility semantics

of ADFs. That is, by Theorem 3.49, it is enough to investigate whether $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$, where $\Gamma^i_{D,v}(v_{\mathbf{u}})$ $(i \geq 0)$ is a sequence of strongly admissible interpretations constructed based on $v$ in $D$. Example 3.50 illustrates the role of Theorem 3.49 and the sequence of strongly admissible interpretations constructed based on a given interpretation.

**Example 3.50** *Consider the ADF given in Example 3.2, i.e.,* $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$. *To investigate whether interpretation* $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$ *is a strongly admissible interpretation, we follow the method presented in Theorem 3.49 by constructing the sequence of strongly admissible interpretations constructed based on* $v$, *as in Definition 3.45. That is, we investigate whether there exists an* $m$ *such that* $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$. *The sequence of strongly admissible interpretations constructed based on* $v$ *is as follows.*
$v_1 = \Gamma_{D,v}(v_{\mathbf{u}}) = \Gamma_D(v_{\mathbf{u}}) \sqcap_i v = \{a, \neg d\} \sqcap_i \{a, \neg c, \neg d\} = \{a, \neg d\}$,
$v_2 = \Gamma^2_{D,v}(v_{\mathbf{u}}) = \Gamma_{D,v}(v_1) = \Gamma_D(v_1) \sqcap_i v = \{a, \neg c, \neg d\} \sqcap_i \{a, \neg c, \neg d\} = \{a, \neg c, \neg d\}$
*Since* $v$ *is the limit of the sequence* $v_1, v_2$, *i.e.,* $v = \Gamma^2_{D,v}(v_{\mathbf{u}})$, *(i.e.,* $v$ *is a least fixed point of* $\Gamma_{D,v}$*), interpretation* $v$ *is a strongly admissible interpretation of* $D$.

*On the other hand, we investigate that* $v' = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}\}$ *is not a strongly admissible interpretation of* $D$. *The sequence of interpretation constructed based on* $v'$ *are as follow.*
$v_1 = \Gamma_D(v_{\mathbf{u}}) \sqcap_i v' = \{a, \neg d\} \sqcap_i \{a, b\} = \{a\}$,
$v_2 = \Gamma_D(v_1) \sqcap_i v' = \{a, \neg d\} \sqcap_i \{a, b\} = \{a\}$.
*Thus, the sequence of interpretations constructed based on* $v'$ *leads to* $v_2 = \{a\}$, *which is not equal to* $v'$, *i.e.,* $v' \neq v_2$ *(that is,* $v'$ *is not a least fixed point of* $\Gamma_{D,v'}$*). Hence,* $v'$ *is not a strongly admissible interpretation of* $D$. *The reason is that the truth value of* $b$ *is presented in* $v'$, *however, with a similar reason as was presented in Example 3.2, it is easy to show that* $b$ *is not strongly acceptable in* $v'$.

Lemma 3.46 and Theorem 3.49 lead us to present an algorithm to answer the decision problem of whether interpretation $v$ is a strongly admissible interpretation, presented in Algorithm 1.
The results of this section lead to an alternative definition for strongly admissible semantics of ADFs, presented in Definition 3.51.

**Definition 3.51** *Let* $D$ *be an ADF and let* $v$ *be an interpretation of* $D$. *Let* $\Gamma^i_{D,v}(v_{\mathbf{u}})$ *(for* $i \geq 0$*) be the sequence of strongly admissible interpretations constructed based on* $v$ *in* $D$. *Interpretation* $v$ *is a strongly*

**Algorithm 1** Algorithm to decide whether $v$ is a strongly admissible interpretation of $D$

Input: $D$ is an ADF,
$v$ is an interpretation of $D$
Output: $v$ is (not) a strongly admissible interpretation of $D$
**for** $i \geq 0$ **do**
$w = \Gamma^{i}_{D,v}(v_{\mathbf{u}})$

**if** $\Gamma^{i+1}_{D,v}(v_{\mathbf{u}}) = v$ **then**
Print: $v$ is a strongly admissible interpretation of $D$
**else if** $\Gamma^{i+1}_{D,v}(v_{\mathbf{u}}) = w$ **then**
Print: $v$ is not a strongly admissible interpretation of $D$
break
**else**
Pass
**end if**
**end for**

*admissible interpretation if $v$ is the limit of the sequence of $\Gamma^{i}_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$), (i.e., if there exists an $m$ such that $v = \Gamma^{m}_{D,v}(v_{\mathbf{u}})$).*

In the current section we presented an alternative definition (i.e., Definition 3.51) of strongly admissible interpretations of a given ADF in which there is no need to investigate that all the arguments the truth values of which are presented in a given interpretation are strongly justifiable. If a given interpretation $v$ is a strongly admissible interpretation of $D$, then it is clear that $a$ is strongly acceptable in $v$ if $v(a) = \mathbf{t}$ and it is strongly deniable in $v$ if $v(a) = \mathbf{f}$. In contrast, when $v$ is not strongly admissible, it may contain some arguments that are strongly justifiable in $v$. For instance, in Example 3.50, $a$ is strongly acceptable in $v'$, however, $v'$ is not a strongly admissible interpretation of $D$, because $b$ is not strongly acceptable in $v'$. Algorithm 2 presents a method to answer whether an argument is strongly justifiable in a given interpretation. Note that in this method, presented in Definition 3.53, in contrast with Definition 3.1 there is no need to find a set of ancestors of a given argument to answer the decision problem. Theorem 3.52 shows why the method presented in Algorithm 2 and Definition 3.53 works to answer the decision problem whether an argument is strongly justifiable in a given interpretation.
Theorem 3.52 is a direct result of Definition 3.3 and Theorem 3.49.

**Algorithm 2** Algorithm to decide whether $a$ is a strongly justified argument in $v$

Input: $D$ is an ADF,
$v$ is an interpretation of $D$
Is $a$ strongly justified in $v$?
Output: $a$ is (not) strongly justified in $v$
$v'$ is the limit of the sequence $\Gamma^i_{D,v}(v_{\mathbf{u}})$
**if** $v(a) \in \{\mathbf{t}, \mathbf{f}\}$ and $v'(a) = v(a)$ **then**
Print: $a$ is strongly justified in $v$
**else**
Print: $a$ is not strongly justified in $v$
**end if**

**Theorem 3.52** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Let $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) be the sequence of strongly admissible interpretations constructed based on $v$ in $D$. Assume that $v'$ is the limit of the sequence of $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$). It holds that $v'(a) \in \{\mathbf{t}, \mathbf{f}\}$ if and only if $a$ is a strongly justified argument in $v$.*

Theorem 3.52 leads to an alternative definition of strong acceptability/deniability of arguments, presented in Definition 3.53, in which to answer whether a given argument is a strongly justified argument in a given interpretation there is no need to find a proper set $P$ of parents of the argument in question to satisfy the conditions of Definition 3.1.

**Definition 3.53** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. Let $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) be the sequence of strongly admissible interpretations constructed based on $v$ in $D$. Assume that $v'$ is the limit of the sequence of $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$). Argument $a$, with $v(a) \in \{\mathbf{t}, \mathbf{f}\}$, is a strongly justified argument in $v$ if $v(a) = v'(a)$.*

## 3.5 Sequence of Strongly Admissible Extensions for AFs and ADFs

In Section 3.3 we showed that the concept of strongly admissible semantics of ADFs forms a generalization of the concept of strongly admissible semantics of AFs. Furthermore, we indicated that there is no one-to-one relation between the set of strongly admissible extensions of AF $F$ and the set of strongly admissible interpretations of the associated

ADF $D_F$. In this section we clarify the relation between the sequence of strongly admissible extensions of a given AF $F$ and the sequence of strongly admissible interpretations of the associated ADF $D_F$.

In Lemma 3.46, we constructed a sequence of strongly admissible interpretations $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$) based on a given interpretation $v$ in an ADF. In Theorem 3.49, we proved that $v$ is a strongly admissible interpretation of a given ADF if and only if $v$ is the limit of the sequence $\Gamma^i_{D,v}(v_{\mathbf{u}})$ (for $i \geq 0$). There is a similar method to indicate whether a given extension is a strongly admissible extension of a given AF, presented in (Caminada and Dunne, 2019). In the following, we first recall the necessary notations from (Caminada and Dunne, 2019). Then we investigate the relation between the sequence of extensions presented in (Caminada and Dunne, 2019) for an AF and the sequence of interpretations for the associated ADF presented in the current work.

In (Caminada and Dunne, 2019, Lemma 2), it is presented that, in a given AF $F$, for an arbitrary extension $S$ of $F$, each extension in the sequence $H^0 = \emptyset$, $H^{i+1} = \Gamma_F(H^i) \cap S$ is a strongly admissible extension of $F$. We recall this lemma in Lemma 3.54. Note that in the following, $\Gamma_F(.)$ is the characteristic function of AFs, as it is defined in Definition 2.13, i.e., $\Gamma_F(S) = \{a \mid a \text{ is defended by } S\}$.

**Lemma 3.54** *(Caminada and Dunne, 2019, Lemma 2) Let $F = (A, R)$ and let $S \subseteq A$. Let $H^0 = \emptyset$ and $H^{i+1} = \Gamma_F(H^i) \cap S$ ($i \geq 0$). For each $i > 0$ it holds that*

- *$H^i \subseteq H^{i+1}$;*
- *$H^i$ is strongly admissible;*
- *$H^i$ strongly defends each of its arguments.*

Let $S$ be a strongly admissible extension of $F$ and let $v = \mathit{Ext2Int}_F(S)$. The main goal of the rest of this section is to show the exact relation between the sequence of strongly admissible extensions of $F$, namely $H^i$ as in Lemma 3.54, and the sequence of the strongly admissible interpretations of $D_F$, namely $v^i$, as in Definition 3.51. In Theorem 1 of (Caminada and Dunne, 2019), it was shown that $S$ is a strongly admissible extension of $F$ if and only if $S = \bigcup_{i=0}^{\infty} H^i$; we rephrase it in Theorem 3.55.

**Theorem 3.55** *(Caminada and Dunne, 2019, Theorem 1) Let $F = (A, R)$, let $S \subseteq A$ and let $H^i$ ($i \geq 0$) be as in Lemma 3.54. $S$ is strongly admissible iff $S = \bigcup_{i=0}^{\infty} H^i$.*

Since the number of arguments of $F$ is finite and $H^i \subseteq H^{i+1}$, we conclude that there exists $j \geq 0$ such that $S = H^j$ iff $S$ is a strongly admissible extension of $F$. It is easy to check that $Ext2Int_F$ is a monotonic function over $H^i$, that is, if $H^i \subseteq H^j$, then $Ext2Int_F(H^i) \leq_i Ext2Int_F(H^j)$. Before presenting the formal relation between the sequence of strongly admissible extensions of $F$ (in the sense of Theorem 3.55) and the sequence of strongly admissible interpretations of $D_F$ (in the sense of Theorem 3.49), we clarify this relation by an example, in Example 3.56.

**Example 3.56** *Consider $F = (\{a, b, c, d\}, \{(a, b), (b, c), (c, d)\})$ and extension $S = \{a, c\}$. We show that $S$ is a strongly admissible extension of $F$ via Theorem 3.55. That is, we construct the sequence of extensions $H^i$ and we show that $S = \bigcup_{i=0}^{\infty} H^i$. Furthermore, in the right-hand column we show the associated interpretation to each extension via Definition 3.34.*

$$H^0 = \{\} \qquad Ext2Int(H^0) = \{\}$$
$$H^1 = F(H^0) \cap S = \{a\} \cap \{a, c\} = \{a\} \qquad Ext2Int(H^1) = \{a, \neg b\}$$
$$H^2 = F(H^1) \cap S = \{a, c\} \cap \{a, c\} = \{a, c\} \quad Ext2Int(H^2) = \{a, \neg b, c, \neg d\}$$

*Since $S = H^2$, i.e., $S$ is a unique fixed point of $\bigcup_{i=0}^{\infty} H^i$, it is a strongly admissible extension of $F$. The interpretation associated with $S$ in $D_F$ via Definition 3.34 is $Ext2Int(S) = v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}, d \mapsto \mathbf{f}\}$. By Theorem 3.41, we already know that $v$ is a strongly admissible interpretation of $D_F$. In other words, we illustrate that $v$ is a strongly admissible interpretation of $D_F$ via Definition 3.51. To this end, we construct the sequence of strongly admissible interpretations $\Gamma^i_{D,v}(v_{\mathbf{u}})$ based on $v$. It is expected that there is a one-to-one relation between $Ext2Int(H^i)$ and the elements of the sequence of $\Gamma^i_{D,v}(v_{\mathbf{u}})$ in Definition 3.51.*

$$v_0 = v_{\mathbf{u}}$$
$$v_1 = \Gamma_{D_F}(v_0) \sqcap v = \{a\} \sqcap v = \{a\}$$
$$v_2 = \Gamma_{D_F}(v_1) \sqcap v = \{a, \neg b\} \sqcap v = \{a, \neg b\}$$
$$v_3 = \Gamma_{D_F}(v_2) \sqcap v = \{a, \neg b, c\} \sqcap v = \{a, \neg b, c\}$$
$$v_4 = \Gamma_{D_F}(v_3) \sqcap v = \{a, \neg b, c, \neg d\} \sqcap v = \{a, \neg b, c, \neg d\}$$

*In fact, by Theorem 3.49, it holds that $v$ is a strongly admissible interpretation of $D_F$, since $v = \Gamma^4_{D_F,v}(v_{\mathbf{u}})$. However, the guess that each $\Gamma^i_{D_F,v}(v_{\mathbf{u}})$, for $i \geq 0$, is equal to $Ext2Int(H^i)$ was wrong. Since as we can see, on*

*the one hand, Ext2Int($H^1$) contains the truth value of initial arguments that are in $S$ and the arguments that are attacked by $H^1$, i.e., the children of arguments of $H^1$. On the other hand, $\Gamma^1_{D_F,v}(v_{\mathbf{u}})$ contains the truth values of initial arguments that are presented in $v$ and do not contain the truth values of children of initial arguments presented in $v$. In other words, $\Gamma_{D_F}(v_1)$ produces the truth values of children of initial arguments presented in $v_1$. That is, $v_2 = \Gamma_{D_F}(v_1) \sqcap_i v$ contains the truth values of initial arguments and their children that are presented in $v$. Thus, it seems that, for $i \geq 0$, it holds that each Ext2Int($H^i$) is equal to $v_{2i}$. For instance, in the current example Ext2Int($H^1$) $= v_2$ and Ext2Int($H^2$) $= v_4$. Further, in this example the projection of each strongly admissible extension of $F$ via Ext2Int(.) is a strongly admissible interpretation of $D_F$. However, for instance, $v_1$ is not a projection of any strongly admissible extension. This shows that Ext2Int(.) is not a surjective function, as presented earlier in Example 3.43.*

**Theorem 3.57** *Let $F$ be an AF and let $D_F$ be its associated ADF. Let $S$ be a strongly admissible extension of $F$ and let $H^i$ be as in Lemma 3.54. Let Ext2Int($S$) $= v$ and let $v_i = \Gamma^i_{D,v}(v_{\mathbf{u}})$ be the sequence of interpretations as in Lemma 3.46. Then it holds that Ext2Int($H^i$) $= v_{2i}$, for each $i \geq 0$.*

**Proof** For $i \geq 0$, we show that $Ext2Int(H^i) = v_{2i}$ by induction on $i$.

Base case: It is obvious that $Ext2Int(H^0) = v_{\mathbf{u}} = v_0$.

Induction hypothesis: assume that $Ext2Int_F(H^j) = v_{2j}$ for each $j$ with $0 \leq j < i$.

Inductive step: we have to show that $Ext2Int_F(H^i) = v_{2i}$. To show the inductive step, we show that $a \mapsto x \in Ext2Int_F(H^i)$ if and only if $a \mapsto x \in v_{2i}$, for $x \in \{\mathbf{t}, \mathbf{f}\}$.

1. First we show that $a \mapsto \mathbf{t} \in Ext2Int_F(H^i)$ if and only if $a \mapsto \mathbf{t} \in v_{2i}$. We claim that $a \mapsto \mathbf{t} \in Ext2Int_F(H^i)$ if and only if $a \mapsto \mathbf{t} \in v_{2i-1}$. Assume that $a \mapsto \mathbf{t} \in Ext2Int_F(H^i)$, i.e., $a \in H^i$. Since $H^i = F(H^{i-1}) \cap S$, it holds that $a \in F(H^{i-1})$ and $a \in S$. Relation $a \in F(H^{i-1})$ implies that $a$ is defended by $H^{i-1}$. Thus, for each $p$ such that $(p, a) \in R$, there exists a defender of $a$, namely $c_p$, such that $(c_p, p) \in R$, $c_p \in H^{i-1}$ and since $H^{i-1}$ is a strongly admissible extension, $c_p \neq a$. Since defender $c_p$ belongs to $H^{i-1}$, it holds that $c_p \mapsto \mathbf{t} \in Ext2Int_F(H^{i-1})$ and each parent of $a$ is assigned to $\mathbf{f}$ in $Ext2Int_F(H^{i-1})$. By the induction hypothesis, $Ext2Int_F(H^{i-1}) = v_{2(i-1)} = v_{2i-2}$. That is, for each $p$ such that $(p, a) \in R$, it holds that

$p \mapsto \mathbf{f} \in v_{2i-2}$. Thus, $\varphi_a^{v_{2i-2}} \equiv \top$, that is, $a \mapsto \mathbf{t} \in \Gamma_{D_F}(v_{2i-2})$. Since $a \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(S)$, it holds that $a \mapsto \mathbf{t} \in \Gamma_{D_F}(v_{2i-2}) \sqcap_i v$. Thus, $a \mapsto \mathbf{t} \in v_{2i-1}$.

We have already shown that if $a \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(H^i)$, then $a \mapsto \mathbf{t} \in v_{2i-1}$. Since all the relations are equivalence relations, the other direction works as well. That is, if $a \mapsto \mathbf{t} \in v_{2i-1}$ then $a \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(H^i)$. Thus, $a \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(H^i)$ iff $a \mapsto \mathbf{t} \in v_{2i-1}$. We use this equation in the proof of the next item. Further, since the characteristic function is a monotonic function, $a \mapsto \mathbf{t} \in v_{2i-1}$ implies that $a \mapsto \mathbf{t} \in \Gamma_{D_F}(v_{2i-1})$. Hence, $a \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(H^i)$ if and only if $a \mapsto \mathbf{t} \in (\Gamma_{D_F}(v_{2i-1}) \sqcap_i v) = v_{2i}$.

2. We show that $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(H^i)$ if and only if $a \mapsto \mathbf{f} \in v_{2i}$. Assume that $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(H^i)$. By the definition of the $\mathit{Ext2Int}_F$ function, there exists a parent of $a$, namely $p$, such that $p \in H^i$, i.e., $p \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(H^i)$. As shown in the first item, $p \mapsto \mathbf{t} \in \mathit{Ext2Int}_F(H^i)$ if and only if $p \mapsto \mathbf{t} \in v_{2i-1}$. Thus, $\varphi_a^{v_{2i-1}} \equiv \bot$. Hence, $a \mapsto \mathbf{f} \in \Gamma_{D_F}(v_{2i-1})$. On the other hand, $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(H^i)$ implies that $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(S)$, since $\mathit{Ext2Int}_F(H^i) = \mathit{Ext2Int}_F(F(H^{i-1})) \sqcap \mathit{Ext2Int}_F(S)$. That is, $a \mapsto \mathbf{f} \in (\Gamma_{D_F}(v_{2i-1}) \sqcap_i v) = v_{2i}$. Thus, if $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(H^i)$, then $a \mapsto \mathbf{f} \in v_{2i}$. Since all the relations are equivalence relations, $a \mapsto \mathbf{f} \in \mathit{Ext2Int}_F(H^i)$ if and only if $a \mapsto \mathbf{f} \in v_{2i}$.

To conclude, $\mathit{Ext2Int}_F(H^i) = v_{2i}$, for $i \geq 0$. □

In Section 3.3, we showed that an extension is a strongly admissible extension of AF $F$ if and only if $\mathit{Ext2Int}_F(S)$ is a strongly admissible interpretation of the associated ADF $D_F$. Thus, the map of each $H^i$ via $\mathit{Ext2Int}_F(.)$, such that $H^i$ is as in Lemma 3.54, is a strongly admissible interpretation of $D_F$. However, by Theorem 3.57, $\Gamma^j_{D,v}(v_{\mathbf{u}})$ as in Lemma 3.46 is a map of an $H^i$ if and only if $j$ is an even number, i.e., $\mathit{Ext2Int}_F(H^i) = \Gamma^{2i}_{D,v}(v_{\mathbf{u}})$, for $i$ with $i \geq 0$.

## 3.6 Conclusion

In this chapter we have defined the concept of strong admissibility semantics for ADFs, based on the concept of strongly justified arguments. From a theoretical perspective, in Section 3.2, we observe that the strongly admissible interpretations of a given ADF form a lattice with the trivial

interpretation as the unique minimal element and the grounded interpretation as the unique maximal element. Furthermore, in Section 3.3 we prove that the concept of strong admissibility semantics of ADFs forms a proper generalization of the concept of strong admissibility semantics of AFs (Baroni and Giacomin, 2007; Caminada, 2014).

In addition, in Section 3.4 we have presented an alternative definition for an interpretation being strongly admissible without checking whether all the arguments that are presented in that interpretation are strongly justified. This leads to an algorithm for answering the verification problem under strong admissibility semantics of ADFs (Algorithm 1). Moreover, based on the new definition of strongly admissible interpretations of ADFs, we have presented an alternative definition for an argument being strongly justified in a given interpretation, in which there is no need to find a set of arguments that satisfies the conditions of Definition 3.1, for a given argument. This definition leads to Algorithm 2, which answers the decision problem whether a given argument is a strongly justified argument in a given interpretation.

In Section 3.5, we have indicated further relations between the sequence of strongly admissible extensions of an AF $F$ and the sequence of strongly admissible interpretations of the associated ADF $D_F$. That is, we have shown that there is no one-to-one relation between the sequence of strongly admissible extensions constructed based on a strongly admissible extension $S$ of a given AF $F$ and the sequence of strongly admissible interpretations constructed based on $\mathit{Ext2Int}(S)$ of the associated ADF $D_F$.

It is a possible topic of future research to show that the notion of strong admissibility semantics of ADFs can be reused for those generalizations of AFs that can be represented in ADFs, namely SETAFs (Nielsen and Parsons, 2006) and bipolar AFs (Cayrol and Lagasquie-Schiex, 2005).

In addition, we have proven that each grounded interpretation is the unique maximal element of the lattice of strongly admissible interpretations. Thus, it seems that the concept of strong admissibility can play a significant role in the dialectical proof procedures that we have introduced for grounded semantics in (Keshavarzi Zafarghandi et al., 2020), (see Chapter 6).

The idea of the grounded discussion games presented in (Keshavarzi Zafarghandi et al., 2020, Chapter 6) is that a discussion game can serve as an explanation why a particular argument should be accepted/denied in the grounded interpretation of a given ADF. Since the grounded semantics is the unique maximal element of the lattice constructed by strongly admissible interpretations, the concept of strong admissibility is related to grounded

semantics in a similar way as the concept of admissibility is related to preferred semantics. That is, to answer the credulous decision problem of an ADF under the grounded semantics, there is no need to construct the full grounded interpretation of the given ADF. Instead, it is enough to construct a strongly admissible interpretation of the given ADF that satisfies the decision problem. In other words, answering the credulous decision problem of ADFs under the grounded semantics is equivalent to answering the same decision problem under the strong admissibility semantics. Similarly, to answer the credulous decision problem of ADFs under preferred semantics, it is enough to investigate whether there exists an admissible interpretation in order to solve the decision problem. We used this method in preferred discussion games in (Keshavarzi Zafarghandi et al., 2019a) to answer the credulous decision problem of ADFs under preferred semantics.

On the other hand, the grounded discussion game (GDG) presented in (Keshavarzi Zafarghandi et al., 2020) was defined over ADFs without any redundant links, to answer whether a given argument is credulously justifiable under the grounded semantics of an ADF. However, the concept of strongly admissible semantics is presented for all kinds of ADFs. Thus, we will investigate whether the concept of strongly admissible semantics is at the basis of the proof procedures of the grounded discussion games for ADFs without any redundant links.

Intuitively, it seems that the grounded discussion game and a strongly admissible interpretation are two sides of the same coin to investigate whether a queried argument is credulously justified in the grounded interpretation of an ADF. Moreover, both methods can be used to explain 'why is an argument credulously justified in the grounded interpretation?' Thus, a motivation for future work is to study the relation between these two approaches.

In addition, another possible question that can be answered using the concept of strongly admissible semantics of ADFs is whether a grounded discussion game contains the least possible amount of information about the truth values of ancestors of the argument in question to answer the credulous decision problem under grounded semantics; in other words, whether the grounded discussion game presents the shortest explanation that answers the credulous decision problem under strongly admissible/grounded semantics for a given argument in an ADF. This question is interesting, since specifically when the proponent wins the game, we are eager to know whether the proponent presents the least possible amount of infor-

mation to convince the opponent about the truth of their initial claim, i.e., acceptance/denial of an argument in the grounded interpretation.

Computational complexity results of different semantics of AFs and ADFs are presented in (Dvořák and Dunne, 2018; Nouioua, 2013; Linsbichler et al., 2018; Strass and Wallner, 2015). Computational complexity of strongly admissible semantics of AFs is studied in (Dvořák and Wallner, 2020). Further, in (Caminada and Dunne, 2020), the computational complexity of identifying strongly admissible labellings with bounded or minimal size is studied. As a future work, it would be interesting to clarify the computational complexity of investigating: 1. whether a given interpretation is a strongly admissible interpretation, 2. whether a given argument is credulously/skeptically justifiable under the strongly semantics of a given ADFs.

# Chapter 4

# Complexity of Strong Admissibility

The notion of strong admissibility semantics has been introduced for ADFs in Chapter 3. In the current chapter, we study the computational complexity of several reasoning tasks under strong admissibility semantics. We address the following problems:

1. the credulous decision problem;
2. the skeptical decision problem;
3. the verification problem;
4. the strong justification problem; and
5. the problem of finding a smallest witness of strong justification of a queried argument.

## 4.1 Introduction

Interest and attention in argumentation theory has been increasing among artificial intelligence researchers (Bench-Capon and Dunne, 2007). Applications of argumentation theory are based on a variety of argumentation formalisms and methods of evaluating arguments (Atkinson et al., 2017; Baroni et al., 2018b; van Eemeren et al., 2014). Dung's abstract argumentation frameworks (Dung, 1995) (AFs for short) have received notable attention, also thanks to their simple syntax that can model and evaluate a number of non-monotonic reasoning tasks. Semantics of AFs single

out coherent subsets of arguments that fit together, according to specific criteria (Baroni et al., 2011).

AFs model individual attack relations among arguments. Abstract dialectical frameworks (ADFs) are expressive generalizations of AFs in which the logical relations among arguments can be represented. ADFs were first introduced in (Brewka and Woltran, 2010), and were further refined in (Brewka et al., 2013, 2017a, 2018a).

Often a new semantics is a refinement of an already existing one by introducing further restrictions on the set of accepted arguments or possible attackers. One of the main types of semantics of AFs is the grounded semantics. Its characteristics include that 1. each AF has a unique grounded extension; 2. the grounded extension collects all the arguments about which no one doubts their acceptance; 3. the grounded extension is often a subset of the set of extensions of other types of AF semantics. Thus, it is important to investigate whether an argument belongs to the grounded extension of a given AF. The notion of strong admissibility is introduced for AFs to answer the query 'Why does an argument belong to the grounded extension?'.

While the grounded extension collects all the arguments of a given AF that can be accepted without any doubt, a strongly admissible extension explains why an argument belongs to the grounded extension, without presenting further information about each argument in the grounded extension. Thus, the strong admissibility semantics can be the basis for an algorithm that can be used not only for answering the credulous decision problem but also for human-machine interaction that requires an explainable outcome (cf. (Caminada and Uebis, 2020; Booth et al., 2018)).

In AFs, the concept of strong admissibility semantics has first been defined in the work of Baroni and Giacomin (2007), and later in (Caminada, 2014). Furthermore, in (Caminada and Dunne, 2019), Caminada and Dunne presented a labelling account of strong admissibility to answer the decision problems of AFs under grounded semantics. Moreover, Caminada showed in (Caminada, 2018, 2014) that strong admissibility plays a role in discussion games for AFs under grounded semantics. In addition, the computational complexity of strong admissibility of AFs has been analyzed (Caminada and Dunne, 2020; Dvořák and Wallner, 2020).

Because of the specific structure of ADFs, the definition of strong admissibility semantics of AFs cannot be directly reused in ADFs. Thus the concept of strong admissibility for ADFs has been introduced (Keshavarzi Zafarghandi et al., 2021b). This concept fulfils properties that are related

to those of the strong admissibility semantics for AFs, as follows:

1. Strong admissibility is defined in terms of strongly justified arguments.
2. Strongly justified arguments are recursively reconstructed from their strongly justified parents.
3. Each ADF has at least one strongly admissible interpretation.
4. The set of strongly admissible interpretations of ADFs forms a lattice with as least element the trivial interpretation and as maximum element the grounded interpretation.
5. The strong admissibility semantics can be used to answer whether an argument is justifiable under grounded semantics.
6. The strong admissibility semantics of ADFs is different from the admissible, conflict-free, complete and grounded semantics of ADFs.
7. The strong admissibility semantics for ADFs is a proper generalization of the strong admissibility semantics for AFs.

Whereas several fundamental properties of strong admissibility semantics for ADFs have been established, the computational complexity under strong admissibility semantics has not been studied. This chapter fills this gap by studying the complexity of the central reasoning tasks under the strong admissibility semantics of ADFs, as follows.

1. The credulous decision problem, i.e., whether there exists a strongly admissible interpretation that satisfies the queried argument, is co-NP-complete.
2. The skeptical decision problem, i.e., whether all strongly admissible interpretations satisfy a queried argument, is trivial.
3. The verification problem, i.e., whether a given interpretation is a strongly admissible interpretation of an ADF, is co-NP-complete.
4. The strong justification problem for an argument in an interpretation, i.e., whether an argument is strongly justified in an interpretation, is co-NP-complete.
5. The problem of finding a smallest witness of strong justification of an argument, i.e, whether there exists a minimal strongly admissible interpretation that satisfies a queried argument, is $\Sigma_2^{\mathsf{P}}$-complete.

## 4.2 Algorithm for Strongly Admissible Interpretations of ADFs

For the ease of the readers we repeat the algorithms presented in Section 3.4. To this end, we also rephrase the concept of strong admissibility semantics for ADFs from Chapter 3, which is defined based on the notion of strongly justifiable arguments (i.e., strongly acceptable/deniable arguments). Below $v_{|_P}$ is equal to $v(p)$ for any $p \in P$, and assigns all arguments that do not belong to $P$ to $\mathbf{u}$, i.e., $v_{|_P} = v_{\mathbf{u}}|^{p \in P}_{v(p)}$.

**Definition 4.1** *Let $D = (A, L, C)$ be an ADF and let $v$ be an interpretation of $D$. Argument $a$ is a strongly justified argument in interpretation $v$ w.r.t. set $E$ if one of the following conditions hold:*

- *$v(a) = \mathbf{t}$ and there exists a subset of parents of $a$ excluding $E$, namely $P \subseteq par(a) \setminus E$ such that (a) $a$ is acceptable with respect to $v_{|_P}$ and (b) all $p \in P$ are strongly justified in $v$ w.r.t. set $E \cup \{p\}$.*

- *$v(a) = \mathbf{f}$ and there exists a subset of parents of $a$ excluding $E$, namely $P \subseteq par(a) \setminus E$ such that (a) $a$ is deniable with respect to $v_{|_P}$ and (b) all $p \in P$ are strongly justified in $v$ w.r.t. set $E \cup \{p\}$.*

*An argument $a$ is* strongly acceptable*, resp.* strongly deniable*, in $v$ if $v(a) = \mathbf{t}$, resp. $v(a) = \mathbf{f}$, and $a$ is strongly justified in $v$ w.r.t. set $\{a\}$. We further say that an argument is strongly justified in $v$ if it is either strongly acceptable or deniable in $v$.*

Note that in Definition 4.1, the set of parents of $a$ can be the empty set, i.e., $P = \emptyset$. If the set of parents of an argument,is empty, then $v|_P = v_{\mathbf{u}}$. In this case, $a$ is strongly acceptable/deniable in $v$ if $\varphi_a^{v_{\mathbf{u}}}$ is irrefutable/unsatisfiable, respectively. We say that *$a$ is not strongly justified in an interpretation $v$* if there is no such a set of parents of $a$ that satisfies the conditions of Definition 4.1 for $a$. Akin to AFs, the notion of strong admissibility semantics for ADFs is presented in Definition 4.2 based on strongly justified arguments.

**Definition 4.2** *Let $D = (A, L, C)$ be an ADF and let $v$ be an interpretation of $D$. An interpretation $v$ is a strongly admissible interpretation if for each $a$ such that $v(a) = \mathbf{t}/\mathbf{f}$, it holds that $a$ is a strongly justified argument in $v$.*

Example 4.3 clarifies the notion of strong admissibility semantics of ADFs, which is a repeat of Example 3.4 presented in Chapter 3.

**Example 4.3** *Considering ADF* $D = (\{a, b, c, d\}, \{\varphi_a : \top, \varphi_b : a \wedge \neg c, \varphi_c : \neg b \wedge d, \varphi_d : \bot\})$. *Let* $v = \{b, \neg c, \neg d\}$. *Since* $\varphi_d : \bot$, *it holds that* $d$ *is strongly deniable in* $v$. *Let* $P = \{d\}$, *since* $\varphi_c^{v|_P} \equiv \bot$, *it holds that* $c$ *is strongly deniable in* $v$. *Thus,* $c$ *and* $d$ *are strongly justified in* $v$. *However,* $b$ *is not strongly justified in* $v$, *since* $\varphi_b^v \not\equiv \top$. *Thus,* $v$ *is not a strongly admissible interpretation of* $D$. *However, for instance,* $v_1 = \{a\}$, $v_2 = \{\neg c, \neg d\}$ *and* $v_3 = \{a, b, \neg c, \neg d\}$ *are strongly admissible interpretations of* $D$. *We show that* $b$ *is strongly acceptable in* $v_3$. *To this end, let* $P = \{a, c\}$ *be a set of parents of* $b$. *First, it holds that* $\varphi_b^{v_3|_P} \equiv \top$. *Thus, the first condition of Definition 4.1 is satisfied for* $b$. *We also have to check whether each parent of* $b$ *is strongly justified in* $v_3$. *To this end, we show that* $a$ *is strongly acceptable in* $v_3$ *and* $c$ *is strongly deniable in* $v_3$. *The latter one is obvious as we show it in the beginning of this example. In addition,* $\varphi_a^{v_\mathbf{u}} \equiv \top$, *thus,* $a$ *is strongly acceptable in* $v_3$. *Hence,* $b$ *and* $a$ *are strongly justified in* $v_3$. *Further,* $v_3$ *is a unique grounded interpretation of* $D$.

In Example 4.3, if we choose a set of parents of $c$ equal to $\{b\}$, then we cannot show that $c$ is strongly deniable in interpretation $v$. The reason is that $b$ is not strongly justified in $v$, as it is presented in Example 4.3. This shows the importance of choosing a right set of parents that satisfies the conditions of Definition 4.1 for a queried argument. However, there exists an alternative definition for strongly justified of arguments, we recall it in Algorithm 4, in which there is no need of indicating a set of parents of a queried argument. First, we rephrase an alternative method, presented in Section 3.5, to answer the verification problem under strong admissibility semantics of $D$.

**Definition 4.4** *Let* $D = (A, L, C)$ *be a given ADF and let* $v, w$ *be interpretations of* $D$. *Let* $\Gamma_{D,v}(w) = \Gamma_D(w) \sqcap_i v$ *where* $\Gamma^n_{D,v}(w) = \Gamma_{D,v}(\Gamma^{n-1}_{D,v}(w))$ *for* $n$ *with* $n \geq 1$. *Note that* $\Gamma^0_{D,v}(w) = w$. *We call the collection of the interpretations of* $\Gamma^n_{D,v}(v_\mathbf{u})$ *for* $n \geq 1$, *the set of interpretations constructed based on* $v$ *in* $D$.

We rephrase the properties of the sequence of interpretation constructed based on an interpretation $v$ in Lemma 4.5.

**Lemma 4.5** *Let* $D = (A, L, C)$ *be a given ADF and let* $v$ *be an interpretation of* $D$. *Let* $\Gamma^n_{D,v}(v_\mathbf{u})$ *be the sequence of interpretation constructed based on* $v$, *as in Definition 4.4. For each* $i$ *it holds that;*

- $\Gamma^i_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^{i+1}_{D,v}(v_{\mathbf{u}})$,
- $\Gamma^i_{D,v}(v_{\mathbf{u}})$ *is a strongly admissible interpretation of* $D$,
- *if* $\Gamma^i_{D,v}(v_{\mathbf{u}})(a) = \mathbf{t}/\mathbf{f}$, *then* $a$ *is strongly justifiable in* $\Gamma^i_{D,v}(v_{\mathbf{u}})$.

**Remark 4.5.1** *the sequence of interpretations* $\Gamma^i_{D,v}(v_{\mathbf{u}})$ *as Definition 4.4, are named the sequence of strongly admissible interpretations constructed based on* $v$ *in* $D$.

Theorem 4.6 presents a method to investigate whether an interpretation $v$ is a strongly admissible interpretation based on the sequence of interpretations constructed based on $v$.

**Theorem 4.6** *Let* $D$ *be an ADF and let* $v$ *be an interpretation of* $D$. *Let* $\Gamma^i_{D,v}(v_{\mathbf{u}})$ *(for* $i \geq 0$*) be the sequence of strongly admissible interpretations constructed based on* $v$ *in* $D$. *The following conditions hold:*

- *there is an* $m$ *with* $m \geq 0$ *s.t.* $\Gamma^m_{D,v}(v_{\mathbf{u}}) = \Gamma^{m+1}_{D,v}(v_{\mathbf{u}})$;
- $v$ *is a strongly admissible interpretation of* $D$ *if and only if there exists an* $m$ *s.t.* $v = \Gamma^m_{D,v}(v_{\mathbf{u}})$.

Theorem 4.6 presents a powerful method to answer the verification problem under the strong admissibility semantics by which there is no need of examining whether all the arguments which are presented in $v$ are strongly justifiable in $v$. Lemma 4.5 and Theorem 4.6 lead us to present an algorithm to answer verification problem under the strong admissibility semantics, presented in Algorithm 3.

If a given interpretation $v$ is a strongly admissible interpretation of $D$, then it is clear that $a$ is strongly justifiable in $v$ if $v(a) = \mathbf{t}/\mathbf{f}$. In contrast, when $v$ is not strongly admissible, it may contain some arguments that are strongly justifiable in $v$. For instance, in Example 4.3, $c$ and $d$ are strongly deniable in $v$, however, $v$ is not a strongly admissible interpretation of $D$, because $b$ is not strongly acceptable in $v$. Algorithm 4 presents a method to answer whether an argument is strongly justified in a given interpretation. Note that in this method, presented in Algorithm 4, in contrast with Definition 4.1 there is no need of finding a proper set of parents of a queried argument to answer the decision problem.

**Theorem 4.7** *Let* $D$ *be an ADF and let* $v$ *be an interpretation of* $D$. *Let* $\Gamma^i_{D,v}(v_{\mathbf{u}})$ *(for* $i \geq 0$*) be the sequence of strongly admissible interpretations constructed based on* $v$ *in* $D$. *Let* $v'$ *be a limit of this sequence. It holds that* $v'(a) = \mathbf{t}/\mathbf{f}$ *if and only if* $a$ *is strongly justifiable in* $v$.

**Algorithm 3** Algorithm to decide whether $v$ is a strongly admissible interpretation of $D$

Input: $D$ is an ADF
$v$ is an interpretation of $D$
Output: $v$ is (not) a strongly admissible interpretation of $D$
**for** $i \geq 0$ **do**
$w = \Gamma^{i}_{D,v}(v_{\mathbf{u}})$

**if** $\Gamma^{i+1}_{D,v}(v_{\mathbf{u}}) = v$ **then**
Print: $v$ is a strongly admissible interpretation of $D$
**else if** $\Gamma^{i+1}_{D,v}(v_{\mathbf{u}}) = w$ **then**
Print: $v$ is not a strongly admissible interpretation of $D$
break
**else**
Pass
**end if**
**end for**

## 4.3 Computational Complexity

We analyse the complexity under strong admissibility semantics for (a) the standard reasoning tasks of ADFs (Dvořák and Dunne, 2018) and (b) two problems specific to strong admissibility semantics, i.e., the small witness problem introduced for AFs in (Dvořák and Wallner, 2020; Caminada and Dunne, 2020) and the strong justification problem.

For a given ADF $D$ we consider the following problems:

1. *The credulous decision problem*: whether an argument $a$ is credulously justifiable with respect to the strong admissibility semantics of $D$. That is, if there exists a strongly admissible interpretation of $D$ in which $a$ is strongly justified. This reasoning task is denoted as $Cred_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D)$ and is presented formally as follows:

$$Cred_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D) = \begin{cases} \text{yes} & \text{if } \exists v \in sadm(D) \text{ s.t.} \\ & v(a) = \mathbf{t}/\mathbf{f}, \\ \text{no} & \text{otherwise} \end{cases}$$

2. *The skeptical decision problem*: whether an argument $a$ is skeptically justified with respect to the strong admissibility semantics of $D$. That

**Algorithm 4** Algorithm to decide whether $a$ is a strongly justified in $v$

Input: $D$ is an ADF
$v$ is an interpretation of $D$
Is $a$ strongly justified in $v$?
Output: $a$ is (not) strongly justified in $v$
$v'$ is a limit of the sequence $\Gamma^i_{D,v}(v_{\mathbf{u}})$
**if** $v(a) \in \{\mathbf{t}, \mathbf{f}\}$ and $v'(a) = v(a)$ **then**
Print: $a$ is strongly justified in $v$
**else**
Print: $a$ is not strongly justified in $v$
**end if**

is, if $a$ is strongly justified in all strongly admissible interpretations of $D$, denoted as $Skept_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D)$, which is presented formally as follows:

$$Skept_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D) = \begin{cases} \text{yes} & \text{if } \forall v \in sadm(D): \\ & v(a) = \mathbf{t}/\mathbf{f} \text{ holds,} \\ \text{no} & \text{otherwise} \end{cases}$$

3. *The verification problem*: whether a given interpretation $v$ is a strongly admissible interpretation of $D$, denoted by $Ver_{sadm}(v, D)$, which is presented formally as follows:

$$Ver_{sadm}(v, D) = \begin{cases} \text{yes} & \text{if } v \in sadm(D), \\ \text{no} & \text{otherwise} \end{cases}$$

4. *The strong justification problem:* The problem whether a given argument $a$ is strongly justified in a given interpretation $v$ is denoted as $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D)$, which is presented formally as follows:

$$StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D) = \begin{cases} \text{yes} & \text{if } a \text{ is strongly} \\ & \text{justified in } v, \\ \text{no} & \text{otherwise} \end{cases}$$

5. *The small witness problem:* We are interested in computing a strongly admissible interpretation that has the least information of the ancestors of a given argument, namely $a$, where $v(a) = \mathbf{t}/\mathbf{f}$. The decision version of this problem is the $k$-Witness problem, denoted

by $k$-$Witness_{sadm}$, indicating whether a given argument is strongly justified in at least one $v$ such that $v \in sadm(D)$ and $|v^{\mathbf{t}} \cup v^{\mathbf{f}}| \leq k$. Note that $k$ is part of the input of this problem. This decision problem is presented formally as follows:

$$k\text{-}Witness_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D) = \begin{cases} \text{yes} & \text{if } \exists v \in sadm(D) \\ & \text{s.t. } v(a) = \mathbf{t}/\mathbf{f} \\ & \&|v^{\mathbf{t}} \cup v^{\mathbf{f}}| \leq k, \\ \text{no} & \text{otherwise} \end{cases}$$

### 4.3.1 The Credulous/Skeptical Decision Problems

In this section we study the credulous/skeptical problem under the strong admissibility semantics for ADFs. That is, we show the complexity of deciding whether an argument in question is credulously/skeptically justifiable in at least one/all strongly admissible interpretation(s) of a given ADF.

We show that $Cred_{sadm}$ is coNP-complete and $Skept_{sadm}$ is trivial. To this end, we use the fact that the set of strongly admissible interpretations of a given ADF $D$ forms a lattice with respect to the $\leq_i$-ordering, with the maximum element being $grd(D)$. Thus, any strongly admissible interpretation of $D$ has at most an amount of information equal to $grd(D)$. Thus, answering the credulous decision problem under the strong admissibility semantics coincides with answering the credulous decision problem under the grounded semantics.

**Theorem 4.8** *$Cred_{sadm}$ is coNP-complete.*

**Proof** We have that $Cred_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D) = Cred_{grd}(a \mapsto \mathbf{t}/\mathbf{f}, D)$ and the latter has been shown to be coNP-complete in (Wallner, 2014, Proposition 4.1.3.). □

Concerning skeptical acceptance, notice that the trivial interpretation is the least strongly admissible interpretation in each ADF. Thus, $Skept_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D)$ is trivially *no*.

**Theorem 4.9** *$Skept_{sadm}$ is a trivial problem.*

### 4.3.2 The Verification Problem

In this section, we settle the complexity of $Ver_{sadm}(v, D)$, i.e., of deciding whether a given interpretation $v$ is a strongly admissible interpretation of an ADF $D$.

Already, there exist two approaches to answer $Ver_{sadm}(v, D)$; one is presented in Algorithm 3, and the other one is to transfer a given ADF $D$ to another ADF $D'$ and using (Wallner, 2014, Theorem 4.1.4) to answer $Ver_{grd}(v, D')$. However, there exists an alternative method presented in this section, which shows that $Ver_{sadm}(v, D)$ is coNP-complete. Why do we present this alternative method? Based on the method, presented in this section, it turns out that $Ver_{sadm}(v, D)$ can be solved within coNP, however,

1. the evaluation of the operator $\Gamma_D$, in Algorithm 3, is costly, namely a $\mathsf{P}^{\mathsf{NP}}$-algorithm.
2. To investigate the complexity of the latter method, we first sketch a simple translation-based approach that reduces the verification problem of strongly admissible semantics to the verification problem of grounded semantics. In order to reduce $Ver_{sadm}(v, D)$ to $Ver_{grd}(v, D')$, we modify the acceptance conditions $\varphi_a$ of $D$ to $\varphi'_a = \neg a$ if $v(a) = \mathbf{u}$ and $\varphi'_a = \varphi_a$ otherwise. We then have that $v \in sadm(D)$ iff $v \in grd(D')$, so that we can use the DP procedure for $Ver_{grd}(v, D')$ (Wallner, 2014, Theorem 4.1.4). This gives a DP procedure.

Intuitively, since the grounded interpretation is the maximum element of the lattice of strongly admissible interpretations and the credulous decision problem under grounded semantics is coNP-complete, it seems that the verification problem under the strong admissibility semantics has to be coNP-complete. However, having the positive answer for $Cred_{grd}(a \mapsto \mathbf{t}/\mathbf{f}, D)$ for each $a$ with $v(a) = \mathbf{t}/\mathbf{f}$ does not lead to the positive answer of $Ver_{sadm}(v, D)$. This is because $v \leq_i grd(D)$ does not imply that $v$ is a strongly admissible interpretation of $D$ (see Example 4.10 below).

**Example 4.10** *Let $D = (\{a, b\}, \{\varphi_a : \top, \varphi_b : a \vee b\})$. The grounded interpretation of $D$ is $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}\}$. Furthermore, the interpretation $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}\}$ is an admissible interpretation of $D$ such that $v \leq_i grd(D)$. However, $v$ is not a strongly admissible interpretation of $D$. As we know, the answer of $Cred_{grd}(b \mapsto \mathbf{t}, D)$ is* yes*, but $b$ is not strongly acceptable in $v$. Thus, $v$ is not a strongly admissible interpretation of $D$, i.e., the answer to $Ver_{sadm}(v, D)$ is* no*.*

To show that $Ver_{sadm}$ is coNP-complete, we modify and combine both the fixed-point iteration from Algorithm 3, and the grounded algorithm from

(Wallner, 2014). To this end, we need some auxiliary results that are shown in Lemmas 4.11 and 4.13.

**Lemma 4.11** *Given an ADF $D$ with $n$ arguments, the following statements are equivalent:*

1. *$v$ is a strongly admissible interpretation of $D$;*
2. *$v = \Gamma^n_{D,v}(v_{\mathbf{u}})$;*
3. *for each $w \leq_i v$, it holds that $v = \Gamma^n_{D,v}(w)$.*

**Proof**

- $1 \leftrightarrow 2$ : by Theorem 4.6.
- $2 \mapsto 3$ : Assume that $v = \Gamma^n_{D,v}(v_{\mathbf{u}})$ and that $w \leq_i v$. We show that $v = \Gamma^n_{D,v}(w)$. Since $v_{\mathbf{u}} \leq_i w \leq_i v$, and $\Gamma_D$ is monotonic and thus also $\Gamma_{D,v}$ monotonic, we have $\Gamma^n_{D,v}(v_{\mathbf{u}}) \leq_i \Gamma^n_{D,v}(w) \leq_i \Gamma^n_{D,v}(v)$. Now using that $v = \Gamma^n_{D,v}(v_{\mathbf{u}})$, we obtain $v \leq_i \Gamma^n_{D,v}(w) \leq_i \Gamma^{2n}_{D,v}(v_{\mathbf{u}})$. Because $\Gamma_{D,v}$ is a monotonic operator, the fixed-point is reached after at most $n$ iterations and thus $\Gamma^{2n}_{D,v}(v_{\mathbf{u}}) = \Gamma^n_{D,v}(v_{\mathbf{u}}) = v$. Hence, $\Gamma^n_{D,v}(w) = v$.
- $3 \mapsto 2$ : Assume that for each $w \leq_i v$ it holds that $v = \Gamma^n_{D,v}(w)$. Thus, since $v_{\mathbf{u}} \leq_i v$, it holds that $v = \Gamma^n_{D,v}(v_{\mathbf{u}})$.

□

In the following, let $v^* = v^{\mathbf{t}} \cup v^{\mathbf{f}}$. The notions of completion of an interpretation and model are presented in Definition 4.12, used in Lemma 4.13.

**Definition 4.12** *Let $w$ be an interpretation. We define the* completion *of $w$ as the set of all two-valued extensions of $w$, denoted by $[w]_2$ where: $[w]_2 = \{u \mid w \leq_i u$ and $u$ is a two-valued interpretation$\}$.*

*Furthermore, a two-valued interpretation $u$ is said to be a* model *of formula $\varphi$, if $u(\varphi) = \mathbf{t}$, denoted by $u \models \varphi$.*

**Lemma 4.13** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. $v \notin sadm(D)$ if and only if there exists an interpretation $w$ of $D$ that satisfies all the following conditions:*

1. *$w <_i v$;*
2. *For each $a \in w^{\mathbf{u}} \cap v^{\mathbf{t}}$ there exists $u_a \in [w]_2$ s.t. $u_a \not\models \varphi_a$;*

*3. For each $a \in w^{\mathbf{u}} \cap v^{\mathbf{f}}$ there exists $u_a \in [w]_2$ s.t. $u_a \models \varphi_a$.*

**Proof** $\Leftarrow$: Assume that $v$ and $w$ are interpretations of $D$ that satisfy all of the items 1, 2, 3 presented in the lemma. We show that $v \notin sadm(D)$. Toward a contradiction assume that $v \in sadm(v)$. Let $a$ be an argument such that $a \in w^{\mathbf{u}} \cap v^{\mathbf{t}}$, thus, since $w$ satisfies the conditions of the lemma, it holds that there exists $u_a \in [w]_2$ such that $u_a \not\models \varphi_a$, i.e., $u_a(a) = \mathbf{f}$. Furthermore, since $v(a) = \mathbf{t}$ and $v \in sadm(D)$, for any $j \in [v]_2$ it holds that $j \models \varphi_a$. Since $w <_i v$, it holds that $j \in [w]_2$. We have shown that there are two two-valued extensions of $w$ that differ in their truth value on $a$. Hence, based on the definition of the characteristic operator $\Gamma_D(w)(a)$ is neither unsatisfiable nor irrefutable, i.e., $\Gamma_D(w)(a) = \mathbf{u}$. The proof method for the case that $a \in w^{\mathbf{u}} \cap v^{\mathbf{f}}$ is similar, i.e., if $a \in w^{\mathbf{u}} \cap (v^{\mathbf{t}} \cup v^{\mathbf{f}})$, then $\Gamma_D(w)(a) = \mathbf{u}$. Thus, for $a \in w^{\mathbf{u}} \cap v^*$ we have $\Gamma_{D,v}(w)(a) = (\Gamma_D(w) \sqcap v)(a) = \mathbf{u}$. In other words, $\Gamma_{D,v}(w) \leq_i w$ and thus, by the monotonicity of $\Gamma_{D,v}(w)$ also $\Gamma^n_{D,v}(w) \leq_i w <_i v$.

Thus, since $\Gamma^n_{D,v}(w) \neq v$ the third item of Lemma 4.11 does not hold for $w$ with $w <_i v$. Thus, $v \notin sadm(D)$.
$\Rightarrow$: Assume that $v \notin sadm(D)$. That is, for the fixed point $w = \Gamma^n_{D,v}(v_{\mathbf{u}})$ we have $w <_i v$. Consider $a \in w^{\mathbf{u}} \cap v^{\mathbf{t}}$. Because $w$ is a fixed point, we have that $\Gamma_{D,v}(w)(a) \neq \mathbf{t}$ and thus $\Gamma_D(w) \neq \mathbf{t}$. That is, there is a $u_a \in [w]_2$ such that $u_a \not\models \varphi_a$. Similar reasoning applies to $a \in w^{\mathbf{u}} \cap v^{\mathbf{f}}$. □

Lemma 4.14 shows that the verification problem is a coNP-problem, and Lemma 4.15 shows the hardness of this problem.

**Lemma 4.14** *$Ver_{sadm}$ is a coNP-problem for ADFs.*

**Proof** Let $D$ be an ADF and let $v$ be an interpretation of $D$. For membership, consider the co-problem. By Lemma 4.13, if there exists an interpretation of $w$ that satisfies the condition of Lemma 4.13, then $v$ is not a strongly admissible interpretation of $D$. Thus, guess an interpretation $w$, together with interpretations $u_a \in [w]_2$ for each $a \in v^*$, and check whether they satisfy the conditions of Lemma 4.13. Note that since $w <_i v$ we have to check the second and the third items of Lemma 4.13 a total of $|v^* \setminus w^{\mathbf{u}}|$ number of times. That is, this checking has to be done at most $|v^*|$ number of times, when $w$ is the trivial interpretation. Thus, this checking step is linear in the size of $v^*$. Therefore, the procedure of guessing of $w$ and checking if it satisfies 1, 2, 3 of Lemma 4.13 is an NP-problem. Thus, if a $w$ satisfies the items of Lemma 4.13, then the answer to $Ver_{sadm}(v, D)$ is *no*. Otherwise, if we check all interpretations $w$ such that $w <_i v$ and

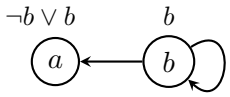


**Figure 4.1:** Reduction used in Lemma 4.15 and 4.23, for $\psi = \neg b \vee b$.

none of them satisfies the conditions of Lemma 4.13, then the answer to $Ver_{sadm}(v, D)$ is *yes*. Thus, $Ver_{sadm}(v, D)$ is a coNP-problem. □

**Lemma 4.15** *$Ver_{sadm}$ is* coNP*-hard for ADFs.*

**Proof** For hardness of $Ver_{sadm}$, we consider the standard propositional logic problem of VALIDITY. Let $\psi$ be an arbitrary Boolean formula and let $X = atom(\psi)$ be the set of atoms in $\psi$. Let $a$ be a new atom, i.e., $a \notin X$. Construct ADF $D = (\{X \cup \{a\}\}, L, C)$ where $\varphi_x : x$ for each $x \in X$ and $\varphi_a : \psi$. We show that $\psi$ is valid if and only if $v = v_{\mathbf{u}}|^a_{\mathbf{t}}$ is a strongly admissible interpretation of $D$. An illustration of the reduction for the formula $\psi = \neg b \vee b$ to the ADF $D = (\{a, b\}, L, \varphi_a : \psi, \varphi_b : b)$ is shown in Figure 4.1. Assume that $\psi$ is a valid formula. We show that $v$ is the grounded interpretation of $D$. By the acceptance condition of each $x$, for $x \in X$ it is clear that $x$ is assigned to $\mathbf{u}$ in the grounded interpretation of $D$. Further, since $\psi$ is a valid formula, it holds that ${\varphi_a}^{v_{\mathbf{u}}} \equiv \top$. Thus, the interpretation $v = v_{\mathbf{u}}|^a_{\mathbf{t}}$ is the grounded interpretation of $D$. Hence, $v \in sadm(D)$.

On the other hand, assume that $\psi$ is not valid. Then there exists a two-valued interpretation $v$ of $atom(\psi)$ such that $v \not\models \psi$. This implies that $a \mapsto \mathbf{t}$ does not belong to the grounded interpretation of $D$. Since the grounded interpretation of $D$ is the maximum element of the lattice of strongly admissible interpretations, it holds that $a$ is not strongly acceptable in any strongly admissible interpretation of $D$, that is, $v \notin sadm(D)$. □

Theorem 4.16 is a direct result of Lemmas 4.14–4.15.

**Theorem 4.16** *$Ver_{sadm}$ is* coNP*-complete for ADFs.*

### 4.3.3 Strong Justification of an Argument

Note that it is possible that an interpretation $v$ contains some strongly justified arguments but $v$ is not strongly admissible itself. Example 4.17 presents such an interpretation. Thus, the problem $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D)$

of deciding whether an argument is strongly justified in a given interpretation of an ADF is different from the previously discussed decision problems. We show that *StrJust* is coNP-complete.

**Example 4.17** *Let* $D = (\{a, b, c, d\}, \{\varphi_a : \bot, \varphi_b : \neg a \wedge c, \varphi_c : d, \varphi_d : \top\})$ *be an ADF. Let* $v = \{b, c, d\}$ *be an interpretation of* $D$*. It is easy to check that* $c$ *and* $d$ *are strongly acceptable in* $v$*. However,* $b$ *is not strongly acceptable in* $D$*. Thus,* $v$ *is not a strongly admissible interpretation of* $D$*. However, there exists a strongly admissible interpretation of* $D$ *in which* $c$ *and* $d$ *are strongly acceptable and that has less information than* $v$*, namely,* $v' = \{c, d\}$*.*

Algorithm 4 presents a straightforward method of deciding whether $a$ is strongly justified in a given interpretation $v$. That is, $a$ is strongly acceptable/deniable in $v$ if it is acceptable/deniable by the least fixed point of the operator $\Gamma_{D,v}$ (which is equal to $\Gamma^n_{D,v}(v_{\mathbf{u}})$ for sufficiently large $n$).

However, the repeated evaluation of $\Gamma_D$ is a costly part of this algorithm and results in a $\mathsf{P}^{\mathsf{NP}}$ algorithm. We will next discuss a more efficient method to answer this reasoning task. To this end, we translate a given ADF $D$ to ADF $D'$, presented in Definition 4.18, such that the queried argument is strongly justifiable in a given interpretation of $D$ if and only if it is credulously justifiable in the grounded interpretation of $D'$. As shown in (Wallner, 2014, Proposition 4.1.3), the credulous decision problem for ADFs under grounded semantics is a coNP-problem. Thus, verifying whether a given argument is strongly justified in an interpretation is a coNP-problem, since the translation can be done in polynomial time with respect to the size of $D$.

**Definition 4.18** *Let* $D = (A, L, C)$ *be an ADF and let* $v$ *be an interpretation of* $D$*. The translation of* $D$ *under* $v$ *is* $D' = (A', L', C')$ *such that* $A' = A \cup \{x, y\}$ *where* $x, y \notin A$*. Furthermore, for each* $a \in A'$ *we define the acceptance condition of* $a$ *in* $D'$*, namely* $\varphi'_a$ *as follows:*

- $\varphi'_x : x$*;*
- $\varphi'_y : y$*;*
- *if* $v(a) = \mathbf{u}$*, then* $\varphi'_a : \neg a$*;*
- *if* $v(a) = \mathbf{t}$*, then* $\varphi'_a = \varphi_a \vee x$*;*
- *if* $v(a) = \mathbf{f}$*, then* $\varphi'_a = \varphi_a \wedge y$*.*

Notice that our reduction ensures that arguments with $v(a) = \mathbf{u}$ will always be $\mathbf{u}$ in $D'$, arguments with $v(a) = \mathbf{t}$ will be assigned to either $\mathbf{t}$ or $\mathbf{u}$ during the least fixed-point computation and arguments with $v(a) = \mathbf{f}$ will be assigned to either $\mathbf{f}$ or $\mathbf{u}$. That is we introduced arguments $x$, $y$ to ensure that arguments in $v^*$ are not assigned to the opposite truth value during the iteration of $\Gamma_{D'}$ that leads to the grounded interpretation of $D'$. Lemmas 4.19 and 4.20 show the correctness of the reduction.

**Lemma 4.19** *Let $D$ be an ADF, let $v$ be an interpretation of $D$, and let $D'$ be the translation of $D$, via Definition 4.18. It holds that if $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D) = yes$, then $Cred_{grd}(a \mapsto \mathbf{t}/\mathbf{f}, D') = yes$.*

**Proof** We assume that $StrJust(a \mapsto \mathbf{t}, v, D) = yes$, and we show that $Cred_{grd}(a \mapsto \mathbf{t}, D') = yes$. The proof for the case that $StrJust(a \mapsto \mathbf{f}, v, D) = yes$ is similar.

Assume that $v_{\mathbf{u}}$ is the trivial interpretation of $D$ and $v'_{\mathbf{u}}$ is the trivial interpretation of $D'$, i.e., $v'_{\mathbf{u}} = v_{\mathbf{u}} \cup \{x \mapsto \mathbf{u}, y \mapsto \mathbf{u}\}$. Assume that $\Gamma^i_{D,v}(v_{\mathbf{u}})$ is a sequence of strongly admissible interpretations constructed based on $v$ in $D$, as in Definition 4.4. Let $w$ be the limit of the sequence of $\Gamma^i_{D,v}(v_{\mathbf{u}})$.

$StrJust(a \mapsto \mathbf{t}, v, D) = yes$ implies that $w(a) = \mathbf{t}$. Since $w$ is a strongly admissible interpretation of $D$, it holds that $a \mapsto \mathbf{t}$ in the grounded interpretation of $D$, i.e., there exists a natural number $n$ such that $\Gamma^n_D(v_{\mathbf{u}})(a) = \mathbf{t}$. By induction on $n$, it is easy to show that $\Gamma^n_{D'}(v'_{\mathbf{u}})(a) = \mathbf{t}$. That is, $a$ is assigned to $\mathbf{t}$ in the grounded interpretation of $D'$. Thus, $Cred_{grd}(a \mapsto \mathbf{t}, D') = \text{yes}$.

□

**Lemma 4.20** *Let $D$ be an ADF, let $v$ be an interpretation of $D$, and let $D'$ be the translation of $D$ via Definition 4.18. It holds that if $Cred_{grd}(a \mapsto \mathbf{t}/\mathbf{f}, D) = yes$, then $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D) = yes$.*

**Proof** Assume that $a$ is justified in the grounded interpretation of $D'$, namely $w$. Thus, there exists a $j$ such that $w = \Gamma^j_{D'}(w_{\mathbf{u}})$ for $j \geq 0$, where $w_{\mathbf{u}}$ is the trivial interpretation of $D'$. By induction we prove the claim that for all $i$, if $a \mapsto \mathbf{t}/\mathbf{f} \in \Gamma^i_{D'}(w_{\mathbf{u}})$, then $a$ is strongly justified in $v$.

Base case: Assume that $a \mapsto \mathbf{t}/\mathbf{f} \in \Gamma^1_{D'}(w_{\mathbf{u}})$. By the acceptance conditions of $x$ and $y$ in $D'$, both of them are assigned to $\mathbf{u}$ in $w$. Then it has to be the case that either $\varphi'_a = \varphi_a \vee x$ or $\varphi'_a = \varphi_a \wedge y$ in $D'$. Thus, $a \mapsto \mathbf{t}/\mathbf{f} \in \Gamma^1_{D'}(w_{\mathbf{u}})$ implies that ${\varphi'_a}^{w_{\mathbf{u}}} \equiv \top/\bot$. Thus, $w(x/y) = \mathbf{u}$, $\varphi'_a = \varphi_a \vee x/\varphi_a \wedge y$ and ${\varphi'_a}^{w_{\mathbf{u}}} \equiv \top/\bot$ together imply that ${\varphi_a}^{w_{\mathbf{u}}} \equiv \top/\bot$.

Hence, $\varphi_a^{v_\mathbf{u}} \equiv \top/\bot$ where $v_\mathbf{u}$ is the trivial interpretation of $D$. That is, $a$ is strongly justified in $v$.

Induction hypothesis: Assume that for all $j$ with $1 \leq j \leq i$, if $a \mapsto \mathbf{t}/\mathbf{f} \in \Gamma_{D'}^{j}(w_\mathbf{u})$, then $a$ is strongly justified in $v$.

Inductive step: We show that if $a \mapsto \mathbf{t}/\mathbf{f} \in \Gamma_{D'}^{i+1}(w_\mathbf{u})$, then $a$ is strongly justified in $v$. Because $x/y \mapsto \mathbf{u} \in w$, we have that $\varphi_a^w \equiv \top/\bot$ implies that $\varphi_a^v \equiv \top/\bot$. Further, $a \mapsto \mathbf{t}/\mathbf{f} \in \Gamma_{D'}^{i+1}(w_\mathbf{u})$ says that there exists a set of parents of $a$, namely $P$, where $P \subseteq w^\mathbf{t} \cup w^\mathbf{f}$, such that, $\varphi_a^{w_{|P}} \equiv \top/\bot$. Thus, $\varphi_a^{v_{|P}} \equiv \top/\bot$. By induction hypothesis, each $p \in P$ is strongly justified in $v$. Thus, $a$ is strongly justified in $v$. □

Theorem 4.21 is a direct result of Lemmas 4.19 and 4.20.

**Theorem 4.21** *Let $D$ be an ADF, let $v$ be an interpretation of $D$, and let $D'$ be the translation of $D$, via Definition 4.18. It holds that $Cred_{grd}(a \mapsto \mathbf{t}/\mathbf{f}, D) = yes$ iff $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D) = yes$.*

We use the auxiliary Theorem 4.21 to present the main result of this section, i.e., to show that *StrJust* is coNP-complete.

**Lemma 4.22** *Let $D$ be an ADF, let $a$ be an argument, and let $v$ be an interpretation of $D$. Deciding whether $a$ is strongly justified in $v$, i.e., whether $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D)$, is a coNP-problem.*

**Proof** It is shown in (Wallner, 2014, Proposition 4.1.3) that the credulous decision problem under grounded semantics, i.e., $Cred_{grd}$, is a coNP-problem. Further, the translation of a given ADF $D$ to $D'$ via Definition 4.18 can be done in polynomial time. By Theorem 4.21, it holds that $Cred_{grd}(a \mapsto \mathbf{t}/\mathbf{f}, D) = yes$ iff $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D) = yes$. Thus, deciding whether a given argument is strongly justified in interpretation $v$, i.e., $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D)$ is a coNP-problem. □

**Lemma 4.23** *Let $D$ be an ADF, let $a$ be an argument, and let $v$ be an interpretation of $D$. Deciding whether $a$ is strongly justified in $v$, i.e., $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D)$, is coNP-hard.*

**Proof** Let $\psi$ be any Boolean formula and let $X = atom(\psi)$ be the set of atoms in $\psi$. Let $a$ be a new variable that does not appear in $X$. Construct $D = (\{X \cup \{a\}\}, L, C)$, such that $\varphi_x : x$ for each $x \in X$ and $\varphi_a : \psi$. ADF $D$ can be constructed in polynomial time with respect to the size of $\psi$. We show that $a$ is strongly acceptable in any $v$ where $v(a) = \mathbf{t}$ if and only if $\psi$

is a valid formula. An illustration of the reduction for a formula $\psi = \neg b \vee b$ to the ADF $D = (\{a, b\}, L, \varphi_a : \psi, \varphi_b : b)$ is depicted in Figure 4.1.

Assume that $a$ is strongly acceptable in $v$, thus by Definition 4.1, there exists a set of parents of $a$, namely $P$, such that $\varphi_a^{v_{|P}} \equiv \top$ and for each $p \in P$ it holds that $p$ is strongly justified in $v$. By the definition of $D$ the acceptance condition of each parent of $a$, namely $p$ is $\varphi_p : p$, thus, by the acceptance condition of $p$, it is not strongly justifiable in $v$. Thus, the only case in which $a$ is strongly acceptable in $v$ is that $P = \emptyset$, i.e., $\varphi_a^{v_\mathbf{u}} \equiv \top$. Hence, for any two-valued interpretation $u$ of $X \cup \{a\}$ it holds that $u \models \psi$. Moreover since the atom $a$ does not appear in $\psi$ we obtain that for any two-valued interpretation $u$ of $X$ it holds that $u \models \psi$. Hence, $\psi$ is a valid formula and it is a *yes* instance of the VALIDITY problem of classical logic.

On the other hand, assume that $\psi$ is a valid formula. Then it is clear that the interpretation $v$ that assigns $a$ to $\mathbf{t}$ and $x$ to $\mathbf{u}$, for each $x \in X$, is the grounded interpretation of $D$. Thus, the answer to the strong acceptance problem of $a$ in any $v$ with $v(a) = \mathbf{t}$ is *yes*.

For credulous denial of $a$, it is enough to present the acceptance condition of $a$ equal to the negation of $\psi$ in $D$, i.e., $\varphi_a : \neg\psi$, and follow a similar method. That is, $a$ is strongly deniable in $v$, where $v(a) = \mathbf{f}$, if and only if $\psi$ is a valid formula. □

Theorem 4.24 is a direct result of Lemmas 4.22 and 4.23.

**Theorem 4.24** *Let $D$ be an ADF, let $a$ be an argument, and let $v$ be an interpretation of $D$. Deciding whether $a$ is strongly justified in $v$, i.e., $StrJust(a \mapsto \mathbf{t}/\mathbf{f}, v, D)$ is* coNP*-complete.*

### 4.3.4 Smallest Witness of Strong Justification

Assume that an argument $a$, its truth value, and a natural number $k$ are given. We are eager to know whether there exists a strongly admissible interpretation $v$ that satisfies the truth value of $a$ and $|v^\mathbf{t} \cup v^\mathbf{f}| < k$. This reasoning task is denoted by $k\text{-}Witness_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D)$. We show that $k\text{-}Witness_{sadm}$ is $\Sigma_2^\mathsf{P}$-complete. Lemma 4.25 shows that this problem is a $\Sigma_2^\mathsf{P}$-problem and Lemma 4.26 indicates the hardness of this reasoning task.

**Lemma 4.25** *Let $D = (A, L, C)$ be an ADF, let $a$ be an argument, let $x \in \{\mathbf{t}, \mathbf{f}\}$, and let $k$ be a natural number. Deciding whether there exists a strongly admissible interpretation $v$ of $D$ where $v(a) = x$ and $|v^\mathbf{t} \cup v^\mathbf{f}| < k$ is a $\Sigma_2^\mathsf{P}$-problem, i.e., $k\text{-}Witness_{sadm}$ is a $\Sigma_2^\mathsf{P}$-problem.*

**Proof** For membership, non-deterministically guess an interpretation $v$ and verify whether this interpretation satisfies the following items:

1. $v \in sadm(D)$;
2. $v(a) = x$;
3. $|v^{\mathbf{t}} \cup v^{\mathbf{f}}| < k$.

If $v$ satisfies all the items, then the answer to the decision problem is *yes*, i.e., $k\text{-}Witness_{sadm}(a \mapsto \mathbf{t}/\mathbf{f}, D) = yes$. The complexity of each of the above items is as follows.

1. Verifying strong admissibility of $v$ is co-NP-complete, as is presented in Section 4.3.2.
2. Verifying if $v$ contains the claim, i.e., if $v(a) = x$, can clearly be done in polynomial time.
3. Collecting $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ and checking whether $|v^{\mathbf{t}} \cup v^{\mathbf{f}}| < k$ takes only polynomial time.

That is, the algorithm first non-deterministically guesses an interpretation $v$ and then performs checks that are in coNP to verify that $v$ satisfies the requirements of the decision problem. Thus, this gives an $\mathsf{NP}^{\mathsf{coNP}} = \Sigma_2^{\mathsf{P}}$ procedure. □

**Lemma 4.26** *Let $D = (A, L, C)$ be an ADF, let $a$ be an argument, let $x \in \{\mathbf{t}, \mathbf{f}\}$, and let $k$ be a natural number. Deciding whether there exists a strongly admissible interpretation $v$ of $D$ where $v(a) = x$ and $|v^{\mathbf{t}} \cup v^{\mathbf{f}}| < k$ is $\Sigma_2^{\mathsf{P}}$-hard, i.e., $k\text{-}Witness_{sadm}$ is $\Sigma_2^{\mathsf{P}}$-hard.*

**Proof** Consider the following well-known problem on quantified Boolean formulas. Given a formula $\Theta = \exists Y \forall Z\, \theta(Y, Z)$ with atoms $X = Y \cup Z$ (and $Y \cap Z = \emptyset$) and propositional formula $\theta$. Deciding whether $\Theta$ is valid is $\Sigma_2^{\mathsf{P}}$-complete (see e.g. (Arora and Barak, 2009)). We can assume that $\theta$ is of the form $\psi \wedge \bigwedge_{y \in Y}(y \vee \neg y)$, where $\psi$ is an arbitrary propositional formula over atoms $X$, and that $\theta$ is satisfiable. Moreover, we can assume that the formula $\theta$ only uses $\wedge$, $\vee$, $\neg$ operations and negations only appear in literals. Let $\bar{Y} = \{\bar{y} : y \in Y\}$, i.e., for each $y \in Y$ we introduce a new argument $\bar{y}$.

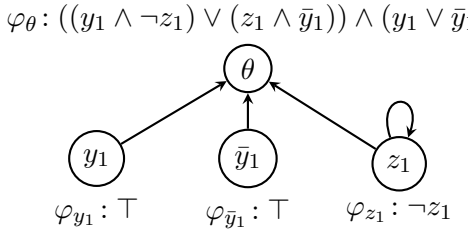


**Figure 4.2:** Illustration of the reduction from the proof of Lemma 4.26 for $\Theta = \exists y_1 \forall z_1((y_1 \wedge \neg z_1) \vee (z_1 \wedge \neg y_1)) \wedge (y_1 \vee \neg y_1)$.

We construct an ADF $D_\Theta = (A, L, C)$ with

$$\begin{aligned} A =& Y \cup \bar{Y} \cup Z \cup \{\theta\} \\ C =& \{\varphi_y : \top \mid y \in Y\} \cup \{\varphi_{\bar{y}} : \top \mid y \in Y\} \\ & \cup \{\varphi_z : \neg z \mid z \in Z\} \cup \{\varphi_\theta : \theta[\neg y/\bar{y}]\} \end{aligned}$$

It is easy to verify that the grounded interpretation $g$ of $D_\Theta$ sets all arguments $Y \cup \bar{Y}$ to $\mathbf{t}$ and all arguments $Z$ to $\mathbf{u}$. Moreover, $g(\theta) \in \{\mathbf{t}, \mathbf{u}\}$. An illustration of the reduction for a formula $\theta = ((y_1 \wedge \neg z_1) \vee (z_1 \wedge \neg y_1)) \wedge (y_1 \vee \neg y_1)$ to the ADF $D = (A, L, C)$ is shown in Figure 4.2, where: $A = \{y_1, \bar{y}_1, z_1, \theta\}$, $\varphi_{y_1} : \top, \varphi_{\bar{y}_1} : \top, \varphi_{z_1} : \neg z$ and $\varphi_\theta : ((y_1 \wedge \neg z_1) \vee (z_1 \wedge \bar{y}_1)) \wedge (y_1 \vee \bar{y}_1)$. We show that there is a strongly admissible interpretation $v$ with $v(\theta) = \mathbf{t}$ and $|S| = |Y| + 1$ where $S = v^{\mathbf{t}} \cup v^{\mathbf{f}}$ iff $\Theta$ is a valid formula.

- Assume that $\Theta$ is a valid formula. We show that there exists a strongly admissible interpretation $v$ with $|S| = |Y| + 1$. Since $\Theta$ is a valid formula, there exists an interpretation $I_Y$ of $Y$ such that for any interpretation $I_Z$ of $Z$, it holds that $I_Y \cup I_Z \models \theta(Y, Z)$, i.e., $\theta$ is true. Specifically, it holds that $I_Y \models \theta(Y, Z)$.

  We define a three-valued interpretation $v$ of $A$ such that $v(y) = \mathbf{t}$ if $I_Y(y) = \mathbf{t}$, $v(\bar{y}) = \mathbf{t}$ if $I_Y(y) = \mathbf{f}$, $v(\theta) = \mathbf{t}$, and $v(x) = \mathbf{u}$ otherwise. It is easy to check that $v$ is a strongly admissible interpretation of $D$ where $|S| = |Y| + 1$. Thus, $\theta$ is strongly acceptable in a strongly admissible interpretation $v$ where $|S| = |Y| + 1$.

- Let $v$ be the strongly admissible interpretation with $v(\theta) = \mathbf{t}$ and $|S| \leq |Y| + 1$. Let $g$ be the unique grounded interpretation of $D$. It holds that $v \leq_i q$. For each $z \in Z$, since $c_z : \neg z$, it is clear that $v(z) = \mathbf{u}$ in any strongly admissible interpretation $v$ of $D$. Moreover, because $\theta$ is of the form $\psi \wedge \bigwedge_{y \in Y}(y \vee \neg y)[\neg y/\bar{y}]$, we have that for each

| | $Cred_{sadm}$ | $Skept_{sadm}$ | $Ver_{sadm}$ | $StrJust$ | $k\text{-}Witness_{sadm}$ |
|---|---|---|---|---|---|
| AFs | P | trivial | P | n.a. | NP-c |
| ADFs | coNP-c | trivial | coNP-c | coNP-c | $\Sigma_2^P$-c |

**Table 4.1:** Complexity under the strong admissibility semantics of AFs and ADFs ($\mathscr{C}$-c denotes completeness for class $\mathscr{C}$)

$y \in Y$ either $v(y) = \mathbf{t}$ or $v(\bar{y}) = \mathbf{t}$ and thus $|S| = |Y| + 1$. Because of this, we also have that not both $v(y) = \mathbf{t}$ or $v(\bar{y}) = \mathbf{t}$ can be simultaneously true. We can thus define the following interpretation $I_Y$ of $Y$ such that $I_Y(y) = \mathbf{t}$ if $v(y) = \mathbf{t}$ and $I_Y(y) = \mathbf{f}$ if $v(\bar{y}) = \mathbf{t}$. Since $\theta$ is strongly accepted with respect to $v$, we have that for each interpretation $I_Z$ of $Z$, the formula $\theta$ is satisfied by $I_Y \cup I_Z$. That is, the QBF $\Theta$ is valid.

□

Theorem 4.27 is a direct result of Lemmas 4.25 and 4.26.

**Theorem 4.27** *$k\text{-}Witness_{sadm}$ is $\Sigma_2^P$-complete.*

In Table 4.1, we summarize our results on the complexity of strong admissibility semantics in ADFs and compare them with the corresponding results for AFs (Caminada and Dunne, 2020; Dvořák and Wallner, 2020).

## 4.4 Conclusion

We studied the computational properties of the strong admissibility semantics of ADFs. When compared to AFs, computational complexity for ADFs increases by one step in the polynomial hierarchy (Stockmeyer, 1976) for nearly all reasoning tasks (Strass and Wallner, 2015; Dvořák and Dunne, 2018). We have shown that, similarly, ADFs have higher computational complexity under the strong admissibility semantics when compared to AFs (Table 4.1).

From a theoretical perspective we observe that:

1. The credulous decision problem under the strong admissibility semantics of ADFs is coNP-complete, while this decision problem is tractable in AFs.
2. Since the trivial interpretation is the least strongly admissible interpretation for each ADF, the skeptical decision problem is trivial, which is similar for AFs.

3. The verification problem for ADFs is coNP-complete, while it is tractable for AFs.
4. Since an argument can be strongly justified in an interpretation that is not a strongly admissible interpretation, we defined a new reasoning task in Section 4.3.3, called the strong justification problem. The complexity of this decision problem, which investigates whether a queried argument is strongly justified in a given interpretation, is coNP-complete.
5. The problem of finding a smallest witness of strong justification of an argument investigates whether there exists a strongly admissible interpretation that assigns a minimum number of arguments to $\mathbf{t}/\mathbf{f}$ and satisfies a queried argument is $\Sigma_2^{\mathsf{P}}$-complete, while this reasoning task is NP-complete for AFs.

We next highlight an interesting difference in the complexity landscapes of AFs and ADFs. When relating the complexity of grounded and strong admissibility semantics, we have that for AFs the verification problems can be (log-space) reduced to each other, while for ADFs there is a gap between the coNP-complete $Ver_{sadm}$ problem and the DP-complete $Ver_{grd}$ problem. That is, on the ADF level the step of proving arguments to be $\mathbf{u}$ in the grounded interpretation adds an NP part to the complexity; a similar effect can be observed for admissible and complete semantics.

As future work, it would be interesting to analyse the computational complexity of the current reasoning tasks for strong admissibility semantics over subclasses of ADFs, in particular bipolar ADFs (Brewka and Woltran, 2010) and acyclic ADFs (Diller et al., 2020).

# Chapter 5

# Semi-Stable Semantics

In ADFs stable semantics both shows the 'black-and-white' character of a knowledge representation and indicates support-cycle among arguments. However, similar to AFs, an ADF may not have any stable model. In the case that an AF has no stable extension, the notion of semi-stable semantics of AFs has been proposed as a way of approximating stable semantics of AFs.

However, the notion of semi-stable semantics as studied for AFs has received little attention for ADFs, as a remedy of approximating stable semantics if an ADF does not have any stable model. In the current chapter, we present the concepts of semi-two-valued models and semi-stable models for ADFs. We show that these two notions satisfy a set of plausible properties required for semi-stable semantics of ADFs. Furthermore, we show that semi-two-valued and semi-stable semantics of ADFs form proper generalization of the semi-stable semantics of AFs, just like two-valued model and stable semantics for ADFs are generalizations of stable semantics for AFs.

## 5.1 Introduction

Formalisms of argumentation have been introduced to model and evaluate argumentation. Abstract argumentation frameworks (AFs) as introduced by Dung (1995) are a core formalism in formal argumentation. A popular line of research investigates extensions of Dung's AFs that allow for a richer syntax (see, e.g. Brewka et al. 2014).

In this chapter, we investigate a generalisation of Dung's AFs, namely, abstract dialectical frameworks (ADFs) (Brewka et al., 2018a), which are

known as an advanced abstract formalism for argumentation covering several generalizations of AFs (Brewka et al., 2014; Polberg, 2017; Dvořák et al., 2020). This is accomplished by acceptance conditions which specify, for each argument, its relation to its neighbour arguments via propositional formulas. These conditions determine the links between the arguments which can be, in particular, attacking or supporting.

In formal argumentation one is interested in investigating 'How is it possible to evaluate arguments in a given formalism?' Answering this question leads to the introduction of several types of semantics. In AFs, one starts with selecting a set of arguments without any conflicts. Conflict-freeness is a main characteristic of all types of semantics of AFs. Very often a new semantics is an improvement of an already existing one by introducing further restrictions on the set of accepted arguments or possible attackers. A list of semantics of AFs is presented in (Dung, 1995), namely conflict-free, admissible, complete, preferred, and stable semantics. Further semantics for AFs have been introduced later on, for instance, stage semantics (Verheij, 1996), semi-stable semantics first in (Verheij, 1996) (under a different name) then further investigated in (Caminada, 2006), ideal semantics (Dung et al., 2007), and eager semantics (Caminada, 2007b). Each semantics presents a point of view on accepting arguments.

Most of the semantics of AFs have been defined for ADFs and it has been shown that semantics of ADFs are generalizations of semantics of AFs (Brewka et al., 2018a; Gaggl et al., 2021). In this work, we focus on semi-stable semantics for ADFs, in a way that follows the same idea of semi-stable semantics of AFs. To this end, we first present a weaker version of the two-valued models of ADFs, which we call semi-two-valued models. Then we define semi-stable models for ADFs as a special case of semi-two-valued models of a given ADF. The relation between semi-two-valued and semi-stable models is similar to the relation between two-valued and stable models for ADFs. The difference is that a stable model is chosen among two-valued models, however, a semi-stable model will be chosen among semi-two-valued models of a given ADF.

Some of the semantics have become popular in the domain of argumentation, such as grounded semantics, preferred semantics and stable semantics. Each AF has a unique grounded extension, and one or more preferred extensions. However, it is possible that an AF does not have any stable extension. Because of this shortcoming of stable semantics, in order to pick at least one set of arguments, preferred and grounded semantics become more popular in argumentation. In contrast, stable semantics still

enjoys a strong support in logic programming (Gelfond and Lifschitz, 1988) and answer set programming (Gelfond and Lifschitz, 1991), since it is preferred to have no outcome as opposed to an imperfect one. On the one hand, in argumentation a grounded extension presents the least amount of information about the acceptance of arguments. That is, a grounded extension collects a set of arguments about which there is no doubt. In other words, the grounded extension of a given AF is very skeptical. On the other hand, it is possible that an AF has a stable extension but the set of preferred extensions and stable extensions are not equal.

To overcome this deficiency, semi-stable semantics have been introduced for AFs. Semi-stable semantics is a way of approximating stable semantics when a given AF does not have any stable extension. Key characteristics of semi-stable semantics in AFs are:

1. It is placed between stable semantics and preferred semantics;
2. If an AF has at least one stable extension, then the set of stable extensions and the set of semi-stable extensions coincide;
3. Each finite AF has at least one semi-stable extension.[1]

Computational complexity of semi-stable semantics is studied in (Dunne and Caminada, 2008). Furthermore, (Caminada, 2007a) presents an algorithm to compute semi-stable semantics of AFs.

In this chapter we propose a notion of semi-stable semantics for ADFs. First we discuss required properties for such a semantics in order to ensure that our notion is a proper generalization of the notion of semi-stable semantics for AFs. Then we define our notion of semi-stable semantics for ADFs and study its properties. It turns out that our notion fulfills the required properties presented in Section 5.1.1.

### 5.1.1 Requirements of Semi-Stable Semantics

For AFs, the property holds that a semi-stable extension is stable in the AF restricted to the arguments that have a truth value (accepted/rejected, in/out). This holds in general, and in particular also for AFs that have no stable extension. In the current work we follow this same idea to extend the notion of semi-stable semantics of AFs for ADFs.

In ADFs, the notion of stable model is defined based on the notion of two-valued model. An ADF may have no stable model. On the one

[1] (Verheij, 2003b, Example 5.8) shows that existence is not guaranteed for infinite AFs. See also (Caminada and Verheij, 2010).

hand, if a given ADF does not have any two-valued model, then it does not have any stable model. On the other hand, an ADF may have two-valued models, while none of them is a stable model. We focus on the first issue here. To define the notion of semi-stable semantics for ADFs, we follow the same method as for stable semantics of ADFs. That is, first we introduce the notion of semi-two-valued semantics. Subsequently, we pick semi-stable models among semi-two-valued models of a given ADF. A *semi-two-valued model* is a complete interpretation, that is, the number of arguments that are assigned to unknown is $\subseteq$-minimal among all complete interpretations. Further, a *semi-stable model* is a semi-two-valued model $v$ that has a constructive proof for arguments that are assigned to $\mathbf{t}$ in $v$. We show that the semi-stable semantics and semi-two-valued semantics presented in this work will satisfy the following conditions, which are akin to the properties of the notion of semi-stable semantics of AFs.

1. A semi-stable/semi-two-valued model of a given ADF should maximize the union of the sets of the accepted and of the rejected/denied arguments among all complete interpretations, with respect to subset inclusion;
2. Each semi-stable/semi-two-valued model is a preferred interpretation;
3. Each stable model is a semi-stable/semi-two-valued model;
4. Each finite ADF has at least one semi-two-valued model;
5. If an ADF has a stable model, then the set of stable models coincides with the set of semi-stable models;
6. The notion of semi-stable/semi-two-valued semantics for ADFs is a proper generalization of semi-stable semantics for AFs.

This chapter is structured as follows. In Section 5.2.1, we present the relevant background of AFs, i.e., the notion of semi-stable semantics of AFs. Then, in Section 5.2.2, we present definitions of semi-two-valued/semi-stable semantics for ADFs. In this section, we show that the notion of semi-stable semantics and semi-two-valued semantics satisfy the required properties, items $1-5$, presented above in this section. Further, in Section 5.3 we show that the notion of semi-stable/semi-two-valued semantics of ADFs is a proper generalization of the concept of semi-stable semantics of AFs, cf. the 6th property. In Section 5.4, we present the conclusion of our work. Furthermore, we briefly discuss a related research, in particular, (Alcântara and Sá, 2018) has also proposed a notion of semi-stable semantics for ADFs.

## 5.2 Semi-stable Semantics

A main goal of this section is to introduce the notion of semi-stable semantics of ADF. First, in Section 5.2.1 we recall the definition of semi-stable semantics of AFs, presented in (Caminada, 2006). Then, in Section 5.2.2 we propose the notion of semi-stable semantics of ADFs as a way of approximating stable semantics of a given ADF $D$, when $D$ have no stable model.

### 5.2.1 Semi-stable Semantics for AFs

The notion of semi-stable semantics of AFs has been presented in Section 2.3.1, comprehensively. Here we recall the notion briefly. Semi-stable semantics, introduced in (Verheij, 1996) (under a different name) then further investigated in (Caminada, 2006), we recall it in Definition 5.1, is a way of approximating stable semantics when a given AF does not have any stable extension.

**Definition 5.1** *(Caminada, 2006) Let $F = (A, R)$ be an AF and let $S$ be an extension of $F$. For $a \in A$, we write $a^+ = \{b \mid (a, b) \in R\}$ and $S^+ = \cup\{a^+ \mid a \in S\}$. Set $S$ is called a semi-stable extension iff $S$ is a complete extension where $S \cup S^+$ is maximal.*

The set of semi-stable extensions of $F$ is denoted by *semi-stb*$(F)$. Some alternative definitions of semi-stable semantics of AFs are also presented in (Caminada, 2006), as follows. Extension $S$ of $F$ is a semi-stable semantics if:

- $S$ is a preferred extension where $S \cup S^+$ is maximal.
- $S$ is an admissible extension where $S \cup S^+$ is maximal.

Semi-stable semantics in an AF is placed between stable semantics and preferred semantics. Each finite AF has at least one semi-stable extension. Further, if an AF has at least one stable extension, then the set of stable semantics and semi-stable semantics coincide, presented in Theorem 2.21.

### 5.2.2 Semi-stable Semantics for ADFs

The notion of stable semantics for ADFs is defined following similar ideas from logic programming. Stable models extend the concept of minimal model in logic programming by excluding self-justifying cycles of atoms.

The concept of stable semantics of ADFs has been presented in (Brewka et al., 2013, Definition 6) and in (Brewka et al., 2017a, Definition 18); we recall it in Definition 5.2.

**Definition 5.2** *Let $D$ be an ADF and let $v$ be a two-valued model of $D$. Then $v$ is a stable model of $D$ if $v^{\mathbf{t}} = w^{\mathbf{t}}$, where $w$ is the grounded interpretation of the stb-reduct $D^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}}$, $L^v = L \cap (A^v \times A^v)$, and $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$.*

Intuitively, the grounded interpretation collects all the information that is beyond any doubt, thus, it is said that there is a constructive proof for all arguments presented in the grounded interpretation. Hence, a two-valued model $v$ of $D$ is a stable interpretation (model), if there exists a constructive proof for all arguments assigned to true in $v$, in case all arguments that are assigned to false in $v$ are actually false. Example 5.3 clarifies the notion of stable semantics of ADFs.

**Example 5.3** *Let $D = (\{a, b, c\}, \{\varphi_a : \neg b, \varphi_b : b \vee \neg c, \varphi_c : \neg a \vee \neg b\})$ be an ADF, depicted in Figure 5.1. $D$ has two two-valued models, namely $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$ and $v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$. We check whether they are stable models. To investigate whether $v_1$ is a stable model, first we evaluate the stb-reduct of $D$ under $v_1$, namely $D^{v_1} = (A^{v_1}, L^{v_1}, C^{v_1})$. Here $A^{v_1} = \{a, c\}$, $L^{v_1} = \{(a, c)\}$, and $\varphi_a : \neg\bot \equiv \top$ and $\varphi_c : \neg a \vee \neg\bot \equiv \top$. The reduct $D^{v_1}$ is depicted in Figure 5.2 (on the left). Since the unique grounded interpretation of $D^{v_1}$ is $w = \{a \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$, i.e., $w^{\mathbf{t}} = v_1^{\mathbf{t}}$, two-valued model $v_1$ is a stable model of $D$.*

*However, we show that $v_2$ is not a stable model of $D$. To this end, we first evaluate $D^{v_2} = (A^{v_2}, L^{v_2}, C^{v_2})$, where $A^{v_2} = \{b, c\}$, $L^{v_2} = \{(b, b), (b, c), (c, b)\}$, and $\varphi_b : b \vee \neg c$ and $\varphi_c : \neg\bot \vee \neg b \equiv \top$, depicted in Figure 5.2 (on the right). Since the unique grounded interpretation of $D^{v_2}$ is $w = \{b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\}$, i.e., $w^{\mathbf{t}} \neq v_2^{\mathbf{t}}$, two-valued model $v_2$ is not a stable model of $D$. Intuitively, model $v_2$ is not a stable model of $D$, since in $v_2$ the acceptance of $b$ depends on $b$ itself, that is, there is a cyclic justification. Thus, $v_2$ violates the main condition of stable semantics that a stable model should have no self-justifying cycles of atoms. Thus, $stb(D) = \{v_1\}$.*

An ADF may have no stable model. Example 5.4 presents an ADF that has a two-valued model, but no stable model.

**Example 5.4** *Let $D = (\{a, b, c\}, \{\varphi_a : c \vee b, \varphi_b : c, \varphi_c : a \leftrightarrow b\})$, depicted in Figure 5.3. The only two-valued model of $D$ is $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$. However, $w^{\mathbf{t}} = \{\}$ where $w$ is the grounded interpretation of $D^v$. Thus, $w^{\mathbf{t}} \neq v^{\mathbf{t}}$. Hence, $v$ is not a stable model of $D$.*

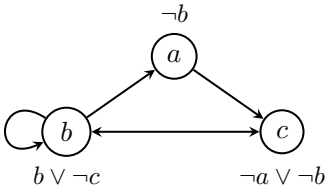


**Figure 5.1:** The ADF of Example 5.3.

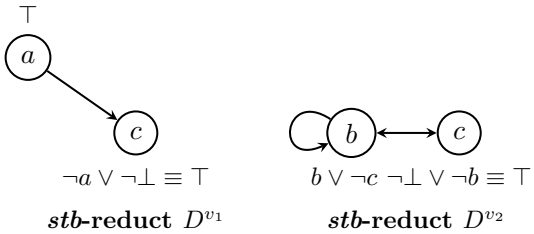


**Figure 5.2:** The reduct of ADF $D$ of Example 5.3.

Before providing the formal definition of semi-stable semantics for ADFs, we present the intuition why an ADF may have no stable models. An ADF $D$ may not have any stable model due to either of the following two reasons:

1. $mod(D) = \emptyset$, i.e., $D$ does not have any two-valued models from which to pick a stable model; or,
2. $mod(D) \neq \emptyset$, but for any $v \in mod(D)$ it holds that $v \notin stb(D)$; that is, when there is no constructive proof for arguments that are assigned to **t** in $v$ where $v \in mod(D)$.

Nonetheless, there are many cases about which one might want to draw a conclusion even when a given ADF does not have any two-valued model or stable model. One option is focusing on other semantics like preferred and grounded semantics that exist for any ADF. However, a unique grounded interpretation presents a piece of information about those arguments about which there is no doubt. That is, it is possible that in a given ADF the grounded interpretation has less information than each of its stable models. In other words, the information of the grounded interpretation is too skeptical. Furthermore, there exists an ADF $D$ such that $stb(D) \neq \emptyset$ but the set of stable semantics of $D$ and the set of preferred semantics of $D$ are not equivalent, i.e., $stb(D) \subsetneq prf(D)$. That is, by preferred semantics some non-stable models may be introduced, even in the case that a stable

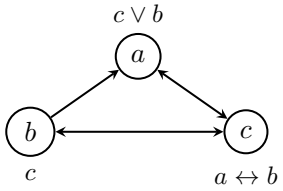


**Figure 5.3:** The ADF of Example 5.4

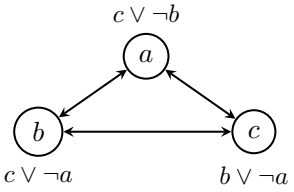


**Figure 5.4:** The ADF of Example 5.5

model exists. Example 5.5 is an instance of an ADF such that $stb(D) \neq \emptyset$ but $stb(D) \subsetneq prf(D)$.

**Example 5.5** *Let* $D = (\{a, b, c\}, \{\varphi_a : c \vee \neg b, \varphi_b : c \vee \neg a, \varphi_c : b \vee \neg a\})$ *be an ADF, depicted in Figure 5.4. The set of preferred interpretations of* $D$ *is* $prf(D) = \{\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}\}$. *Both of the preferred interpretations of* $D$ *are two-valued models of* $D$. *However,* $stb(D) = \{\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}\}$. *That is,* $stb(D) \subsetneq prf(D)$.

*Furthermore, the unique grounded interpretation of* $D$ *is the trivial interpretation that has strictly less information than the stable model of* $D$, *i.e.,* $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$.

Still, it is interesting to present a semantics for ADFs that is equal to stable semantics if there exists a stable model. In ADFs, to define the notion of stable semantics as it is in (Brewka et al., 2018a), first the notion of two-valued semantics is introduced. Then a two-valued model is called a stable model if it satisfies the conditions of Definition 5.2, i.e., if it does not contain any support cycle. Since in AFs there is no support cycle, these two notions are equal. That is, for the associated ADF $D_F$ of a given AF $F$ it holds that $mod(D_F) = stb(D_F)$. Due to this distinction between two-valued models and stable models in ADFs, different levels of semi-stable semantics can be considered in ADFs for the notion of semi-stable semantics of AFs. Here we follow a similar method as presented

in (Brewka et al., 2018a) for stable semantics to present the concept of semi-stable semantics.

The first reason that an ADF $D$ does not have any stable semantics is that $D$ does not have any two-valued model. We focus on this issue to present an alternative semantics for stable semantics of ADFs. In this alternative option, i.e., semi-stable semantics for ADFs, we are looking for a semi-two-valued model, which is a partially two-valued model, presented in Definition 5.6, that satisfies the condition of Definition 5.2, that is, it does not contain any support cycles among arguments. These new points of view of acceptance of arguments, which are called *semi-two-valued semantics* and *semi-stable semantics* of ADFs, have to satisfy the requirements presented in Section 5.1.1. The properties in Section 5.1.1 are akin to the properties of the notion of semi-stable semantics of AFs, presented in Theorem 2.21.

**Definition 5.6** *Let $D$ be an ADF and let $v$ be an interpretation of $D$. An interpretation $v$ is a semi-two-valued model (interpretation) of $D$ if the following conditions hold:*

1. *$v$ is a complete interpretation of $D$;*
2. *$v^{\mathbf{u}}$ is $\subseteq$-minimal among all $w^{\mathbf{u}}$ such that $w$ is a complete interpretation of $D$.*

The set of semi-two-valued models of $D$ is denoted by *semi-mod*$(D)$. Note that when an ADF has a two-valued model, then the set of semi-two-valued models and the set of two-valued models coincide, which is shown in Lemma 5.15. We introduce the concept of semi-stable models over the notion of semi-two-valued models in Definition 5.7.

**Definition 5.7** *Let $D$ be an ADF and let $v$ be a semi-two-valued model of $D$. An interpretation $v$ is a semi-stable model (interpretation) of $D$ if the following condition holds:*

- *$v^{\mathbf{t}} = w^{\mathbf{t}}$ s.t $w$ is the grounded interpretation of sub-reduct $D^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}} \cup v^{\mathbf{u}}$, $L^v = L \cap (A^v \times A^v)$, and $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$.*

The set of semi-stable models of $D$ is denoted by *semi-stb*$(D)$. Note that in Definition 5.7, in sub-reduct $D^v$ we assume that $v$ is a semi-two-valued model (complete interpretation) of $D$, however, in Definition 5.2, in sub-reduct $D^v$ it is assumed that a given interpretation $v$ is a two-valued model

of $D$. Since in Definition 5.7, interpretation $v$ is a semi-two-valued model, it may contain an argument that is assigned to $\mathbf{u}$. Therefore, in sub-reduct $D^v$ in Definition 5.7, we keep those arguments that are assigned to $\mathbf{u}$ in $v$ as well, i.e., $A^v = v^{\mathbf{t}} \cup v^{\mathbf{u}}$. Arguments that are assigned to $\mathbf{u}$ in $v$ will remain in $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$. Intuitively, a complete interpretation $v$ is a semi-stable model of $D$ if $v^{\mathbf{u}}$ is $\subseteq$-minimal among complete interpretations of $D$ and there exists a constructive proof for arguments which are assigned to $\mathbf{t}$ in $v$, in case all arguments which are assigned to false in $v$ are actually false. Corollary 5.8 is a direct result of Definition 5.1, which defines the notion of semi-stable model over the set of semi-two-valued models of a given ADF.

**Corollary 5.8** *Let $D$ be an ADF. Each semi-stable model of $D$ is a semi-two-valued model of $D$.*

Example 5.9 clarifies the notion of semi-stable semantics of ADFs.

**Example 5.9** *Let $D = (\{a, b, c\}, \{\varphi_a : \neg a, \varphi_b : c \wedge (\neg a \vee c), \varphi_c : b \wedge (a \vee b)\})$ be an ADF, depicted in Figure 5.5. The set of preferred interpretations of $D$ is $prf(D) = \{\{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}\}$. None of the preferred interpretations is a two-valued model. Thus, $D$ does not have any stable model. Both $v_1 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$ and $v_2 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$ are complete interpretations of $D$. Furthermore, both $v_1$ and $v_2$ are semi-two-valued models of $D$, since $v_1^{\mathbf{u}} = v_2^{\mathbf{u}} = \{a\}$. However, we show that only $v_2$ is a semi-stable model of $D$. To this end, we first evaluate sub-reduct $D^{v_2}$. Since no argument is assigned to $\mathbf{t}$ in $v_2$ and only $a$ is assigned to $\mathbf{u}$ in $v$, $A^{v_2} = \{a\}$. Thus, $D^{v_2} = (\{a\}, \{\varphi_a : \neg a\})$, depicted in Figure 5.6. It is clear that the unique grounded interpretation $D^{v_2}$ is $w = \{a \mapsto \mathbf{u}\}$. Since $w^{\mathbf{t}} = v_2^{\mathbf{t}} = \emptyset$, it holds that $v_2$ is a semi-stable model of $D$.*

*On the other hand, $v_1$ is not a semi-stable model of $D$. In $v_1$, both $b$ and $c$ are assigned to $\mathbf{t}$ and $a$ is assigned to $\mathbf{u}$, therefore, $A^{v_1} = A$. Since no argument is assigned to $\mathbf{f}$ in $v_1$, we have $D^{v_1} = D$. The grounded interpretation of $D/D^{v_1}$ is $w = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}$. That is, $w^{\mathbf{t}} = \emptyset$. However, $v_1^{\mathbf{t}} = \{b, c\}$, i.e., $w^{\mathbf{t}} \neq v_1^{\mathbf{t}}$. Thus, $v_1$ is not a semi-stable model of $D$.*

Proposition 5.10 shows the first property of semi-two-valued model presented in Section 5.1.1.

**Proposition 5.10** *Let $D$ be an ADF, and let $v$ be a semi-two-valued model of $D$. It holds that $v$ maximizes the union of the sets of the accepted*

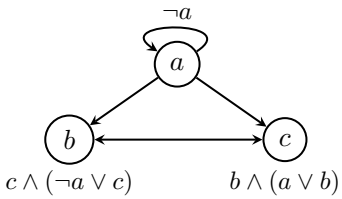


**Figure 5.5:** The ADF of Example 5.9


**Figure 5.6:** reduct $D^{v_2}$ of ADF of Example 5.9

*and of the denied among all complete interpretations of $D$, i.e., $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is maximal with respect to subset inclusion.*

**Proof** Let $D = (A, L, C)$ be a given ADF. Assume that $v$ is a semi-two-valued model of $D$. Toward a contradiction, assume that $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is not $\subseteq$-maximal among all complete interpretations of $D$. Thus, there exists a complete interpretation $w$ such that $w^{\mathbf{t}} \cup w^{\mathbf{f}}$ is $\subseteq$-maximal. Thus, it holds that $v^{\mathbf{t}} \cup v^{\mathbf{f}} \subsetneq w^{\mathbf{t}} \cup w^{\mathbf{f}}$. Hence, it holds that $w^{\mathbf{u}} \subsetneq v^{\mathbf{u}}$, i.e., $v^{\mathbf{u}}$ is not minimal among complete interpretations of $D$. That is, by Definition 5.6, $v$ is not a semi-two-valued model of $D$. This contradicts the assumption that $v$ is a semi-two-valued model of $D$. Thus, the assumption that $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is not $\subseteq$-maximal among complete interpretations is wrong. □

Proposition 5.10 clarifies the distinction between preferred semantics and semi-two-valued models of ADFs. While interpretation $v$ is a preferred interpretation of $D$ if it is $\leq_i$-maximal in $com(D)$, interpretation $v$ is a semi-two-valued model of $D$ if $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal in $com(D)$. Corollary 5.11 is a direct result of Proposition 5.10 and the fact that each semi-stable model is a semi-two-valued model.

**Corollary 5.11** *Let $D$ be an ADF, and let $v$ be a semi-stable model of $D$. It holds that $v$ maximizes the union of the sets of the accepted and of the denied among all complete interpretations of $D$, i.e., $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal in $com(D)$.*

Theorem 5.12 presents the second and the third required properties for semi-stable/semi-two valued semantics for ADFs, presented in Section 5.1.1.

**Theorem 5.12** *Let $D$ be an ADF.*

1. *Each semi-two-valued model of $D$ is a preferred interpretation of $D$;*
2. *Each semi-stable model of $D$ is a preferred interpretation of $D$;*
3. *Each stable model of $D$ is a semi-two-valued model of $D$;*
4. *Each stable model of $D$ is a semi-stable model of $D$.*

**Proof** Let $D$ be an ADF.

- Proof of item 1: assume that $v$ is a semi-two-valued model of $D$. We show that $v$ is a preferred interpretation of $D$. Toward a contradiction, assume that $v \notin \mathit{prf}(D)$. By Definition 5.6, $v$ is a complete interpretation of $D$. That is, if $v$ is not a preferred interpretation, then there exists a preferred interpretation $v'$ such that $v <_i v'$. Thus, $v'^{\mathbf{u}} \subsetneq v^{\mathbf{u}}$. Hence, by Definition 5.6, $v$ is not a semi-two-valued model of $D$. This contradicts the assumption that $v$ is a semi-two-valued model of $D$. Therefore, the assumption that $v$ is not a preferred interpretation of $D$ is wrong.

- Proof of item 2: assume that $v$ is a semi-stable model of $D$. We show that $v$ is a preferred interpretation of $D$. By Corollary 5.8, each semi-stable model of $D$ is a semi-two-valued model of $D$. Thus, $v$ is a semi-two-valued model of $D$. By the first item of this theorem, $v$ is a preferred interpretation of $D$.

- Proof of item 3: Assume that $v$ is a stable model. First, each stable model is a complete interpretation. Thus, the first item of Definition 5.6 is satisfied. Second, each stable model is a two-valued model, i.e., $v^{\mathbf{u}} = \emptyset$. Thus, $v^{\mathbf{u}}$ is $\subseteq$-minimal among all $w^{\mathbf{u}}$, where $w$ is a complete interpretation of $D$. Hence, the second item of Definition 5.6 is satisfied. Thus, $v$ is a semi-two-valued model of $D$.

- Proof of item 4: assume that $v$ is a stable model. By the previous item, $v$ is a semi-two-valued model of $D$. We show that $v$ satisfies the condition of Definition 5.7. Since $v$ is a stable model, by Definition 5.2, $v^{\mathbf{t}} = w^{\mathbf{t}}$ such that $w$ is the grounded interpretation of sub-reduct $D^v = (A^v, L^v, C^v)$. Since $v^{\mathbf{u}} = \emptyset$, in Definition 5.7, $A^v = v^{\mathbf{t}}$. That is, Definition 5.7 (semi-stable model) and Definition 5.2 (stable-model) coincide for $v$. Thus, if $v$ is a stable model of $D$, then $v$ is a semi-stable model of $D$.

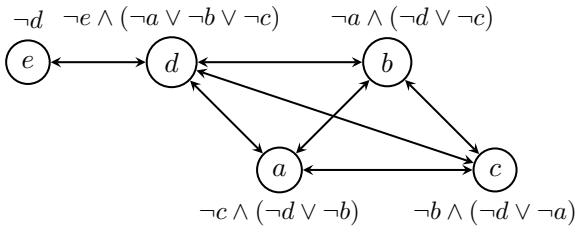


**Figure 5.7:** An ADF with a preferred interpretation that is not a semi-stable/semi-two-valued model

□

The first two items of Theorem 5.12 imply that the set of semi-stable/semi-two-valued models of an ADF $D$ is a subset of the set of preferred interpretations of $D$, i.e., *semi-stb*$(D) \subseteq$ *prf*$(D)$ and *semi-mod*$(D) \subseteq$ *prf*$(D)$. However, Proposition 5.13 indicates that the notion of preferred semantics coincides neither with the notion of semi-stable semantics nor with the notion of semi-two-valued semantics. That is, there exists an ADF $D$ such that *prf*$(D) \not\subseteq$ *semi-stb*$(D)$ and *prf*$(D) \not\subseteq$ *semi-mod*$(D)$.

**Proposition 5.13** *There is an ADF $D$ such that the set of preferred interpretations of $D$ does not coincide with the set of semi-stable models, nor with the set of semi-two-valued models of $D$.*

**Proof** We show that there exists an ADF with a preferred interpretation which is not a semi-two-valued model. To this end, we use the ADF presented in ((Diller et al., 2020, Theorem 6 )). Consider ADF $D = (\{a, b, c, d, e\}, \{\varphi_a : \neg c \wedge (\neg d \vee \neg b), \varphi_b : \neg a \wedge (\neg d \vee \neg c), \varphi_c : \neg b \wedge (\neg d \vee \neg a), \varphi_d : \neg e \wedge (\neg a \vee \neg b \vee \neg c), \varphi_e : \neg d\})$, depicted in Figure 5.7. $D$ has four preferred interpretations, namely $v_1 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f},\ c \mapsto \mathbf{t},\ d \mapsto \mathbf{t},\ e \mapsto \mathbf{f}\}$, $v_2 = \{a \mapsto \mathbf{f},\ b \mapsto \mathbf{t},\ c \mapsto \mathbf{f},\ d \mapsto \mathbf{t},\ e \mapsto \mathbf{f}\}$, $v_3 = \{a \mapsto \mathbf{t},\ b \mapsto \mathbf{f},\ c \mapsto \mathbf{f},\ d \mapsto \mathbf{t},\ e \mapsto \mathbf{f}\}$, and $v_4 = \{a \mapsto \mathbf{u},\ b \mapsto \mathbf{u},\ c \mapsto \mathbf{u},\ d \mapsto \mathbf{f},\ e \mapsto \mathbf{t}\}$. It holds that $v_1$, $v_2$, $v_3$ are semi-two-valued models/two-valued models of $D$, since $v_1^{\mathbf{u}} = v_2^{\mathbf{u}} = v_3^{\mathbf{u}} = \emptyset$. However, $v_4$ is not a semi-two-valued/semi-stable model, since $v_4^{\mathbf{u}} = \{a, b, c\}$, that is, $v_4^{\mathbf{u}}$ is not $\subseteq$-minimal among $v_i^{\mathbf{u}}$, for $1 \leq i \leq 4$. Thus, in ADFs, the notion of preferred semantics is not equal to the notion of semi-stable/semi-two-valued semantics.

□

Proposition 5.14 presents the fourth property required for semi-two-valued semantics, presented in Section 5.1.1.

**Proposition 5.14** *Each ADF has at least one semi-two-valued model.*

**Proof** Let $D$ be an ADF. Each ADF has a unique grounded interpretation. By the facts that the grounded interpretation is the least fixed point of $\Gamma_D$ and the grounded interpretation is a least complete interpretation with respect to the $\leq_i$-ordering, we conclude that each ADF has at least one complete interpretation. By Definition 5.6, each semi-two-valued model $v$ is a complete interpretation where $v^{\mathbf{u}}$ is $\subseteq$-minimal among other complete interpretations of $D$. Since the number of arguments is finite, the set of complete interpretations is finite. That is, there exists a complete interpretation $v$ where $v^{\mathbf{u}}$ is $\subseteq$-minimal among all compete interpretations of $D$. Thus, the set of semi-two-valued models of $D$ is non-empty. □

In Theorem 5.18, we show the fifth property of semi-stable semantics, presented in Section 5.1.1: If an ADF $D$ has a stable model, then the set of stable models and the set of semi-stable models of $D$ coincide. To show this theorem, we need some auxiliary results that are shown in Lemmas 5.15–5.17.

**Lemma 5.15** *Let $D$ be an ADF. Assume that $D$ has a two-valued model. Then, the set of semi-two-valued models of $D$ and the set of two-valued models of $D$ coincide.*

**Proof** Assume that $D$ has a two-valued model $v$. Since $v$ is a two-valued model, it holds that $v^{\mathbf{u}} = \emptyset$. Thus, by Definition 5.6, $v$ is a semi-two-valued model, i.e., $\mathit{mod}(D) \subseteq \mathit{semi\text{-}mod}(D)$. It remains to show that every semi-two-valued model of $D$ is also a two-valued model. Toward a contradiction, assume that $D$ has a semi-two-valued model $w$ which is not a two-valued model. Since $w$ is a semi-two-valued but not a two-valued model, it holds that $w$ is a complete interpretation and $w^{\mathbf{u}} \neq \emptyset$. However, since $D$ has a two-valued model, $w^{\mathbf{u}}$ is not $\subseteq$-minimal among all complete interpretations of $D$. That is, by Definition 5.6, $w$ is not a semi-two-valued model. This contradicts the assumption that $w$ is a semi-two-valued model. That is, if $D$ has a two-valued model, then $\mathit{semi\text{-}mod}(D) \subseteq \mathit{mod}(D)$. Hence, if ADF $D$ has a two-valued model, then $\mathit{semi\text{-}mod}(D) = \mathit{mod}(D)$. □

**Lemma 5.16** *Let $D$ be an ADF. Assume that $D$ has a stable model. Then, the set of semi-two-valued models of $D$ and the set of two-valued models of $D$ coincide.*

**Proof** Let $D$ be an ADF that has a stable model $v$. By the fact that each stable model of a given ADF is a two-valued model, it holds that $v$ is a two-valued model. By Lemma 5.15, if there exists a two-valued model, then the set of two-valued models and the set of semi-two-valued models coincide. So if an ADF has a stable model, then *semi-mod*$(D) = mod(D)$. □

**Lemma 5.17** *Let $D$ be an ADF. Assume that $D$ has a stable model. Then each semi-stable model of $D$ is a two-valued model of $D$.*

**Proof** Let $D$ be an ADF that has at least one stable model $v$. By Lemma 5.16, the set of semi-two-valued models of $D$ coincides with the set of two-valued models of $D$, i.e., *semi-mod*$(D) = mod(D)$. Moreover, by Corollary 5.8, each semi-stable model of $D$ is a semi-two-valued model of $D$, i.e., *semi-stb*$(D) \subseteq$ *semi-mod*$(D)$. Thus, if $D$ has a stable model, then each semi-stable model of $D$ is a two-valued model of $D$, i.e., *semi-stb*$(D) \subseteq$ $mod(D)$. □

**Theorem 5.18** *If ADF $D$ has a stable model, then the sets of stable models and semi-stable models of $D$ coincide.*

**Proof** Let $D$ be an ADF. By the forth item of Theorem 5.12, each stable model of $D$ is a semi-stable model of $D$, i.e., *stb*$(D) \subseteq$ *semi-stb*$(D)$.

Assume that $D$ has a stable model $v$ and a semi-stable model $v'$. We show that $v'$ is a stable model of $D$. Toward a contradiction, assume that $v'$ is not a stable model of $D$. By Lemma 5.17, $v'$ is a two-valued model of $D$, i.e., $v'^{\mathbf{u}} = \emptyset$. If $v'$ is not a stable model of $D$, by Definition 5.2, it has to be held that $v'^{\mathbf{t}} \neq w^{\mathbf{t}}$ where $w$ is the grounded interpretation of the *stb*-reduct $D^{v'} = (A^{v'}, L^{v'}, C^{v'})$, where $A^{v'} = v'^{\mathbf{t}}$, $L^{v'} = L \cap (A^{v'} \times A^{v'})$, and $\varphi_a[p/\bot : v'(p) = \mathbf{f}]$ for each $a \in A^{v'}$. This implies that the condition of Definition 5.7 does not hold for $v'$, since $v'^{\mathbf{u}} = \emptyset$. Thus, $v'$ is not a semi-stable model of $D$. This is a contradiction by the assumption that $v'$ is a semi-stable model of $D$. Hence, the assumption that $D$ has a semi-stable model which is not a stable-model is wrong. Thus, if $D$ has a stable model, then *semi-stb*$(D) \subseteq$ *stb*$(D)$. Hence, if ADF $D$ has a stable model, then *stb*$(D) =$ *semi-stb*$(D)$.

□

Proposition 5.14 says that each ADF has at least one semi-two-valued model. In contrast, Proposition 5.19 shows that an ADF may have no semi-stable model. As we presented in the beginning of this section the notions of semi-two-valued semantics and semi-stable semantics of ADFs together fulfil the properties required for the concept of semi-stable semantics, presented in Section 5.1.1.

**Proposition 5.19** *There exists an ADF that does not have any semi-stable model.*

**Proof** Let $D$ be the ADF presented in Example 5.4, i.e., $D = (\{a, b, c\}, \{\varphi_a : c \vee b, \varphi_b : c, \varphi_c : a \leftrightarrow b\})$. We showed, in Example 5.4, that $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$ is a two-valued model of $D$, however, it is not a stable model of $D$. Thus, by Lemma 5.15, $v$ is a semi-two-valued model of $D$. As we know $grd(D^v) = \{\emptyset\}$, however, $v^{\mathbf{t}} = \{a, b, c\}$. Thus, $v$ is not a semi-stable model. □

**Corollary 5.20** *Let $D$ be an ADF that has a two-valued model. If none of the two-valued models $D$ is a stable model of $D$, then $D$ does not have any semi-stable model.*

**Proof** Let $D$ be an ADF that has a two-valued model. Thus, by Lemma 5.15 the set of two-valued models of $D$ coincides with the set of semi-two-valued models of $D$. That is, for each two-valued model/semi-two-valued model $v$ of $D$ it holds that the condition of semi-stable model in Definition 5.7 coincides with the definition of stable model in Definition 5.2. Thus, if for each $v \in mod(D)$, $v$ is not a stable model, then $v$ is not a semi-stable model of $D$, as well. □

As Corollary 5.20 says, if an ADF has a two-valued model but no stable model, then it will not have any semi-stable model either. As we presented in the beginning of Section 5.2.2, the semi-stable semantics presented in this work deal with the first issue, namely, that an ADF may not have a stable model. That is, semi-stable semantics is a new point of view on the acceptance of arguments if an ADF does not have any two-valued model.

## 5.3 Generalization of the Semi-stable Semantics of AFs

In this section, we show that the notions of semi-stable and semi-two-valued semantics for ADFs satisfy the last property presented in Section 5.1.1,

required for these semantics. To this end, we show that the concept of semi-stable/semi-two-valued semantics for ADFs is a proper generalization of the concept of semi-stable semantics for AFs (Verheij, 1996; Caminada, 2006), in Theorems 5.23 and 5.25. Furthermore, we show that the concepts of semi-stable models and semi-two-valued models coincide for the associated ADF of a given AF, in Proposition 5.24.

Given an AF $F = (A, R)$ and its corresponding ADF $D_F = (A, R, C)$ (see Definition 2.53), the set of all possible conflict-free extensions of $F$is denoted by $\mathscr{E}$ and the set of all possible conflict-free interpretations of $D_F$ is denoted by $\mathscr{V}$. The functions $Ext2Int_F$ and $Int2Ext_{D_F}$ in Definitions 5.21–5.22 are modifications of the labelling functions as given in (Baroni et al., 2018a). Function $Ext2Int_F(e)$ represents the interpretation associated to a given extension $S$ in $F$, and function $Int2Ext_{D_F}(v)$ indicates the extension associated to a given interpretation $v$ of $D_F$.

**Definition 5.21** *Let $F = (A, R)$ be an AF, and let $S$ be an extension of $F$. The truth value assigned to each argument $a \in A$ by the three-valued interpretation $v_S$ associated to $S$ is given by $Ext2Int_F : \mathscr{E} \to \mathscr{V}$ as follows.*

$$Ext2Int_F(S)(a) = \begin{cases} \mathbf{t} & \text{if } a \in S, \\ \mathbf{f} & \text{if } a \in S^+, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

It is shown in (Keshavarzi Zafarghandi et al., 2021d, Proposition 20) that if $S$ is a conflict-free extension of $F$, then $Ext2Int_F(S)$ is well-defined. Moreover, the basic condition that $S$ has to be a conflict-free extension is a necessary condition for $Ext2Int_F(S)$ being well-defined. By Definition 5.1, every semi-stable extension of an AF is a complete extension and a conflict-free extension. Thus, if $S$ is a semi-stable extension of AF $F$, then $Ext2Int_F(S)$ is well-defined. An interpretation of $D_F$ can be represented as an extension via the function $Int2Ext_{D_F}$, presented in Definition 5.22.

**Definition 5.22** *Let $D_F = (A, R, C)$ be the ADF associated with AF $F = (A, R)$, and let $v$ be an interpretation of $D_F$, that is, $v \in \mathscr{V}$. The associated extension $S_v$ of $v$ is obtained via application of $Int2Ext_{D_F} : \mathscr{V} \to \mathscr{E}$ on $v$, as follows:*

$$Int2Ext_{D_F}(v) = \{s \in S \mid s \mapsto \mathbf{t} \in v\}$$

Theorem 5.23 presents that the notion of semi-two-valued model semantics for ADFs is a generalization of the concept of semi-stable semantics for AFs.

**Theorem 5.23** *For any AF $F = (A, R)$ and its associated ADF $D_F = (A, R, C)$, the following properties hold:*

- *if $S$ is a semi-stable extension of $F$, then $Ext2Int_F(S)$ is a semi-two-valued model of $D_F$;*
- *if $v$ is a semi-two-valued model of $D_F$, then $Int2Ext_{D_F}(v)$ is a semi-stable extension of $F$.*

**Proof** Let $F$ be an AF and let $D_F$ be its associated ADF, as in Definition 2.53.

- We assume that $\{S_0, S_1, \ldots, S_k\}$ is the set of all complete extensions of $F$. Since $F$ is a finite AF, the set of complete extensions of $F$ is finite. Assume that $\{v_0, v_1, \ldots, v_k\}$ is the set of corresponding complete interpretations of $D_F$, i.e., $v_i = Ext2Int_F(S_i)$ for $i$ with $0 \leq i \leq k$.

  Without loss of generality, assume that $S_0$ is a semi-stable extension of $F$. By Definition 5.1, $S_0$ is a complete extension of $F$ such that $S_0 \cup S_0^+$ is maximal. We show that $v_0 = Ext2Int_F(S_0)$ is a semi-two-valued model of $D_F$. Since $v_0$ is a complete interpretation of $D_F$, to show that $v_0$ is a semi-two-valued interpretation of $D_F$, it remains to show that $v_0^{\mathbf{u}}$ is $\subseteq$-minimal among all $v_i{}^{\mathbf{u}}$ for $i$ with $0 < i \leq k$. Toward a contradiction, assume that $v_0{}^{\mathbf{u}}$ is not $\subseteq$-minimal among all $v_i{}^{\mathbf{u}}$ for $i$ with $0 < i \leq k$. Thus, there exists a $j$ for $0 < j \leq k$ such that $v_j{}^{\mathbf{u}} \subsetneq v_0{}^{\mathbf{u}}$. Thus, there exists an $a$ such that $a \notin v_j{}^{\mathbf{u}}$ and $a \in v_0{}^{\mathbf{u}}$. Thus, by Definition 5.21, it holds that, for each such an $a$, $a \in S_j \cup S_j{}^+$ but $a \notin S_0 \cup S_0{}^+$. Thus, $S_0 \cup S_0{}^+$ is not maximal. This contradicts the assumption that $S_0$ is a semi-stable extension of $F$. Hence, $v_0$ is a semi-two-valued model of $D_F$.

- Assume that $v$ is a semi-two-valued model of $D_F$; we show that $S = Int2Ext_{D_F}(v)$ is a semi-stable extension of $F$. To show that $S$ is a semi-stable extension of $F$, we show that $S \cup S^+$ is maximal. Toward a contradiction, assume that $S \cup S^+$ is not maximal. Thus, there exists a complete extension of $F$, namely $S'$, with $S' \cup S'^+$ is maximal, i.e., $S \cup S^+ \subsetneq S' \cup S'^+$. Thus, by Definition 5.21, it holds that $v \leq_i v'$, where $v' = Ext2Int(S')$. Thus, $v'$ is a complete interpretation of $D_F$ such that $v'^{\mathbf{u}} \subsetneq v^{\mathbf{u}}$. Hence, $v$ is not a semi-two-valued model of $D_F$. This contradicts the assumption that $v$ is a semi-two-valued model of $D_F$. Thus, the assumption that $S \cup S^+$ is

not maximal among all complete extensions of $F$ is wrong. Hence, $S$ is a semi-stable extension of $F$.

□

**Proposition 5.24** *Let $F = (A, R)$ be an AF and let $D_F$ be its associated ADF. The semi-two-valued semantics of $D_F$ coincide with the semi-stable models of $D_F$.*

**Proof** Let $F = (A, R)$ be an AF and let $D_F$ be its associated ADF. By Corollary 5.8, *semi-stb*$(D_F) \subseteq$ *semi-mod*$(D_F)$. Thus it remains to show that *semi-mod*$(D_F) \subseteq$ *semi-stb*$(D_F)$.

Assume that $v$ is a semi-two-valued model of $D_F$. To show that $v$ is a semi-stable model of $D_F$, we show that $v^{\mathbf{t}} = w^{\mathbf{t}}$, where $w$ is the grounded interpretation of sub-reduct $D_F^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}} \cup v^{\mathbf{u}}$. We show that $v^{\mathbf{t}} \subseteq w^{\mathbf{t}}$. Assume that $a \mapsto \mathbf{t} \in v$. Since $D_F$ is an associated ADF to AF $F$, $\varphi_a : \bigwedge_{b \in par(a)} \neg b$. Thus, if $a \in v^{\mathbf{t}}$, then either $a$ is an initial argument of $D_F$ or for each $b \in par(a)$ it holds that $b \in v^{\mathbf{f}}$. In both cases, it is clear that $\varphi_a[p/\bot : v(p) = \mathbf{f}] \equiv \top$. Therefore, $a \in w^{\mathbf{t}}$. Thus, $v^{\mathbf{t}} = w^{\mathbf{t}}$. Hence, $v$ is a semi-stable model of $D_F$. □

**Theorem 5.25** *For any AF $F = (A, R)$ and its associated ADF $D_F$, the following properties hold:*

- *if $S$ is a semi-stable extension of $F$, then $Ext2Int_F(S)$ is a semi-stable model of $D_F$;*
- *if $v$ is a semi-stable model of $D_F$, then $Int2Ext_{D_F}(v)$ is a semi-stable extension of $F$.*

**Proof** [Sketch] The theorem is a direct result of combining Theorem 5.23, which says that semi-two-valued semantics of ADFs are a generalization of semi-stable semantics of AFs, and Proposition 5.24, which says that in the associated ADF $D_F$ of a given AF $F$, the notions of semi-stable semantics and semi-two-valued semantics coincide. □

## 5.4 Conclusion

In this chapter, we have defined the semi-stable and semi-two-valued semantics for finite ADFs. From a theoretical perspective, in Sections 5.2.2 and 5.3, we observe that the notions of semi-stable and semi-two-valued semantics for ADFs fulfil the requirements for these two notions presented in Section 5.1.1.

An ADF may have no stable model, for one of two reasons:

1. $D$ does not have any two-valued model; or
2. each two-valued model contains a support cycle.

The condition presented in Definition 5.2 characterizes the stable semantics for ADFs. The condition says that a two-valued model is stable if it does not contain any support cycle, i.e., if there exists a constructive proof for the arguments that are assigned to $\mathbf{t}$. Thus, to present an alternative definition for stable semantics we focus on the first reason that an ADF does not have a stable model, and we present a partial two-valued semantics in Section 5.2.2, called semi-two-valued semantics in Definition 5.6. Then we define the notion of semi-stable semantics over semi-two-valued semantics in Definition 5.7.

In Section 5.2.2, we show that the notions of semi-two-valued/semi-stable semantics of ADFs presented in this work satisfy the main requirements presented in Section 5.1.1. Specifically:

1. Proposition 5.10 and Corollary 5.11 say that if $v$ is a semi-two-valued/semi-stable model of $D$, then $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal among all complete interpretations of $D$.
2. Theorem 5.12 says that each semi-stable/semi-two-valued model is a preferred interpretation and each stable model of an ADF is a semi-stable/semi-two-valued model of that ADF.
3. Proposition 5.14 says that each ADF has at least one semi-two-valued model.
4. Theorem 5.18 says that if an ADF has a stable model, then the sets of stable models and semi-stable models coincide.

In Section 5.3, we show that the notions of semi-stable/semi-two-valued semantics of ADFs are proper generalizations of the notion of semi-stable semantics of AFs. In Proposition 5.24, we show that the concepts of semi-stable and semi-two-valued semantics coincide in the associated ADF of a given AF, intuitively, since in AFs there cannot be a support cycle.

Alcântara and Sá (2018) have also considered the semi-stable semantics for ADFs. To prevent confusion with the notion of semi-stable semantics presented in the current work, we call their notion semi-stable2 semantics, abbreviated SSS2. A key difference between our notion and SSS2 is that ours is compatible with the standard ADF definitions. In particular, in their discussion, the characteristic operator $\Gamma_D$ and in addition, the semantics of ADFs, and specifically the complete semantics, have not been presented in the way as introduced by Brewka and Woltran (Brewka et al., 2018a; Brewka and Woltran, 2010). For instance, by their deviating definition of complete labelling (Alcântara and Sá, 2018), only $\{\neg a, \neg b\}$ is a complete labelling/grounded model of $D = (\{a, b\}, \{\varphi_a : b, \varphi_b : a\})$. Hence—unlike the standard definitions—the set of preferred labellings of $D$ is in their approach not a subset of the set of complete labellings of $D$, and the unique grounded labelling $\{\}$ is not a complete labelling.

The computational complexity of semantics of AFs and ADFs is presented in Dvořák and Dunne (2018). Computational complexity of semi-stable semantics of AFs is studied in Dunne and Caminada (2008). As a future work, it would be interesting to clarify the computational complexity of investigating:

1. whether a given interpretation is a semi-stable model , or a semi-two-valued model,
2. whether a given argument is credulously acceptable/deniable under semi-stable/semi-two-valued semantics of a given ADFs,
3. whether a given argument is skeptically acceptable/deniable under semi-stable/semi-two-valued semantics of a given ADFs.

# Part III

# Discussion Games

# Chapter 6

# A Discussion Game for the Grounded Semantics

The reasoning tasks that can be defined for the several semantics for ADFs have been presented (Dvořák and Dunne, 2018). Also dialectical methods have a critical role in evaluating arguments. In ADFs this role is not obvious via the definition of semantics. In this chapter, we focus on the grounded semantics of ADFs and provide the grounded discussion game. Because each ADF has a unique grounded interpretation, no one has any doubt on the truth value of arguments of the grounded interpretation, and the grounded interpretation is the least complete interpretation. Thus, it is reasonable to ask 'why is an argument justifiable under the grounded semantics of a given ADF?' We handle this issue by presenting a discussion game under grounded semantics of ADFs. We show that an argument is acceptable (deniable) in the grounded interpretation of an ADF if and only if the proponent of a claim has (respectively, does not have) a winning strategy in the grounded discussion game. Furthermore, we study the relation between grounded discussion games and strong admissibility semantics of ADFs, presented in Chapter 3.

## 6.1 Introduction

Argumentation has received increased attention within artificial intelligence, since the remarkable paper of Dung (1995), in which abstract argumentation frameworks (AFs) are presented. Abstract dialectical frameworks (ADFs) introduced in (Brewka and Woltran, 2010) are expressive generalizations of AFs in which the logical relations among arguments can be

represented. Applications of ADFs have been presented in legal reasoning (Al-Abdulkarim et al., 2016, 2014), online dialog systems (Neugebauer, 2017, 2019) , the instantiation of defeasible theories (Strass, 2014), and text exploration (Cabrio and Villata, 2016).

Although dialectical methods have a role in determining semantics of both AFs and ADFs, the roles are not immediately obvious from the definition of semantics. To cover this gap, quite a number of works have been presented to show that semantics of AFs can be interpreted in terms of structural discussion (Jakobovits and Vermeir, 1999; Prakken and Sartor, 1997; Modgil and Caminada, 2009; Caminada, 2018; Dung and Thang, 2007; van Eemeren et al., 2014). Furthermore, in (Booth et al., 2018) it is shown that the structural discussion method has been used in human-machine interaction.

Because of the special structure of ADFs, existing methods used to interpret semantics of AFs cannot be reused in ADFs. To address this problem, we have presented the first existing game for ADFs (Keshavarzi Zafarghandi et al., 2019a). That game characterizes the preferred semantics. In this work we focus on the grounded semantics of ADFs.

In ADFs, a key question is 'How is it possible to evaluate arguments in a given ADF ?' Answering this question leads to the introduction of several types of semantics, defined based on three-valued interpretations. Different semantics reflect different types of point of view about the acceptance or denial of arguments. In ADFs an interpretation is called *admissible* if it does not contain any unjustifiable information. Most of the semantics of ADFs are based on the concept of admissibility. An interpretation is *complete* if it exactly contains justifiable information. In addition, an interpretation is *grounded* if it collects all the information that is beyond any doubt. Each ADF contains the unique grounded interpretation that can be the trivial interpretation. Furthermore, in the hierarchy, grounded semantics have the lowest computational complexity (Strass and Wallner, 2015). However, by indicating whether an argument is credulously acceptable (deniable) in a given ADF under grounded semantics we have the answer of the skeptical decision problem of the argument in question under complete semantics.

In this chapter, we present a game that can answer both the credulous and the skeptical decision problem of a given ADF, called *grounded discussion game*. In (Polberg, 2017) it is shown that each ADF is equivalent with an ADF without any redundant links. Thus, without loss of generality, the current game is presented over the subclass of ADFs that do not have redundant links. This game works locally by considering those ancestors of

an argument in question that can affect the evaluation of the argument in the grounded interpretation. In this way, the grounded decision problem can be answered without constructing the full grounded interpretation. Furthermore, the current methodology can be used to answer the decision problems under grounded semantics of formalisms that can be represented as ADFs, such as AFs.

In Section 6.2, we present the *grounded discussion game* that can capture the notion of grounded semantics. Furthermore, in Section 6.3 we present soundness and completeness of the method. Finally, in Section 6.4 we study the relation between the grounded discussion games of ADFs, presented in this chapter, and strong admissibility semantics of ADFs, presented in Chapter 3.

## 6.2 Grounded Discussion Games

In this section we present a discussion game to answer the credulous (skeptical) decision problem under grounded semantics in a given ADF $F$ that does not have any redundant relation, without loss of generality, since any ADF has an equivalent of ADF of this kind; see (Polberg, 2017). Below we first present an informal definition of the grounded discussion game, however, if the reader prefers to start with the formal definition, they can skip the next part and start with Definition 6.2.

A grounded discussion game (GDG) is a dispute between a proponent (P) and an opponent (O). A GDG is started by a claim of P about the truth value of argument $a$ in the grounded interpretation of a given ADF. That is, P believes that the trivial interpretation $g_0 = v_{\mathbf{u}}$ can be extended to the grounded interpretation that contains the initial claim. O challenges P by asking whether $a$ is an initial argument. If P finds that $a$ is an initial argument and presents the truth value of $a$ to O, then O has to check whether this value is the same as the initial claim. In this case P wins if the checking of O leads to a positive answer. On the other hand, if P answers $a$ is not an initial argument, then O asks whether an ancestor of $a$ is an initial argument. If P finds that there is no initial argument in the ancestors of $a$, then the game is stopped and O wins the game.

However, if $a$ is not an initial argument but P finds that $b$ is an ancestor of $a$ which is also an initial argument, then P updates the information of $g_0$ with $g = g_0|_x^b$, such that $x$ is the truth value of $b$ in the grounded interpretation $F$. Furthermore, in this step a set of arguments in the shortest paths, between $a$ and $b$, are presented by P to O. Note that it

is possible that there exists more that one shortest path between two arguments. Actually, by presenting $g$, P says that $g$ can be extended to the grounded interpretation of $F$.

Now, O checks a piece of information presented in $g$ and the initial claim. If $g$ contains the initial claim, then the game halts and P wins the games. If the information of $g$ is in contradiction with the initial claim, then O wins the game. Since $a$ is not an initial argument, this checking step by O does not lead to acceptance or rejection of the initial claim. That is, presenting of $g$ by P did not convince O about the initial claim.

Thus, O asks P whether P can extend the information of $g$ to an interpretation that contains the initial argument. To this end, P evaluates the acceptance conditions of the children of the argument presented in $g$ under the information of $g$ and presents $g'$. Then, the game continues by O. If O indicates that $g'$ contains the initial claim, then the game stops. If $g$ and $g'$ contain the same piece of information, O asks P for a new initial ancestor of $a$. Otherwise, O asks P to extend $g'$ more.

The game continues between P and O alternately. P tries to extend the information of $g_0$ to an interpretation that contains the initial claim to support the belief. O tries to challenge P by either: 1. checking the information of the interpretation which is presented by P as an answer, or 2. asking whether the initial claim is an initial argument, or 3. requesting P to find an ancestor of $a$ which is an initial argument, or 4. requesting P to extend the information of the answer given by P to an interpretation that contains the initial claim.

In Example 6.1 we show how the game works before presenting the formal definitions and the algorithm.

**Example 6.1** *Let* $F = (\{a, b, c, d, e, f\}, \{\varphi_a : \bot, \varphi_b : \neg a \vee \neg e, \varphi_c : b \wedge f, \varphi_d : e \wedge \neg c, \varphi_e : \neg f, \varphi_f : \top\})$ *be a given ADF, depicted in Figure 6.1. We know that* $grd(F) = \mathbf{fttfft}$. *P claims that d is deniable with respect to the grounded interpretation of F. That is, by the initial claim P believes that* $d \mapsto \mathbf{f}$ *belongs to the grounded interpretation. In other words, the claim of P says that* $g_0 = v_{\mathbf{u}}$ *can be extended to the grounded interpretation that contains the initial claim.*

- *P says* $g_0 = v_{\mathbf{u}}$ *can be extended to the grounded interpretation of F that contains* $d \mapsto \mathbf{f}$.
- *O asks P whether d is an initial argument.*
- *P checks the acceptance condition of d and the answer is 'no, d is not*

*an initial argument'. Thus, the information of $g_0$ does not change. For technical reasons we let $g_1 = g_0$.*

- *O challenges P by asking whether any of the ancestors of d is an initial argument.*

- *P checks the acceptance conditions of the parents of d, namely c and e; neither of them is an initial argument. Then, P goes one step further and checks the parents of c and e, which are b and f. Here, f is an initial argument. Since P finds an ancestor of a which is an initial argument, P stops searching. By $\varphi_f : \top$, f is acceptable in the grounded interpretation of F. Thus, P presents interpretation $g_2 = g_1|^f_{\mathbf{t}} = \mathbf{uuuuut}$ and set $Ancestors(d, g_1) = \{d, e, c, f\}$, which contains the arguments on the shortest paths between the initial claim d and the initial argument f, that is presented in $g_2$ but not in $g_1$. P claims that $g_2$ can be extended to the grounded interpretation of F that contains the initial claim.*

- *Then O checks the information that is presented by $g_2$. Since $g_2$ does not contain any information about the initial claim, O asks P whether P can extend $g_2$.*

- *To this end, P evaluates the truth value of the children of f that are in $Ancestors(d, g_1)$ under $g_2$. The children of f that appear in that set are c and e. Thus, P evaluates $\varphi_c^{g_2} \equiv b \wedge \top \equiv b$ and $\varphi_e^{g_2} \equiv \bot$. That is, e is deniable with respect to the grounded interpretation of F. Thus, P presents $g_3 = g_2|^e_{\mathbf{f}} = \mathbf{uuuuft}$ to O as an extension of $g_2$ and P claims that $g_3$ can be extended to the grounded interpretation of F that contains the initial claim.*

- *O finds that $g_3$ extends the information of $g_2$ and it does not present any information in contrast with the initial claim. However, $g_3$ does not contain any information about the initial claim. Thus, O asks P whether P can extend $g_3$ to an interpretation that contains the initial claim.*

- *Again P evaluates the only child of e in set $Ancestors(a, g_1)$, namely d, under $g_3$. This attempts leads to $g_4 = \mathbf{uuufft}$.*

- *O checks the information given by $g_4$. Since $g_4$ contains the initial claim, the discussion between P and O halts here and P wins the game.*

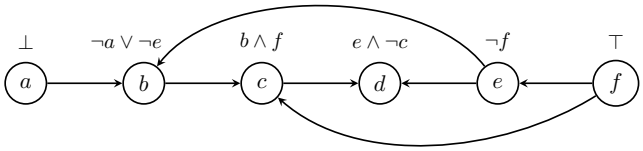


**Figure 6.1:** ADF of Examples 6.1

*Here, P does not present the grounded interpretation of F, however, P presents a constructive proof for the initial claim. That is, to indicate the initial claim, P works on the truth value of the argument in question locally. Thus, the grounded discussion game can answer the credulous decision problem under the grounded semantics of an ADF without indicating the truth value of all arguments in the grounded interpretation.*

**Definition 6.2** *Let $F = (A, R, C)$ be an ADF, let $a$ be an argument and let $S$ be a set of arguments. Function $Par(S)$ shows the set of parents of the elements of $S$; function $child(a)$ designates the set of children of $a$; and function $anc(a)$ presents the set all ancestors of $a$, defined formally in the following.*

- $Par(S) = \bigcup_{a \in S} par(a)$,
- $child(a) = \{b \mid (a, b) \in R\}$,
- *Let $m$ be a smallest integer such that: $Par^m(a) \subseteq \bigcup_{i=1}^{m-1} Par^i(a)$. Now we define: $anc(a) = \bigcup_{n=1}^{m} Par^n(a)$.*

Note that whenever $S$ contains only one argument $a$, $Par(S) = par(a)$ and we also write $Par(a)$ for $Par(\{a\})$. The aim of $anc(a)$ is to collect all ancestors of $a$ and condition $Par^m(a) \subseteq \bigcup_{i=1}^{m-1} Par^i(a)$ is a guarantee that the function does not go into a loop. If $b \in anc(a)$ is an initial argument, then we call it an *initial ancestor of $a$*.

The grounded discussion game is defined based on the following moves; some of them have an argument as a parameter and some of them are binary functions, defined on arguments and interpretations.

- $IniClaim(a)$: with this move P presents her/his beliefs about the truth value of $a$ in the grounded interpretation of $F$.
- $Ini(a)$: with this move O asks P whether $a$ is an initial argument.

- *CheckIni*($a$): with this move P checks whether $a$ is an initial argument.
- *Check*($g_{i-1}, g_i$): with this move O compares the information presented in $g_{i-1}$ and $g_i$.
- *IniAnc*($a, g$): with this move O asks P to present at least one initial ancestor of $a$ which is not presented in $g$, together with its truth value.
- *NewIniAnc*($a, g$): with this move P presents initial ancestors of $a$ which are requested by O in *IniAnc*($a, g$).
- *Ancestors*($a, g$): with this move P presents the set of arguments in the shortest paths between $a$ and the elements of *NewIniAnc*($a, g$).
- *Extend*($g$): with this move O requests P to extend the information of $g$.
- *Eval*($g$): with this move P evaluates the truth value of the children of the arguments presented in $g$ which appears in the last *Ancestors*($a, -$) under $g$.

In the game, P has the responsibility of constructing a proof for the initial claim. On the other hand, O aims to block the discussion by finding a contradiction or challenging P in such a way that P cannot answer the challenge.

- The game between P and O starts with *IniClaim*($a$) by which P presents a belief about the truth value of argument $a$ in the grounded interpretation of $F$. In this step, intuitively, P believes that $g_0 = v_{\mathbf{u}}$ can be extended to the grounded interpretation that contains the claim.
- Then, O applies *Ini*($a$), asks whether $a$ is an initial argument.
- Now, it is P's turn to apply *CheckIni*($a$) to check the acceptance condition of $a$. If $a$ is an initial argument, then the output of *CheckIni*($a$) is $g_1 = g_0|^a_{\mathbf{t/f}}$. Otherwise, $g_1 = g_0$.
- By *Check*($g_{i-1}, g_i$), O checks whether $g_{i-1} <_i g_i$ or $g_{i-1} = g_i$.
  - If $g_{i-1} <_i g_i$ and $g_i$ contains the initial claim, then the game stops.

- If $g_{i-1} <_i g_i$ but $g_i$ contains the negation of the initial claim, then the game stops.
- If $g_{i-1} <_i g_i$ and $g_i$ does not contain any information about the initial claim, then O requests P to extend $g_i$. That is, O applies $\mathit{Extend}(g_i)$.
- If $g_{i-1} = g_i$,
  - if $g_i$ is the output of either $\mathit{CheckIni}(a)$ or $\mathit{Eval}(g_{i-1})$, then O asks P to present a new initial ancestor of $a$. That is, O applies $\mathit{IniAnc}(a, g_{i-1})$,
  - if $g_i$ is the output of $\mathit{NewIniAnc}(a, g_{i-1})$, then the game stops.

- After move $\mathit{IniAnc}(a, g_i)$ by O, P applies $\mathit{NewIniAnc}(a, g_i)$ to find new initial ancestors of $a$. The output of this function is interpretation $g_{i+1}$ with $g_{i+1} = g_i|^b_{\mathbf{t/f}}$ such that $b$ is an initial ancestor of $a$, that was not presented in $g_i$. This function will be defined precisely in the following.

- Subsequently, after move $\mathit{IniAnc}(a, g_i)$ presented by O, P presents a set of arguments between the initial claim and the elements of $\mathit{NewIniAnc}(a, g_i)$, with the shortest distance, by applying $\mathit{Ancestors}(a, g_i)$. If there are more than one shortest path between the initial claim and an element of $\mathit{NewIniAnc}(a, g_i)$, then $\mathit{Ancestors}(a, g_i)$ presents the arguments of all paths with the shortest length.

- After move $\mathit{Extend}(g_i)$ presented by O, P applies $\mathit{Eval}(g_i)$. The output of this function is interpretation $g_{i+1}$ with $g_{i+1} = g_i|^b_{\varphi_b^{g_i}}$ such that $b$ is a child of an argument that is presented in $g_i$ which also appears in the last output of $\mathit{Ancestors}(a, -)$.

The only function that needs more explanation is $\mathit{NewIniAnc}(a, g)$, by which P tries to find the truth values of the initial ancestors of $a$ that are not presented in $g$. To this end, P uses the modification of the function *anc*, defined in Definition 6.2, which is called $\mathit{NewAnc}(a, g)$. This function is a binary function that takes the argument $a$ and interpretation $g$, and returns the set of ancestors of $a$. However, if there exists an initial ancestor of $a$, the truth value of which is not indicated in $g$, then the function stops. This is the reason why this function is called the new ancestors of $a$ with respect to $g$.
Let $m$ be a smallest integer such that:

- $Par^m(a) \subseteq Par^{m-1}(a)$; or
- $\exists p \in Par^m(a)$ such that $\varphi_p \equiv \top/\bot$, and $p$ was not presented in $g$.

Now we can define $NewAnc(a, g) = \bigcup_{n=1}^{m} Par^n(a)$.
Then among the elements of $NewAnc(a, g)$, P looks for the initial arguments. Function $NewIniAnc(a, g)$, presented in the following, takes $a$ and $g$, and updates $g$ by adding the truth values of the initial ancestors of $a$ that appear in $NewAnc(a, g)$.
$NewIniAnc(a, g) = g|^{b}_{\varphi_b^g}$ such that $b \in NewAnc(a, g)$ and $b$ is an initial argument.

**Definition 6.3** *Let $F = (A, R, C)$ be an ADF. A grounded discussion game for credulous acceptance (denial) of $a \in A$ is a sequence $[g_0, \ldots, g_n]$ such that the following conditions hold:*

- $g_0 = v_{\mathbf{u}}$;
- $g_1 = CheckIni(a)$;
- *for* $0 \leq i < n$, $g_i \leq_i g_{i+1}$;
- $g_n$ *contains either*
  - *the initial claim, or*
  - *the negation of the initial claim, or*
  - $g_{n-1}$ *is the output of* $NewIniAnc(a, g_{n-2})$ *and* $g_{n-1} = g_n$.
- *for* $1 < i < n$, *if* $g_{i-1} <_i g_i$ , *then* $g_{i+1}$ *is the output of* $Eval(g_i)$;
- *for* $0 < i < n$, *if* $g_{i-1} = g_i$, *then* $g_{i+1}$ *is the output of* $NewIniAnc(a, g_i)$.

**Definition 6.4** *Let $F$ be a given ADF. Let $[g_0, \ldots, g_n]$ be a grounded discussion game for credulous acceptance (denial) of an argument.*

- *P wins the game if $g_n$ satisfies the initial claim,*
- *O wins the game if $g_n$ satisfies the negation of the initial claim, or* $g_{n-1} =$ $NewIniAnc(a, g_{n-2})$ *and* $g_{n-1} = g_n$.

Example 6.5 is an instance of a game in which O wins.

**Example 6.5** *Let $F = (\{a, b, c\}, \{\varphi_a : \neg b, \varphi_b : \neg c, \varphi_c : \neg a\})$ be an ADF. We know that $grd(F) = v_{\mathbf{u}}$. P claims that b is acceptable in the grounded interpretation of F.*

- *$IniClaim(b) = g_0 = v_{\mathbf{u}}$: P believes that $g_0$ can be extended to the grounded interpretation of F in which b is acceptable.*
- *O asks $Ini(b)$.*
- *P applies $CheckIni(b)$ to answer the challenge. The output of $CheckIni(b)$ is $g_1 = g_0$.*
- *O applies $Check(g_0, g_1)$. Since $g_0 = g_1$ and $g_1$ is the output of $CheckIni(b)$, O requests $IniAnc(b, g_1)$.*
- *To answer $IniAnc(b, g_1)$, P applies $NewIniAnc(b, g_1)$. To this end, first P computes $NewAnc(b, g_1) = \{a, b, c\}$. Since none of them is an initial argument, then the output of $NewIniAnc(b, g_1)$ is $g_2 = g_1$.*
- *O applies $Check(g_1, g_2)$, which leads to $g_1 = g_2$. Since $g_2$ is an output of function $NewIniAnc(b, g_1)$, the game stops and by Definition 6.4, O wins the game.*

*That is, the initial claim of P that b is acceptable with respect to the grounded interpretation of F is false. This corresponds with the fact that the grounded interpretation $v_{\mathbf{u}}$ of F does not satisfy the belief of P.*

## 6.3 Soundness and Completeness

In this section we show that the presented method is sound and complete. To show the completeness, first we show that in an ADF without any redundant links, the grounded interpretation assigns the truth value of an argument to **t** or **f** if it is either an initial argument or its truth value is affected by the initial ancestors. This corollary is the direct result of Lemma 6.6.

**Lemma 6.6** *Let F be an ADF without any redundant link, that does not have any initial argument. Then the grounded interpretation of F is $v_{\mathbf{u}}$.*

**Proof** Toward a contradiction, assume that $F$ does not contain an initial argument and $grd(F) \neq v_{\mathbf{u}}$. Let $a$ be an arbitrary argument. We show that $\varphi_a^{v_{\mathbf{u}}}$ is neither irrefutable nor unsatisfiable. Since $F$ does not have any initial argument, $a$ has a parent.

- Consider that $a$ has a parent $b$ such that $(b, a)$ is a dependent link. By the definition of dependent link, there are two-valued interpretations $v$, $w$ such that $v(\varphi_a) = \mathbf{t}$ and $v|^b_{\mathbf{t}}(\varphi_a) \neq \mathbf{t}$, and $w(\varphi_a) = \mathbf{f}$ and $w|^b_{\mathbf{t}}(\varphi_a) \neq \mathbf{f}$. Thus, $v, w \in [v_{\mathbf{u}}]_2$ and $v(\varphi_a) \neq w(\varphi_a)$. Therefore, $\varphi_a^{v_{\mathbf{u}}}$ is neither irrefutable nor unsatisfiable.

- Consider that none of the parents of $a$ is dependent. Construct the two-valued interpretation $v$ in which 1. $b \mapsto \mathbf{f}$ if $(b, a)$ is an attacker, and 2. $b \mapsto \mathbf{t}$ if $(b, a)$ is a supporter. Construct the two-valued interpretation $w$ in which 1. $b \mapsto \mathbf{t}$ if $(b, a)$ is an attacker, and 2. $b \mapsto \mathbf{f}$ if $(b, a)$ is a supporter. That is, $v, w \in [v_{\mathbf{u}}]_2$. If either $a \notin \mathit{par}(a)$ or $(a, a)$ is a supporter, then $v(\varphi_a) \equiv \mathbf{f}$ and $w(\varphi_a) \equiv \mathbf{t}$. Thus, $\varphi_a^{v_{\mathbf{u}}}$ is neither irrefutable nor unsatisfiable. If $a \in \mathit{par}(a)$ and $(a, a)$ is an attacker, then $v(\varphi_a) = w(\varphi) = \mathbf{u}$. Thus, $\varphi_a^{v_{\mathbf{u}}}$ is neither irrefutable nor unsatisfiable.

Thus, the assumption that $a \mapsto \mathbf{t}/\mathbf{f} \in \mathit{grd}(F)$ is wrong. Hence, the unique grounded interpretation of $F$ is $v_{\mathbf{u}}$. □

Note that in Lemma 6.6, the assumption that $F$ does not have any redundant links is a necessary condition. Example 6.7 presents an instance of ADFs in which none of the arguments are initial arguments, however, the grounded interpretation of which is not the trivial interpretation.

**Example 6.7** *Let $F = (\{a, b\}, \{\varphi_a : b \vee \neg b, \varphi_b : b\})$ in which $(b, a)$ is a redundant link. $F$ does not have any initial argument. However, the grounded interpretation of $F$ is $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}\}$.*

**Lemma 6.8** *Let $F$ be an ADF without any redundant link, let $a$ be an argument that does not have any initial ancestor but has a parent, and let $v$ be the grounded interpretation of $F$. Then it holds that $v(a) = \mathbf{u}$.*

**Proof** Towards a contradiction, assume that $v(a) \neq \mathbf{u}$. Since $a$ is not an initial argument, it holds that $\mathit{par}(a) \neq \emptyset$. By a similar method as a proof of Lemma 6.6, we conclude that for any ancestors of $a$, namely $b$, i.e., $b \in \mathit{anc}(a)$, it holds that $\varphi_b^{v_{\mathbf{u}}}$ is neither irrefutable nor unsatisfiable. Thus, for any $b$ with $b \in \mathit{anc}(a)$, it holds that $v(b) = \mathbf{u}$. Hence, $v(a) = \mathbf{u}$. □ The following corollary is a direct results of Lemmas 6.6 and 6.8.

**Corollary 6.9** *Every argument that is acceptable (deniable) with respect to the grounded interpretation of $F$ either is an initial argument or has at least one initial ancestor.*

**Theorem 6.10** *(Soundness) Let $F$ be a given ADF. If there is a grounded discussion game for an initial claim of P in which P wins, then the grounded interpretation of $F$ satisfies the initial claim of P.*

**Proof of Soundness** Suppose that the initial claim of P is that '$a$ is acceptable (deniable) in the grounded interpretation'. Let $[g_0, \ldots, g_n]$ be a grounded discussion game for the initial claim of $P$, that is, $g_n$ satisfies the initial claim. We show that the grounded interpretation $v$ of $F$ satisfies the initial claim. By the definition the grounded interpretation of $F$ is the least fixed point of the characteristic operator. That is, there exists $m$ such that $\Gamma_F^m(v_{\mathbf{u}}) = v$. We show that $g_n \leq_i v$.

In the grounded discussion game if $n = 1$, that is $[g_0, g_1]$, then $a$ is an initial argument. Thus, clearly $g_1 \leq_i \Gamma_F(v_{\mathbf{u}})$. Since $\Gamma$ is a monotonic operator, $g_1 \leq_i v$. Consider that in the grounded discussion game $n > 1$. By induction on $n$ it is easy to show that for each $m$ with $0 \leq m \leq n$, $g_m \leq_i v$ holds.

Therefore, in the grounded discussion game $[g_0, \ldots, g_n]$ for any $i$ with $0 \leq i \leq n$, $g_i \leq_i v$ holds. In specific, $g_n \leq_i v$. Thus, the initial claim of P is satisfied in the grounded interpretation of $F$.

□

**Definition 6.11** *Let $F$ be an ADF. The distance from argument $a$ to $b$ in $F$ is the distance from $a$ to $b$ in the associated directed graph of $F$, denoted by $d(a, b)$. That is, $d(a, b)$ is the length of a shortest directed path from $a$ to $b$ in the directed graph associated to $F$.*

**Theorem 6.12** *(Completeness) Let $F$ be a given ADF without any redundant links. If $a$ is acceptable (deniable) in the grounded interpretation of $F$, then P wins the grounded discussion game for the initial claim of accepting (denying) of $a$.*

**Proof of Completeness** Let $F$ be an ADF and let $v$ be the grounded interpretation of $F$. Furthermore, let $a$ be an argument which is accepted (denied) with respect to $v$. Since $F$ does not have any redundant links, by Corollary 6.9, either $a$ is an initial argument or $a$ has at least one initial ancestor. We construct a grounded discussion game for the initial claim of $a \mapsto \mathbf{t/f}$ in which P wins. Let $g_0 = v_{\mathbf{u}}$. If $a$ is an initial argument, then $g_1 = g_0|^a_{\mathbf{t/f}}$. Thus, $[g_0, g_1]$ is the grounded discussion game, in which $g_1 = \mathit{CheckIni}(a)$, that satisfies the initial claim.

If $a$ is not an initial claim, then let $g_{1_1} = g_0$ and list the set of initial ancestors of $a$, for instance $L = [a_1, \ldots, a_k]$. Assume that $L$ is ordered based on the distance to $a$, increasingly. That is, $d(a_i, a) \leq d(a_{i+1}, a)$, for $i$ with $1 \leq i < k$. Let us categorize $L$ based on the distance of arguments to $a$. For instance, let $L_1 = \{a_1\} \cup B$ such that $B = \{a_i \mid d(a_i, a) = d(a_1, a)\}$. If $B \neq \{\}$, then let $m = |B|$, otherwise, $m = 1$. Let $L_2 = \{a_i \mid d(a_i, a) = d(a_{m+1}, a)\}$. Continue this process. Since $L$ is finite, there exists $p$ such that $L = \bigcup_{i=1}^{p} L_i$.

Let $g_{2_1} = g_{1_1}|_{v(b)}^{b}$ such that $b \in L_1$. For $j \geq 1$, for $i \geq 2$, 1. if $g_{i_j} > g_{i-1_j}$, then let $g_{i+1_j} = g_{i_j}|_{v(b)}^{b}$ such that $b$ is a child of an argument in $L_j$ that is on a path between $a$ and an element of $L_j$. 2. If $g_{i_j} = g_{i-1_j}$, then let $g_{i+1_j} = g_{i_j}|_{v(b)}^{b}$ such that $b \in L_{j+1}$. If any of the $g_{i_j}$ satisfies the initial claim, then stop the above loop.

Because the number of arguments on the paths between $a$ and elements of $L$ is finite, then the above sequence $[g_0, g_{1_1}, \ldots]$ will stop. Consider that the above loop halts in $g_{i_j}$. We claim that $D = [g_0, \ldots, g_{i_j}]$ is the GDG that satisfies the initial claim. To show that $D$ is a GDG it is enough to show that $D$ satisfies the fourth item of Definition 6.3, since all other items are trivial by the way of defining $D$. It is easy to check that $D$ satisfies the first four items of Definition 6.3. Thus, it is enough to show that $g_{i_j}$ satisfies the initial claim. Assume that $a \mapsto \mathbf{t} \in v$. We show that $a \mapsto \mathbf{t} \in g_{i_j}$. Toward a contradiction, assume that $a \mapsto \mathbf{t} \notin g_{i_j}$. That is, either $a \mapsto \mathbf{f} \in g_{i_j}$ or $a \mapsto \mathbf{u} \in g_{i_j}$. Since each element of $D$ is the update of the previous interpretation in $D$ by updating the truth value of a $b$ with $v(b)$, it is not possible that $a \mapsto \mathbf{f} \in g_{i_j}$. On the other hand, $a \mapsto \mathbf{u} \in g_{i_j}$ means that there is $b$ parent of $a$ the truth value of which has effect on the truth value of $a$ and $v(b) = \mathbf{u}$. By continuing this process, it holds that there is $c$ initial ancestor of $a$ that $v(c) = \mathbf{u}$. It is a contradiction that $v$ is the grounded interpretation of $F$. □

## 6.4 Grounded Discussion Games and Strong Admissibility

On the one hand, the grounded discussion game (GDG), presented in Section 6.2, answers the credulous (skeptical) decision problem of an ADF under grounded semantics without constructing the full grounded interpretation. On the other hand, a goal of presenting strong admissibility semantics of ADFs is to explain 'why a queried argument is justified in

the grounded interpretation'. In other words, since the concept of strong admissibility semantics of AFs relates to grounded semantics of AFs in a similar way as admissible semantics of AFs relates to preferred semantics of AFs, to answer the credulous decision problem under grounded semantics one can answer the query under strong admissibility semantics.

In this section we study the relation between interpretations in a grounded discussion game and strongly admissible interpretations of a given ADF. As background for the current section, one needs the primitive notions of strong admissibility semantics of ADFs, presented in Section 3.2 (in Chapter 3). Let $D$ be an ADF, and let $[g_0, \ldots, g_n]$ be a grounded discussion game used to investigate whether argument $a$ is credulously justified under grounded semantics of $D$.

- First we investigate whether each interpretation presented in the game, i.e., each $g_i$ for $i$ with $0 \leq i \leq n$, is a strongly admissible interpretation of $D$.
- Then, we study whether $g_n$ is a least witness of strong justifiability of $a$, as defined in Definition 3.5.

In Theorem 6.14 we show that each $g_i$ is a strongly admissible interpretation of $D$. However, there is no guarantee for $g_n$ being a least witness of strong justifiability of a queried argument. We investigate a counterexample in Example 6.16. The main result of this section is presented in Corollary 6.17 below.

**Proposition 6.13** *Let $D = (A, L, C)$ be an ADF, let $a$ be an initial argument of $A$, i.e., $\varphi_a \equiv \top/\bot$, let $v$ be an interpretation such that $v(a) = \mathbf{t}/\mathbf{f}$. Then $a$ is strongly justified in $v$.*

**Proof** Since $a$ is an initial argument, $\varphi_a^{v_\mathbf{u}}$ is irrefutable/unsatisfiable. Thus, by Definition 4.1, since $v(a) = \mathbf{t}/\mathbf{f}$, it holds that $a$ is strongly justified in $v$. □

**Theorem 6.14** *Let $D$ be an ADF and let $[g_0, \ldots, g_n]$ be a GDG for a credulous acceptance (denial) of $a \in A$. Each $g_i$ for $i$ with $0 \leq i \leq n$ is a strongly admissible interpretation of $D$.*

**Proof** We show the theorem by induction on $i$.

Base case: Let $i = 0$. By Definition 6.3, $g_0 = v_\mathbf{u}$ and by Lemma 3.19, the trivial interpretation is a strongly admissible interpretation. Thus, $g_0$ is a strongly admissible interpretation.

Induction hypothesis: Assume that for each $j$ with $0 \leq j < i < n$, it holds that $g_j$ is a strongly admissible interpretation of $D$.

Inductive step: We show that $g_i$ is also a strongly admissible interpretation. By Definition 6.3, either $g_i$ is the output of $Eval(g_{i-1})$ or it is the output of $NewIniAnc(a, g_{i-1})$. Further, by Definition 6.3, $g_{i-1} \leq_i g_i$, for all $i$ such that $0 < i \leq n$. If $g_{i-1} = g_i$, since by the induction hypothesis $g_{i-1}$ is a strongly admissible interpretation, then $g_i$ is also a strongly admissible interpretation. Thus, we assume that $g_{i-1} <_i g_i$. By Lemma 3.18, it follows that if $g_{i-1}(b) = \mathbf{t}/\mathbf{f}$, then $b$ is also strongly justified in $g_i$. We show that for each $b$ if $g_{i-1}(b) = \mathbf{u}$ and $g_i(b) \neq \mathbf{u}$, then $b$ is a strongly justified argument of $g_i$, thus $g_i$ is a strongly admissible interpretation of $D$.

- Assume that $g_i$ is the output of $NewIniAnc(a, g_{i-1})$. Let $b$ be an argument the truth value of which is presented in $g_i$ but not in $g_{i-1}$, i.e., $g_i(b) = \mathbf{t}/\mathbf{f}$ but $g_{i-1}(b) = \mathbf{u}$. Since $g_i = NewIniAnc(a, g_{i-1})$, by the definition of $NewIniAnc(a, g_{i-1})$, $b$ is an initial ancestor of $a$. Thus, by Proposition 6.13, $b$ is strongly justified in $g_i$.

- Assume that $g_i$ is the output of $Eval(g_{i-1})$. Let $b$ be an argument the truth value of which is presented in $g_i$ but not in $g_{i-1}$. By the definition of $Eval(-)$ function, $b \mapsto \mathbf{t}/\mathbf{f} \in g_i$ if $\varphi_b^{g_{i-1}} \equiv \top/\bot$. That is, there exists a subset of parents of $b$, namely $P$, such that the truth value of each $p \in P$ is presented in $g_{i-1}$. 1. Thus, $\varphi_b^{g_i|_P} \equiv \top/\bot$. 2. Furthermore, by induction hypothesis, each $p \in P$ is strongly justifiable in $g_{i-1}$.

  Thus, the conditions of Definition 4.1 are satisfied for $b$ in $g_i$. Hence, $b$ is a strongly acceptable/deniable argument in $g_i$.

Hence, $g_i$ is a strongly admissible interpretation of $D$. Thus, every interpretation $g_i$ in a GDG $[g_0, \ldots, g_n]$ is a strongly admissible interpretation of $D$. □

Theorem 6.14 implies that for each argument $a$ if $g_i(a) = \mathbf{t}/\mathbf{f}$, then $a$ is strongly justified in $g_i$. Corollary 6.15 is a direct result of Theorem 6.14. Since if $a$ in an initial claim of P and P wins in a GDG of $[g_0, \ldots, g_n]$, then the truth value of $a$ is presented in $g_n$.

**Corollary 6.15** *Let $D$ be an ADF, let 'a is credulously acceptable/deniable in the grounded interpretation of $D$' be an initial claim of P, and let $[g_0, \ldots, g_n]$ be a grounded discussion game in which P wins. Then, $a$ is strongly justified in $g_n$.*

Corollary 6.15 states that $a$ is strongly acceptable/deniable in $g_n$. For instance, in Example 6.1, for the initial claim of '$d$ is deniable with respect to the grounded interpretation of $F$', the GDG is $[g_0 = \mathbf{uuuuuu}, g_1 = \mathbf{uuuuuu}, g_2 = \mathbf{uuuuut}, g_3 = \mathbf{uuuuft}, g_4 = \mathbf{uuufft}]$ implies that 1. $d$ is deniable in the grounded interpretation of $D$, 2. furthermore, $d$ is strongly deniable in $g_4$, 3. moreover, $g_4$ is the least witness of strongly justifiability of $d$. This raises the question whether in any GDG $[g_0, \ldots, g_n]$, it holds that $g_n$ is a least witness of strong justifiability of a queried argument.

We show that there is no guarantee that $g_n$, for the game $[g_0, \ldots, g_n]$ in which P won, is always a least witness of strong justifiability of the queried argument, by presenting a counterexample in Example 6.16.

**Example 6.16** *Let $F = (\{a, b, c, d, e, f\}, \{\varphi_a : \bot, \varphi_b : \neg a \vee \neg e, \varphi_c : b \vee f, \varphi_d : e \wedge \neg c, \varphi_e : \neg f, \varphi_f : \top\})$ be an ADF. Proponent claims that $d$ is credulously deniable in the grounded interpretation of this ADF. The grounded discussion game that clarifies the claim is $[g_0 = \mathbf{uuuuuu}, g_1 = \mathbf{uuuuuu}, g_2 = \mathbf{uuuuut}, g_3 = \mathbf{uutuft}, g_4 = \mathbf{uutfft}]$, by which P wins. Here $d$ is strongly deniable in $g_4$. However, here $g_4$ is not a least witness of strong justfiability of $d$. Because here $g_4$ contains the truth values of an argument, namely $c$, which is not necessary to know to accept the truth value of $d$ in the grounded interpretation.*

*Although $c$ is a parent of $d$ and if we know that $c$ is accepted in the grounded interpretation, we concludes that $d$ is deniable in the grounded interpretation, there exists a strongly admissible interpretation with a fewer piece of information of ancestors of $d$, namely $v = \mathbf{uuufft}$. It is straightforward to check that $v$ is a strongly admissible interpretation of $D$ that has strictly less amount of information than $g_4$. Thus, to know the truth value of $d$ in the grounded interpretation of $F$ there is no need of any further information of the truth value of $c$.*

Example 6.16 shows that the GDG may produce some information about the truth values of some of the ancestors of the argument in question that are not necessary for answering the credulous decision problems under grounded semantics/strong admissibility semantics.

**Corollary 6.17** *Let $D$ be an ADF, let $[g_0, \ldots, g_n]$ be a GDG which satisfies the initial claim of proponent. It holds that $g_n$ is a strongly admissible interpretation of $D$ which satisfies the initial claim of P, however, $g_n$ may not be a least witness of strong justifiability for the queried argument.*

## 6.5 Conclusion

Grounded discussion games between two agents are presented in this work to answer the credulous decision problem of ADFs under grounded semantics (Keshavarzi Zafarghandi et al., 2020). Since each ADF is equivalent with an ADF without any redundant links, we present the game over this subclass of ADFs. A sub-goal of this game is presenting a constructive proof for the truth value of the argument in the grounded interpretation of a given ADF. To this end, there is no need of evaluating the whole grounded interpretation of a given ADF. Furthermore, in Section 6.3 we have shown that the method is sound and complete. In each move, P tries to show that the initial claim can be in an extension of the trivial interpretation, and O tries to challenge P by checking the content of the interpretation presented by P and either finding the initial claim or requesting P to extend the interpretation or find a new initial ancestor. Since the notion of strong admissibility semantics of ADFs presents a point of view of explaining why a queried argument is justified in the grounded interpretation (Keshavarzi Zafarghandi et al., 2021d,b), we have studied the relation between grounded discussion games and strong admissibility semantics of ADFs, in Section 6.4. As future work, it would be interesting to investigate a game for infinite ADFs and for ADFs the acceptance conditions of which are not restricted to propositional formulas.

# Chapter 7

# Discussion Games for Preferred Semantics

As presented in the previous chapter, the main query considered in this part of the thesis is: 'Why is an argument credulously justifiable under a type of semantics in a given ADF?' In the previous chapter, we answered the query by considering grounded semantics of ADFs and presenting grounded discussion games for ADFs, to clarify the dialectical character of grounded semantics for ADFs.

In this chapter, we focus on the preferred semantics of ADFs, and we provide a discussion game as a proof method to show the role of discussion in reasoning in ADFs under preferred semantics. Properties of the preferred semantics of ADFs include that a preferred interpretation represents maximum information about arguments without losing admissibility; that each admissible interpretation is contained in a preferred interpretation; and that the complexity of reasoning tasks under preferred semantics is the highest among the known semantics of ADFs in the polynomial hierarchy. We show that an argument is acceptable (deniable) by an ADF under preferred semantics if and only if there exists a discussion that can defend the acceptance (respectively, denial) of the argument in question. We show that our method is sound and complete.

## 7.1 Introduction

Abstract Dialectical frameworks (ADFs), first introduced in (Brewka and Woltran, 2010) and further refined in (Brewka et al., 2013, 2018a), are expressive generalizations of Dung's widely used argumentation frame-

works (AFs) (Dung, 1995). ADFs are formalisms that abstract away from the content of arguments but are expressive enough to model different types of relations among arguments. Applications of ADFs have been presented in legal reasoning (Al-Abdulkarim et al., 2016, 2014) and text exploration (Cabrio and Villata, 2016).

Basically, the term 'dialectical method' refers to a discussion among two or more people who have different points of view about a subject but are willing to find out the truth by argumentation. That is, in classical philosophy, dialectic is a method of reasoning based on arguments and counter-arguments (Krabbe, 2006; Macoubrie, 2003).

In ADFs, dialectical methods have a role in picking the truth-value of arguments under principles governed by several types of semantics, defined mainly based on three-valued interpretations, a form of labelings. Thus, in ADFs, beyond an argument being *acceptable* (corresponding to *defended* in AFs), there is a symmetric notion of being *deniable*. One of the most common argumentation semantics are the *admissible* semantics, which in ADFs come in the form of interpretations that do not contain unjustifiable information. The other semantics of ADFs fulfil the admissibility property. Maximal admissible interpretations are called *preferred* interpretations. Preferred semantics have a higher computational complexity than other semantics in ADFs (Strass and Wallner, 2015)[1]. That is, answering the decision problems of preferred semantics is more complicated than answering the same problems of other semantics in a given ADF. Therefore, having a structural discussion to investigate whether a decision problem is fulfilled under preferred semantics in a given ADF has a crucial importance.

There exists a number of works in which the relation between semantics of AFs and structural discussions are studied (Jakobovits and Vermeir, 1999; Prakken and Sartor, 1997; Caminada, 2018; Dung and Thang, 2007; Modgil and Caminada, 2009; van Eemeren et al., 2014). As far as we know, the relation between semantics of ADFs and dialectical methods in the sense of discussion among agents has not been studied yet (Barth and Krabbe, 1982).

In this chapter, we introduce the first existing discussion game for ADFs. We focus on preferred semantics and we show that for each argument that is credulously accepted (denied) under preferred semantics in a given ADF, there is a discussion game successfully defending the argument. Given the unique structure of ADFs, standard existing approaches known from the AFs setting could not be straightforwardly reused (Caminada et al., 2014;

[1]provided that the polynomial hierarchy does not collapse.

Cayrol et al., 2003; Vreeswijk and Prakken, 2000; Verheij, 2007). We thus propose a new approach based on interpretations that can be revised by evaluating the truth values of the parents of the argument in question. The current methodology can be reused in other formalisms that can be represented in ADFs, such as AFs (Vreeswijk and Prakken, 2000; Verheij, 2007).

In Section 7.2, we present the *preferred discussion game*, which is a game with perfect information, that can capture the notion of preferred semantics. We show that there exists a proof strategy for arguments that are credulously acceptable (deniable) under preferred semantics in a given ADF and vice versa. Furthermore, we show soundness and completeness of the method.

## 7.2 Discussion Game for Preferred Semantics

In this section, we present the structure of the discussion game for preferred semantics. The aim is to show that an argument is credulously accepted (resp. denied) under preferred semantics in an ADF iff there exists a winning dialogue in a preferred discussion game for the player who starts the game with the corresponding claim of acceptance (resp. denial). A *preferred discussion game*, which is similar to Socrates' form of reasoning (Walton and Krabbe, 1995; Caminada, 2008), is a (non-deterministic) game in which the proponent of an initial claim should establish its initial claim and subsequent commitments. The game starts with a claim by the proponent (P) about credulous acceptance (resp. denial) of an argument under preferred semantics in a given ADF. Then at each step in the dialogue, P makes new commitments in order to establish an interpretation that supports the initial claim. The game is continued as long as there is a new claim by P and there is no contradiction.

Since each preferred interpretation is an admissible interpretation, if we want to investigate whether an argument is credulously acceptable (resp. deniable) under preferred semantics, we study whether the argument is credulously acceptable (resp. deniable) under admissible semantics. The key advantage of the current method is that the credulous acceptance (resp. deniability) problem for preferred semantics in an ADF $F$ can be solved without enumeration of all admissible interpretations of $F$. In the following, Examples 7.1, 7.8, and 7.10 represent preferred discussion games, in which there are winning dialogues for P's belief.

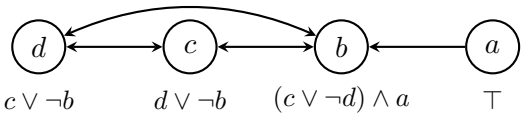


**Figure 7.1:** ADF of Example 7.1

**Example 7.1** *Given an ADF* $F = (\{a, b, c, d\}, \{\varphi_a : \top,\ \varphi_b : (c \vee \neg d) \wedge a,\ \varphi_c : d \vee \neg b,\ \varphi_d : c \vee \neg b\})$*, depicted in Figure 7.1.*

- *Assume that P claims that d is credulously acceptable under preferred semantics. The claim of P consists of information about the truth value of d, and there is no further information about the truth values of other arguments. This initial information of P is represented by the interpretation* $v_1 = \mathbf{uuut}$.

- *We check the consequences of P's claim. Based on the acceptance condition of d, argument d is acceptable in a preferred interpretation iff either c is accepted or b is denied in that interpretation. That is, the information of* $v_1$ *must be extended to one of two interpretations;* $v_2 = \mathbf{uutt}$ *and* $v_2' = \mathbf{ufut}$*, and P must answer the question, 'Does either b have to be assigned to* $\mathbf{f}$ *or c have to be assigned to* $\mathbf{t}$*, if d is assigned to* $\mathbf{t}$ *in a preferred interpretation?'*

- *Since there exist new commitments in both* $v_2$ *and* $v_2'$*, the dialogue must be continued on either one of them. P chooses to work on* $v_2$*, in which the only new challenged argument is c. P checks under which condition c can be accepted in a preferred interpretation. Based on the acceptance condition* $\varphi_c : d \vee \neg b$*, argument c is assigned to* $\mathbf{t}$ *if and only if either d is assigned to* $\mathbf{t}$ *or b is assigned to* $\mathbf{f}$*. That is, the new information of P about the truth values of arguments can be represented by* $v_3 = \mathbf{uutt}$ *and* $v_3' = \mathbf{uftt}$*. In the former one there is no new claim, that is, the dialogue* $v_1, v_2$ *and* $v_3$ *does not have to be continued anymore. P wins this dialogue, since P can defend the initial claim via this dialogue.*

Definitions 7.2–7.5 are needed to define a dialogue, which is presented in Definition 7.6.

**Definition 7.2** *Let k and v be interpretations such that* $k \leq_i v$*. An argument a is* ***recently presented*** *in interpretation v with respect to k*

| $\odot$ | **t** | **f** | **u** |
|---|---|---|---|
| **t** | **t** | **u** | **t** |
| **f** | **u** | **f** | **f** |
| **u** | **t** | **f** | **u** |

**Table 7.1:** Definition of operator $\odot$ on the set $\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$

*iff* $k(a) = \mathbf{u}$ *and* $v(a) \neq \mathbf{u}$. *Note that in the following,* $A_{k,v}$ *shows the set of all arguments that are recently presented in* $v$ *with respect to* $k$*, i.e.,* $A_{k,v} = k^{\mathbf{u}} \cap v^*$.

**Definition 7.3** *Let* $v$ *be an interpretation of an ADF* $F$*, and let* $a$ *be an argument such that* $v(a) \in \{\mathbf{t}, \mathbf{f}\}$*. An interpretation* $w_a^v$ *is called a* ***minimal interpretation around*** $a$ ***in*** $F$ ***with respect to*** $v$*, if* $\Gamma_F(w_a^v)(a) = v(a)$ *and there exists no* $w' <_i w_a^v$ *such that* $\Gamma_F(w')(a) = v(a)$*. The set of all minimal interpretations around* $a$ *in* $F$ *with respect to* $v$ *is denoted by* $W_a^v$.

Since the acceptance condition of each argument is indicated by a propositional formula, argument $a$ may have more than one minimal interpretation around $a$ in $F$. For instance, in Example 7.1, it is assumed that $d$ is credulously accepted, $v_1 = \mathbf{uuut}$. With respect to interpretation $v_0 = v_{\mathbf{u}} = \mathbf{uuuu}$, argument $d$ is recently presented in $v_1$, i.e., $A_{v_0,v_1} = \{d\}$. Based on the acceptance condition of $d$, namely $\varphi_d : c \vee \neg b$, interpretations $w_d^{v_1} = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}, d \mapsto \mathbf{u}\}$ and $w'^{v_1}_d = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}\}$ are minimal interpretations around $d$ in $F$. In Example 7.1, the set of all minimal interpretations around $d$ in $F$ with respect to $v_1$ is, $W_d^{v_1} = \{w_d^{v_1}, w'^{v_1}_d\}$.

We define an operator $\odot$ on the set $\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ of truth values, such that $\mathbf{u} \odot x = x \odot \mathbf{u} = x$ if $x \in \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$, $\mathbf{t} \odot \mathbf{t} = \mathbf{t}$ and $\mathbf{f} \odot \mathbf{f} = \mathbf{f}$, while $\mathbf{f} \odot \mathbf{t} = \mathbf{t} \odot \mathbf{f} = \mathbf{u}$, as depicted in Table 7.1.

We define the binary operator $\odot$ on interpretations pointwise as follows: $(v \odot w)(a) = v(a) \odot w(a)$ for all $a \in A$. Note that $\odot$ is associative, commutative and idempotent. The binary operator $\odot$ on two interpretations can be extended to finite sequences of interpretations. Let $v_1, \ldots v_n$ be interpretations, The definition of $\bigodot$ is given pointwise as follows, for each $a \in A$.

$$\bigodot_{i=1}^{n} v_i(a) = v_1(a) \odot \cdots \odot v_n(a)$$

**Definition 7.4** *Let $v$ and $k$ be interpretations such that $k \leq_i v$. For each $w \in \prod_{a \in A_{k,v}} W_a^v$, we define $\delta(v, w)$, which is called an* ***evaluation of the parents of arguments in*** $A_{k,v}$ ***with respect to*** $v$ ***and*** $w$*, as follows:*

$$\delta(v, w)(b) = v(b) \odot ( \bigodot_{a \in A_{k,v}} w_a(b))$$

*The set of all possible evaluations of the parents of arguments in $A_{k,v}$ is called* ***all evaluations of parents of*** $A_{k,v}$*, and denoted by $\delta_{A_{k,v}}(v)$ such that:*

$$\delta_{A_{k,v}}(v) = \{\delta(v, w) \mid w \in \prod_{a \in A_{k,v}} W_a^v\}$$

In the following, we use $\delta_{A_{k,v}}(v)$ to define the moves of the player. Note that we evaluate $\delta_{A_{k,v}}(v)$ if and only if $A_{k,v} \neq \emptyset$. We explain the reason why we stop a dialogue of the game in which $A_{k,v} = \emptyset$. Furthermore, if $A_{k,v}$ contains only one argument $a$, that is, $v = k|_{\mathbf{t}}^a$ or $v = k|_{\mathbf{f}}^a$, as it is presented in Definition 2.60, and $w = (w_a^v)$ where $w_a^v$ is a minimal interpretation around $a$ with respect to $v$, then we denote $\delta(v, w)$ by $\delta(v, w_a^v)$, and we denote the set of all evaluations of $A_{k,v}$ by $\delta_a(v)$.

In Example 7.1 presented above, the set of all minimal interpretations around $d$ in $F$ with respect to $v_1$ is $W_d^{v_1} = \{w_d^{v_1}, w'^{v_1}_d\}$, where $w_d^{v_1} = \{\mathbf{uutu}\}$ and $w'^{v_1}_d = \{\mathbf{ufuu}\}$. Thus, $w$ is either $w_d^{v_1}$ or $w'^{v_1}_d$. The evaluation of the parents of $d$ with respect to $v_1$ and $w = w_d^{v_1}$ is $\delta(v_1, w) = \mathbf{uutt}$, while with respect to $v_1$ and $w = w'^{v_1}_d$ the interpretation is $\delta(v_1, w) = \mathbf{ufut}$. Therefore, the set of evaluations of the parents of $d$ is $\delta_d(v_1) = \{\mathbf{uutt}, \mathbf{ufut}\}$.

The information of the player in a dialogue can be represented by an interpretation. In the first claim of P, there exists only information about the truth value ($\mathbf{t}$ or $\mathbf{f}$) of the argument that is claimed.

**Definition 7.5** *Let $a$ be an argument and let $v$ be an interpretation in which all arguments are assigned to* $\mathbf{u}$ *except for the argument $a$ which is assigned to either* $\mathbf{t}$ *or* $\mathbf{f}$*. Then $v$ is called an* initial claim*.*

**Definition 7.6** *A* dialogue *is a sequence of interpretations $[v_0, \ldots, v_n]$ such that the following hold:*

- *$v_0$ is the trivial interpretation in which all arguments are assigned to* $\mathbf{u}$*.*
- *$v_1$ is an initial claim (as in Definition 7.5).*

- *For* $i > 1$, $v_i \in \delta_{A_{v_{i-2},v_{i-1}}}(v_{i-1})$ *(as in Definition 7.4).*
- *For* $0 \leq i < n-1$, *it holds that* $v_i <_i v_{i+1}$.
- *It holds that either* $v_{n-1} = v_n$ *or* $v_{n-1} \not\leq_i v_n$.

Connecting the above definition of dialogue to the player, $v_1$ is an initial claim, presented by P, and each step in the dialogue is a move in which P makes new commitments aiming to establish an interpretation that supports the initial claim. Figure 7.2 shows the tree of all possible dialogues for Example 7.1 about the claim that argument $d$ is acceptable.

The last item in Definition 7.6 indicates the conditions under which a dialogue stops. If $v_{n-1} = v_n$, then $A_{v_{n-1},v_n} = \emptyset$. That is, the truth value of no argument is recently presented in $v_n$ and the truth values of all arguments in $v_{n-1}$ are exactly the same as $v_n$. Thus, there is nothing to discuss and the claim by P was successfully defended in the dialogue, and P wins. If $v_{n-1} \not\leq_i v_n$, this means that $v_{n-1}$ and $v_n$ do not agree on the truth value of at least one argument in $v^*_{n-1}$. We say that there is a *contradiction* in a dialogue $[v_0, \ldots, v_n]$ if the dialogue contains interpretations $v_{n-1}$ and $v_n$ such that $v_{n-1} \not\leq_i v_n$. In this case, the claim by P was not successfully defended by P in the dialogue, and P loses.

Actually, a *preferred discussion game* can be represented as a labeled rooted tree in which the root is labeled by the trivial interpretation, namely, $v_0 = v_{\mathbf{u}}$; the node in the first level of the tree is labeled by the initial claim, the interpretation $v_1$. The nodes of depth $i > 0$ are labeled one by one by all interpretations in $\delta(v, w)$, where $v$ is the label of the directly preceding node of the tree with depth $i-1$, and $w \in \prod_{a \in A_{k,v}} W^v_a$ in which $A_{k,v}$ is the set of all arguments that are recently presented in $v$ with respect to the label of the directly preceding node of $v$, namely, $k$. The game tree of Example 7.1, including a winning dialogue for P, is depicted in Figure 7.2. The dialogue $[v_0, v_1, \delta(v_1, w^{v_1}_d), \delta(v_2, w'^{v_2}_c), \delta(v_3, w^{v_3}_b)]$ in Fig. 7.2 leads to a contradiction, since $\delta(v_2, w'^{v_2}_c) \not\leq_i \delta(v_3, w^{v_3}_b)$.

**Definition 7.7** *Let* $[v_0, ..., v_n]$ *be a dialogue. Then we say that the dialogue is* won by the proponent *if* $v_{n-1} = v_n$. *The dialogue is* lost by the proponent *if* $v_{n-1} \not\leq_i v_n$.

We say that P wins the game iff P wins at least one dialogue of the game. P loses the game iff P loses all dialogues of the game. As we see in Figure 7.2, P wins the dialogue $[v_0, v_1 = \mathbf{uuut}, v_2 = \mathbf{uutt}, v_3 = \mathbf{uutt}]$, since $v_2 = v_3$. This dialogue shows that P can defend the initial claim. Thus, after this dialogue there is no need for continuing the game.

A preferred discussion game is here presented as a game tree. Our aim with the game is to establish that there is a winning dialogue for P if and only if the initial claim is satisfied by a preferred interpretation.

Example 7.8 is another instance of a preferred discussion game in which P wins the game, since there is a winning dialogue for the initial claim.

**Example 7.8** *Let $F$ be the ADF given in Example 7.1, i.e.,* $F = (\{a, b, c, d\}, \{\varphi_a : \top,\ \varphi_b : (c \vee \neg d) \wedge a,\ \varphi_c : d \vee \neg b,\ \varphi_d : c \vee \neg b\})$, *depicted in Figure 7.3.*

- *P claims that $d$ is credulously deniable in a preferred interpretation in $F$:* $v_1 = \mathbf{uuuf}$*; here,* $A_{v_0,v_1} = \{d\}$.
- *Applying* $\delta(-,-)$ *on* $v_1$ *and* $w_d^{v_1}$ *leads to* $v_2 = \delta(v_1, w_d^{v_1}) = \mathbf{utff}$.
- *The recently presented arguments are $b$ and $c$, i.e.,* $A_{v_1,v_2} = \{b, c\}$. *The minimal interpretations around $b$ in $F$ with respect to $v_2$ are* $w_b^{v_2} = \{a \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$ *and* $w'^{v_2}_b = \{a \mapsto \mathbf{t}, d \mapsto \mathbf{f}\}$, *and the minimal interpretation around $c$ in $F$ with respect to $v_2$ is* $w_c^{v_2} = \{b \mapsto \mathbf{t}, d \mapsto \mathbf{f}\}$. *Thus,* $v_3 = \delta(v_2, W_{bc}) = \mathbf{ttuf}$ *and* $v_3' = \delta(v_2, W'_{bc}) = \mathbf{ttff}$.
- *In dialogue* $[v_0, v_1, v_2, v_3]$, *since* $v_2 \not\leq_i v_3$, *P loses the dialogue. That is, P cannot defend the initial claim in this dialogue.*
- *However, in dialogue* $[v_0, v_1, v_2, v_3']$ *applying* $\delta(-,-)$ *on* $v_3'$ *and* $w_a^{v_3'}$, *it holds that* $v_4 = \delta(v_3', w_a^{v_3'}) = v_3'$. *Then, dialogue* $[v_0, v_1, v_2, v_3', v_4]$ *is a winning dialogue for P.*

*In dialogue* $[v_0, v_1, v_2, v_3', v_4]$, *P gradually constructs an interpretation making further commitments that support the initial claim, sometimes making a choice between several possibilities. Here P is successful, since no contradiction is encountered at the final dialogue step.*

The ADF of Example 7.1 can also be used as an example in which P loses the game, because P loses all dialogues that start with a certain claim by P; we explain this in Example 7.9.

**Example 7.9** *Given ADF $F$ of Example 7.1, i.e.,* $F = (\{a, b, c, d\}, \{\varphi_a : \top,\ \varphi_b : (c \vee \neg d) \wedge a,\ \varphi_c : d \vee \neg b,\ \varphi_d : c \vee \neg b\})$, *depicted in Figure 7.4.*

- *P claims that $b$ can be denied in a preferred interpretation in $F$,* $v_1 = \mathbf{ufuu}$.

- *There are three different dialogues based on this initial claim;*

  1. $[v_{\mathbf{u}}, v_1 = \mathbf{ufuu}, v_2 = \mathbf{ufft}, v_3 = \mathbf{uufu}]$,
  2. $[v_{\mathbf{u}}, v_1 = \mathbf{ufuu}, v_2 = \mathbf{ufft}, v'_3 = \mathbf{uuuu}]$,
  3. $[v_{\mathbf{u}}, v_1 = \mathbf{ufuu}, v'_2 = \mathbf{ffuu}, v"_3 = \mathbf{ufuu}]$.

*Each of the dialogues of this game ends in a contradiction. That is, in each dialogue P is unsuccessful, and P cannot defend the initial claim. Thus, P loses the game starting from this claim.*

Example 7.10 shows that P may have more than one winning dialogue for an initial claim. Actually, whenever there exists more than one preferred interpretation that satisfies the truth value of the initial claim, P has more than one winning dialogue.

**Example 7.10** *Given ADF $F$ of Example 7.1, i.e., $F = (\{a, b, c, d\}, \{\varphi_a : \top,\ \varphi_b : (c \vee \neg d) \wedge a,\ \varphi_c : d \vee \neg b,\ \varphi_d : c \vee \neg b\})$, a part of the game tree presents two winning dialogues of P, depicted in Figure 7.5.*

- *P claims that $b$ can be accepted in a preferred interpretation in $F$: $v_1 = \mathbf{utuu}$.*

- *There are two different winning dialogues based on this initial claim for P:*

  1. $[v_0, v_1, v_2 = \mathbf{tttu}, v_3 = \mathbf{tttt}, v_4 = \mathbf{tttt}]$,
  2. $[v_0, v_1, v'_2 = \mathbf{ttuf}, v'_3 = \mathbf{ttff}, v'_4 = \mathbf{ttff}]$.

*After P presents the initial claim $v_1$, there are two possibilities since $\delta_b(v_0, v_1) = \{v_2, v'_2\}$. P can choose to extend $v_2$ or $v'_2$ to an admissible interpretation. Here both choices are successful, and each leads to a dialogue won by P. That is, in this game, P has winning dialogues corresponding to these two admissible interpretations.*

In Example 7.10, it holds that $prf(F) = \{\mathbf{tttt}, \mathbf{ttff}\}$, that is, $b$ is skeptically acceptable under preferred semantics of $F$. As we see in this example, there is more than one winning dialogue for P. In Theorem 7.13, we show that if an argument is credulously acceptable (deniable), as it is presented in Definition 2.76, under preferred semantics in $F$, then there is a winning dialogue for P with the initial claim of accepting (denying) of the given argument. The examples above illustrate that P only has to consider the arguments that have been recently presented in the directly preceding move.

Let $F$ be an ADF and let $[v_0, \ldots, v_n]$ be a dialogue of a preferred discussion game of an initial claim of $F$. The *length* of the dialogue is the length of the sequence $[v_0, \ldots, v_n]$, namely, the number of elements of the sequence, in this case $n + 1$.

**Proposition 7.11** *Let $F = (A, L, C)$ be an ADF and let $|A| = n$. Then the length of each dialogue in a preferred discussion game of $F$ is at most $n + 2$.*

**Proof** Remember that for every $i < n$, the dialogue $[v_0, \ldots, v_i]$ is continued in $v_i$ iff $v_{i-1} <_i v_i$. Checking whether $v_{i-1} <_i v_i$ can be done by indicating the truth value of an argument in $v_i$ that is not indicated before, i.e., in $v_{i-1}$, i.e., $A_{v_{i-1},v_i}$. Since the number of arguments of $F$ is $n$ and $v_0 = v_{\mathbf{u}}$, the longest dialogue contains interpretations such that $v_0 < \cdots < v_n$, and in the next step, the parents of arguments of $A_{v_{n-1},v_n}$ will be evaluated. That is, the longest dialogue can be a sequence of $n + 2$ interpretations. Thus, the length of each dialogue cannot be more that $n + 2$. □

Since we assumed in Remark 2.40.1 that each ADF is finite, the immediate result of Proposition 7.11 is that a dialogue is finite.

**Theorem 7.12 (Soundness)** *Let an ADF $F = (A, L, C)$ be given. If there exists a winning dialogue for P in a preferred discussion game with initial claim of accepting (denying) an argument $a$, then $a$ is credulously acceptable (deniable) under preferred semantics in $F$.*

**Proof** Assume that there is winning dialogue $[v_0, \ldots, v_m]$ for P in a preferred discussion game, for accepting (denying) of an argument $a$. To show soundness, it is enough to show that $v_m$ is an admissible interpretation. Towards a contradiction, assume that $v_m$ is not an admissible interpretation, that is, $v_m \not\leq_i \Gamma_F(v_m)$. Thus, there exists an argument $b$ such that $v_m(b) \in \{\mathbf{t}, \mathbf{f}\}$, however, the valuation of the acceptance condition of $b$ under $v_m$ is not the same as $v_m(b)$; we prove the case that $v_m(b) = \mathbf{t}$. The proof method for the case in which $v_m(b) = \mathbf{f}$ is analogous.

Assume that $v_m(b) = \mathbf{t}$, but $\Gamma_F(v_m)(b) \in \{\mathbf{f}, \mathbf{u}\}$. $v_m(b) = \mathbf{t}$ means that there exists $i$ with $0 \leq i < m$ such that $v_i(b) = \mathbf{t}$ and $v_{i-1}(b) = \mathbf{u}$, i.e., $b \in A_{v_{i-1},v_i}$. Furthermore, $v_{i+1}$ contains the truth values of some of the parents of $b$, where $v_{i+1} \in \delta_{A_{v_{i-1},v_i}}$, such that $\Gamma_F(v_{i+1})(b) = \mathbf{t}$. Since P wins this dialogue, $v_{m-1} = v_m$. That is, $\varphi_b^{v_m} \equiv \top$, since $v_m$ contains the

truth values of $par(b)$ presented in $v_{i+1}$. Thus, $\Gamma_F(v_m)(b) = \mathbf{t}$. Therefore, the assumption that $v_m$ is not an admissible interpretation is rejected. □

In the following, the update of an interpretation $v$ with a truth value $x \in \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ for an argument $b$, as it is presented in Definition 2.60, is denoted by $v|_x^b$, where,

$$v|_x^b(a) = \begin{cases} x & \text{for } a = b, \\ v(a) & \text{for } a \neq b. \end{cases}$$

In addition, $v|_P$ is equal to $v(p)$ for any $p \in P$; however, it assigns all other arguments that do not belong to $P$ to $\mathbf{u}$, i.e., $v|_P = v_{\mathbf{u}}|_{v(p)}^{p \in P}$.

**Theorem 7.13 (Completeness)** *Let an ADF $F = (A, L, C)$ be given. If an argument $a$ is credulously acceptable (resp. deniable) under preferred semantics in $F$, then there is a winning dialogue for P in the preferred discussion game with the initial claim of accepting (resp. denying) of $a$.*

**Proof** Assume that an argument $a$ is credulously accepted under preferred semantics in $F$ (the proof method in case $a$ is credulously denied is analogous). Then there is a preferred interpretation $v$ of $F$ in which $a$ is accepted. We show that there exists a winning dialogue $[v_0, \ldots, v_n]$ for P in the preferred discussion game that is based on $v$. That is, we show that based on $v$ we can construct a dialogue $D = [v_0, \ldots, v_n]$ that satisfies the five conditions of Definition 7.6.

To construct the dialogue $D = [v_0, \ldots, v_n]$, we construct $v_i$ for $i$ with $0 \leq i \leq n$ as follows.

Let $v_0 = v_{\mathbf{u}}$. Let $v_1$, the initial claim, be an interpretation in which $a$ is assigned to $\mathbf{t}$ and all other arguments of $A$ are assigned to $\mathbf{u}$. Let $D = [v_0, v_1]$, and let $A_{v_0,v_1} = v_0^{\mathbf{u}} \cap v_1^*$, as it is presented in Definition 7.2. Note that for $i \geq 1$, we extend dialogue $D = [v_0, \ldots, v_i]$ if $v_{i-1} < v_i$. Since $v_0 <_i v_1$, we continue to extend dialogue $D$ by constructing $v_2$. Based on the method of constructing $v_{i+1}$, presented in the following, it holds that $v_i \leq v_{i+1}$. We stop to extend $D$ if $v_{i-1} = v_i$.

We construct $v_{i+1}$ based on $v_i$ and $v_{i-1}$, when $v_{i-1} <_i v_i$, for $i$ with $1 \leq i$, to this end first we construct $t_{i+1}$ as follows:

$$t_{i+1}(b) = \begin{cases} v(b) & b \in par(a), \text{ where } a \in A_{v_{i-1},v_i}, \\ v(b) & b \in v_i^*, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

Note that by the way of constructing of $t_{i+1}$, it holds that for each $b \in A$, if $b \in v_i^*$, then $t_{i+1}(b) = v_i(b) = v(b)$. Thus, it holds that $v_i \leq_i t_{i+1}$. We evaluate $\delta_{A_{v_{i-1},v_i}}(v_i)$ and we check whether $t_{i+1} \in \delta_{A_{v_{i-1},v_i}}(v_i)$.

- If $t_{i+1} \in \delta_{A_{v_{i-1},v_i}}(v_i)$, then let $v_{i+1} := t_{i+1}$ and extend $D$ to $D = [v_0, \ldots, v_{i+1}]$. Since $v_{i+1} \in \delta_{A_{v_{i-1},v_i}}(v_i)$, and $v_i \leq_i v_{i+1}$, by Definition 7.6, $D$ is a dialogue. If $v_i = v_{i+1}$, then we stop to extend $D$.

- If $t_{i+1} \notin \delta_{A_{v_{i-1},v_i}}(v_i)$: pick an element of $\delta_{A_{v_{i-1},v_i}}(v_i)$, namely $k$, such that $v_i \leq_i k < t_{i+1}$ and let $v_{i+1} = k$ and let $D = [v_0, \ldots, v_{i+1}]$. If $v_i = v_{i+1}$, then we stop to extend $D$. Since $v_{i+1} \in \delta_{A_{v_{i-1},v_i}}(v_i)$, and $v_i \leq_i v_{i+1}$, by Definition 7.6, $D$ is a dialogue. We present this step formally in the following, that is, we show that there exists a $k$ such that $k \in \delta_{A_{v_{i-1},v_i}}(v_i)$ and $v_i \leq_i k < t_{i+1}$. We show how we choose such a $k$.

  let $K = \{k \mid k \in \delta_{A_{v_{i-1},v_i}}(v_i),\ v_i \leq_i k <_i t_{i+1}\}$. First we show that $K \neq \emptyset$.

  Since $v \in \mathit{prf}(F)$, for each $a \in A_{v_{i-1},v_i}$ there exists a subset $P_a \subseteq \mathit{par}(a)$ such that $v|_{P_a} \in W_a^{v_i}$. For each $a \in A_{v_{i-1},v_i}$, let $w_a = v|_{P_a}$. Let $\delta(v_i, w) = v_i \odot (\bigodot_{a \in A_{v_{i-1},v_i}} w_a)$, where $w_a = v|_{P_a}$. By the construction of $\delta(v_i, w)$ and by Definition 7.4, it holds that $\delta(v_i, w) \in \delta_{A_{v_{i-1},v_i}}(v_i)$. We show that $v_i \leq_i \delta(v_i, w) <_i t_{i+1}$.

  First, we show that $\delta(v_i, w) <_i t_{i+1}$. For each $a \in A_{v_{i-1},v_i}$, $P_a \subseteq \mathit{par}(a)$ and $\mathit{par}(a) \subseteq t_{i+1}^*$. Thus by the definition of $\delta(v_i, w)$, for each $a \in A_{v_{i-1},v_i}$, it holds that $w_a \leq_i t_{i+1}$. Thus, $\delta(v_i, w) \leq_i t_{i+1}$. If $\delta(v_i, w) = t_{i+1}$, then this is a contradiction by the assumption that $t_{i+1} \notin \delta_{A_{v_{i-1},v_i}}(v_i)$. Thus, $\delta(v_i, w) <_i t_{i+1}$.

  Now, we show that $v_i \leq_i \delta(v_i, w)$. For $i \geq 1$, for each $b \in A$, if $v_i(b) \in \{\mathbf{t}, \mathbf{f}\}$, then $v_i(b) = v(b)$. For each $a \in A_{v_{i-1},v_i}$, for each $b \in A$, if $b \in \mathit{par}(a) \cap P_a$, then $w_a(b) = v(b)$. For each $a \in A_{v_{i-1},v_i}$, for each $b \in A$, if $b \notin \mathit{par}(a) \cap P_a$, then $w_a(b) = \mathbf{u}$. Thus, for each $b$ if $v_i(b) \in \{\mathbf{t}, \mathbf{f}\}$, then for each $a \in A_{v_{i-1},v_i}$, it holds that $w_a(b) \leq_i v_i(b)$. That is, for each $b \in A$, if $v_i(b) \in \{\mathbf{t}, \mathbf{f}\}$, then $v_i(b) = \delta(v_i, w)(b)$. If there exits $a \in A_{v_{i-1},v_i}$ and $b \in A$ such that $w_a(b) \in \{\mathbf{t}, \mathbf{f}\}$ and $v_i(b) = \mathbf{u}$, then $\delta(v_i, w)(b) \in \{\mathbf{t}, \mathbf{f}\}$. This is because $v$ is a preferred interpretation, thus, there exist no $a, a', \in A_{v_{i-1},v_i}$ and $b \in A$ such that $w_a(b) = \mathbf{t}$ and $w_{a'}(b) = \mathbf{f}$. Hence, $v_i \leq_i \delta(v_i, w)$.

Since $\delta(v_i, w) \in \delta_{A_{v_{i-1},v_i}}(v_i)$ and $v_i \leq_i \delta(v_i, w) <_i v_{i+1}$, it holds that $\delta(v_i, w) \in K$, i.e., $K \neq \emptyset$. Note that $K$ may contain more than one interpretation. For each $k \in K$, let $v_{i+1} = k$, and let $D = [v_0, \ldots, v_{i+1}]$. That is, if $|K| > 1$, then we would have more than one sequence of interpretations, namely dialogues for the initial claim.

By the construction of $D$, it holds that $v_i \leq_i v_{i+1}$, for $i$ with $0 \leq i$. Since the number of arguments is finite, this procedure will stop. That is, there exists a $D = [v_0, \ldots, v_n]$ such that $v_i <_i v_{i+1}$ for $i$ with $0 \leq i < n$, $v_{n-1} = v_n$, and $v_i \in \delta_{A_{v_{i-2},v_{i-1}}}(v_{i-1})$ for $i \geq 2$. Thus, by Definition 7.7, $D$ is a winning dialogue for P.

□

We illustrate the above completeness proof by an example. Consider again Example 7.1, i.e., $F = (\{a, b, c, d\}, \{\varphi_a : \top,\ \varphi_b : (c \vee \neg d) \wedge a,\ \varphi_c : d \vee \neg b,\ \varphi_d : c \vee \neg b\})$, shown in Figure 7.1. Argument $d$ is credulously acceptable in $F$ under preferred semantics. Furthermore, $v = \mathbf{tttt}$ is a preferred interpretation of $F$ in which $d$ is accepted. We follow the method presented in the proof of Theorem 7.13 to construct a winning dialogue for P.

Let $v_0 = v_{\mathbf{u}}$, let $v_1 = \mathbf{uuut}$ and let $D = [v_0, v_1]$. Since $A_{v_0,v_1} = \{d\}$ and $par(d) = \{b, c\}$, by the method of constructing $t_2$, presented in Theorem 7.13, it holds that $t_2 = \mathbf{uttt}$. We check whether $t_2 \in \delta_{A_{v_0,v_1}}(v_1)$. As we see, $t_2 \notin \delta_{A_{v_0,v_1}}(v_1) = \{\mathbf{uutt}, \mathbf{ufut}\}$. Let $P_d = \{c\} \subseteq par(d)$, as we see $w_d = v|_{P_d} = \mathbf{uutu} \in W_d^{v_1}$. It holds that $\delta(v_1, w_d) = \mathbf{uutt}$ and $K = \{k \mid k \in \delta_{A_{v_0,v_1}}(v_1), v_1 \leq_i k <_i t_2\} = \{\mathbf{uutt}\}$. Now, let $v_2 = \mathbf{uutt}$ and let $D = [v_0, \mathbf{uuut}, \mathbf{uutt}]$. In this step, $A_{v_1,v_2} = \{c\}$, $par(c) = \{d, b\}$. Thus, $t_3 = \mathbf{uttt}$. We check whether $t_3 \in \delta_{A_{v_1,v_2}} = \{\mathbf{uutt}, \mathbf{uftt}\}$. Since $t_3 \notin \delta_{A_{v_1,v_2}}$, we construct $K = \{k \mid v_1 \leq_i k <_i t_3,\ k \in \delta_{A_{v_1,v_2}}(v_1)\} = \{\mathbf{uutt}\}$. Hence, we pick $v_3 = \mathbf{uutt}$ and we let $D = [v_0, v_1, v_2, v_3]$. Since $v_2 = v_3$, it holds that $D$ is a winning dialogue for the initial claim of P.

## 7.3 Conclusion

In this chapter, preferred discussion games have been considered as a proof method to investigate credulous acceptance (denial) of arguments in an ADF under preferred semantics. Notable results are as follows:

1. The method is sound and complete.

2. Winning one dialogue of the game by P is sufficient to show that there exists a preferred interpretation in which the argument in question is assigned to the truth value which is claimed. Similar proposals for AFs have been studied by (Modgil and Caminada, 2009; Thang et al., 2009; Verheij, 2007).
3. In each move, the proponent has to study the truth value of arguments that were presented in the directly preceding move. In contrast, in (Caminada, 2018), O has to check all past moves of P to find a contradiction.
4. To investigate the credulous decision problem of ADFs under preferred semantics, there is no need to enumerate all preferred interpretations of an ADF.
5. In (Diller et al., 2018) it is shown that in the class of acyclic ADFs, all semantics coincide. Thus, in acyclic ADFs, the presented game can be used to decide the credulous problem for other types of semantics.

As future work, we could investigate structural discussion games for other semantics of ADFs. In addition, we could study discussion games for other decision problems of ADFs. Furthermore, we could investigate whether the presented method is more effective than the methods used in current ADF-solvers, for example (Brewka et al., 2017a; Ellmauthaler and Strass, 2014). This study may lead to new ADF-solvers that work locally on an argument to answer decision problems.

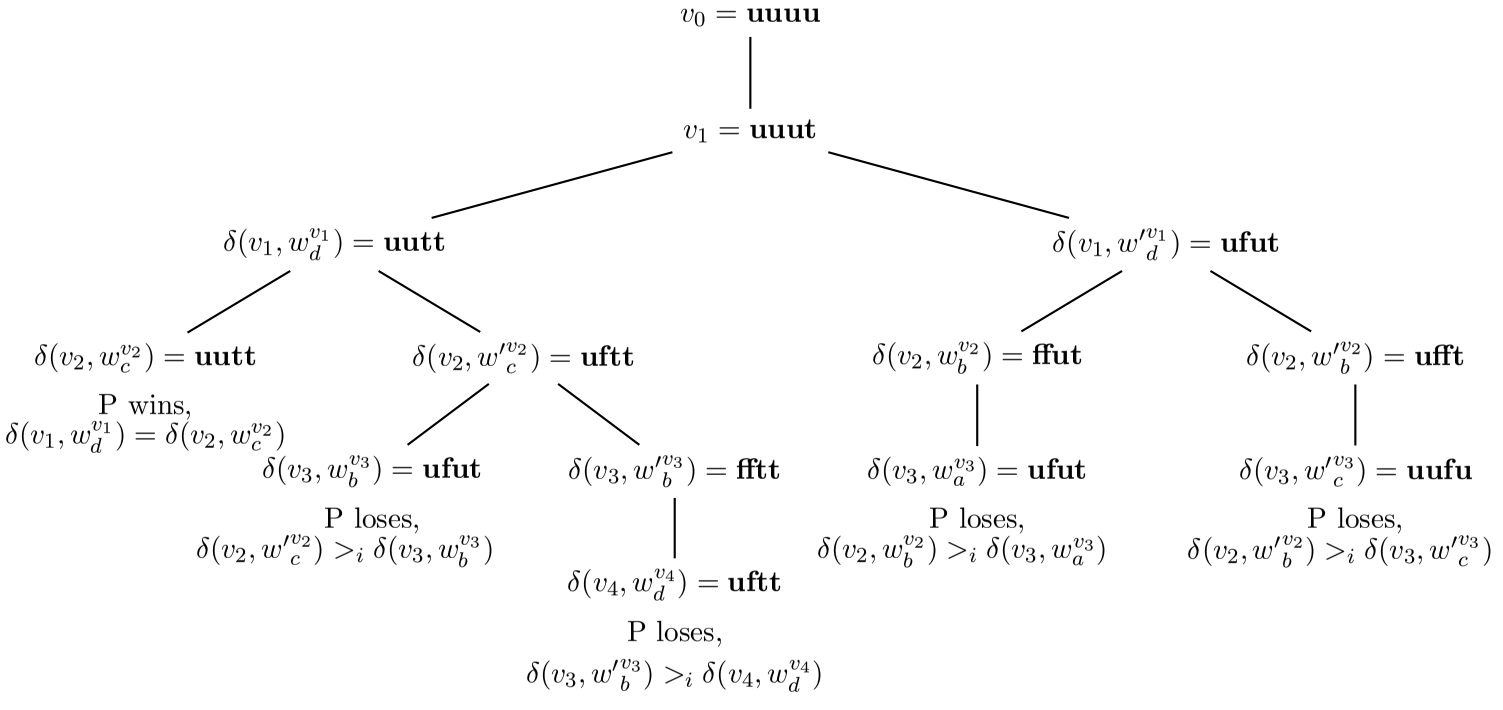


**Figure 7.2:** Associated game tree of the game in Example 7.1

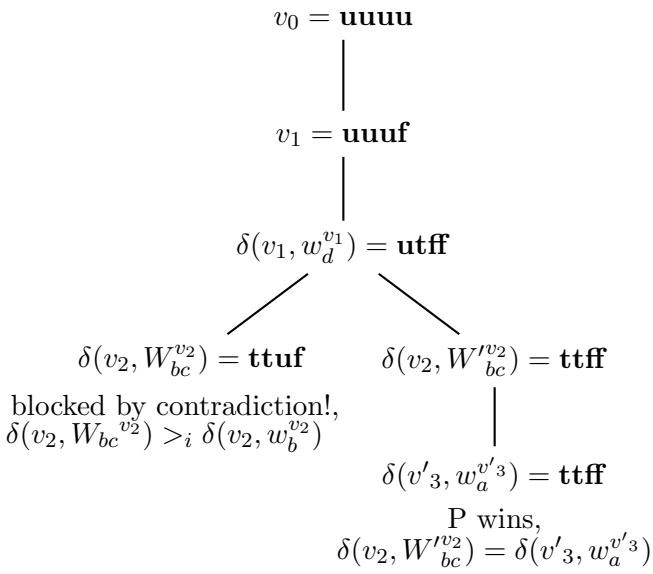


**Figure 7.3:** Associated tree of the game in Example 7.8

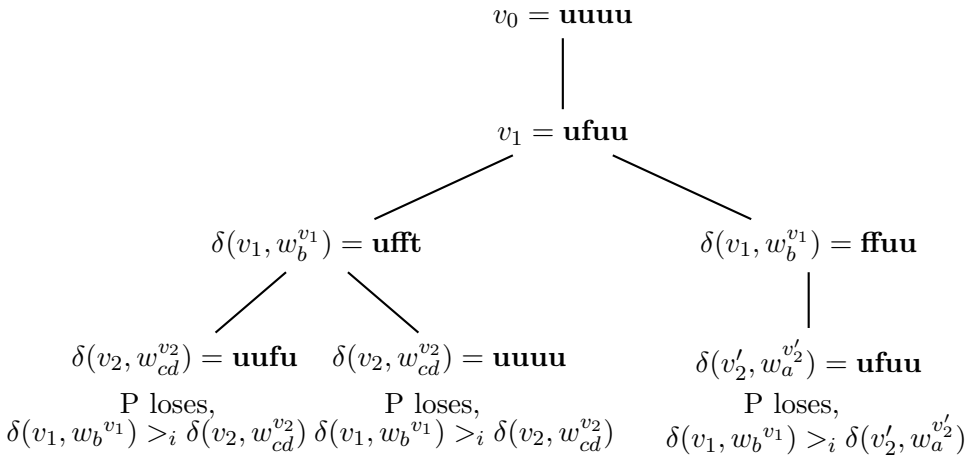


**Figure 7.4:** Associated tree of the game in Example 7.9

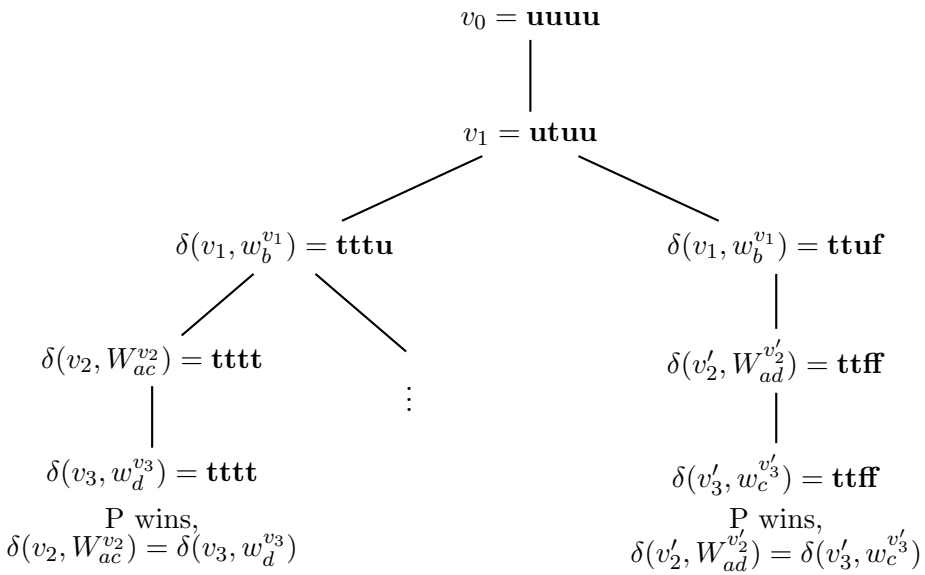


**Figure 7.5:** Associated tree of the game in Example 7.10

# Part IV

# Variations

# Chapter 8

# Investigating Subclasses of ADFs

The additional expressiveness of ADFs, which can formalize arbitrary relationships among arguments, comes at the price of higher computational complexity. Thus, an understanding of potentially easier subclasses is relevant. Compared to Dung's abstract argumentation frameworks, where several subclasses such as acyclic and symmetric frameworks are well understood, there has been no in-depth analysis for ADFs in such direction yet (with the notable exception of bipolar ADFs). In this chapter, we introduce certain subclasses of ADFs and investigate their properties. In particular, we show that for acyclic ADFs, the different semantics coincide. On the other hand, we show that the concept of symmetry is less powerful for ADFs than for AFs, and that further restrictions are required to achieve results that are similar to the known ones for Dung's frameworks. A particular such subclass, namely, support-free symmetric ADFs, turns out to be closely related to argumentation frameworks with collective attacks, also known as SETAFs; we investigate this relation in detail and obtain as a by-product that even for SETAFs, symmetry is less powerful than for AFs. We also discuss the role of odd-length cycles in the subclasses we have introduced. Finally, we analyse the expressiveness of the ADF subclasses that we introduce in terms of signatures.

## 8.1 Introduction

Since the landmark paper by Dung (1995) has been published in 1995, abstract argumentation frameworks (AFs) have gained more and more

significance in the AI domain. First of all, AFs have proven useful to capture the essence of different nonmonotonic formalisms. Moreover, AFs are nowadays an integral concept in several advanced argumentation-based formalisms in the sense that their semantics are defined based on a translation (typically called an instantiation) to Dung AFs. Finally, the relevance of AFs is witnessed by the *International Competition on Computational Models of Argumentation* (ICCMA), where systems for solving different problems on AFs compete on different tracks[1].

The fundamental of Dung is to abstract away from the content of particular arguments and to focus only on conflicts among arguments, where each argument is viewed as an atomic item. Therefore, the only information AFs take into account is whether an argument attacks another one or not. Semantics single out coherent subsets of arguments which "fit" together, according to specific criteria (Baroni et al., 2011). More formally, an AF semantics takes an argumentation framework as input and produces as output a collection of sets of arguments, called extensions. Complexity of the reasoning problems that can be defined for the several semantics for AFs is well understood (Dvořák and Dunne, 2018) and ranges from tractability up to the second level of the polynomial hierarchy. To this end, the analysis of restricted classes of AFs is of importance. In his chapter, Dung already showed that the class of acyclic (also known as well-founded) AFs leads to a collapse of the different semantics. Further studies include symmetric AFs (Coste-Marquis et al., 2005) and AFs under other graph-driven restrictions (Dunne, 2007). Symmetric AFs have been proven to satisfy the property of coherence (preferred and stable semantics coincide) and relatively-groundedness (the grounded extension is given by the intersection of the preferred extensions). Moreover, these restrictions make decision problems often easier from a complexity perspective. A fact that is particularly useful in connection with backdoor approaches (Dvořák et al., 2012) that utilize the distance to an easier fragment. This approach has, for instance, been realised in practice with the cegartix system (Dvořák et al., 2014).

Abstract dialectical frameworks (ADFs) are generalizations of Dung argumentation frameworks where arbitrary relationships among arguments can be formalized via propositional formulas which are attached to the arguments (Brewka and Woltran, 2010; Brewka et al., 2017b). This allows to express notions of support, collective attacks, and even more complicated relations. Due to their flexibility in formalizing relations between

[1] `http://argumentationcompetition.org`

arguments, ADFs have recently been used in several applications (Cabrio and Villata, 2016; Pührer, 2017; Al-Abdulkarim et al., 2016; Neugebauer, 2017). However, this additional expressibility comes with the price of higher computational complexity (Strass and Wallner, 2015). Specifically, reasoning in ADFs spans the first three (rather than the first two, as for AFs) levels of the polynomial hierarchy.

It is thus natural to investigate subclasses of ADFs. Compared to Dung argumentation frameworks, where subclasses like acyclic and symmetric AFs have been thoroughly studied, there has not been a systematic investigation of subclasses of ADFs yet. An exception is the class of bipolar ADFs (Brewka and Woltran, 2010) where the links between arguments are restricted to have either supporting or attacking nature. However, results about structural restrictions on ADFs where different semantics coincide are still lacking.

In this work, we aim to define several subclasses of ADFs and investigate how the restrictions we define influence the semantic evaluation of such ADFs. As a first class, we consider acyclic ADFs (i.e., the link-structure forms an acyclic graph) and show that—analogous to well-founded AFs—the main semantics, namely grounded, complete, preferred, and two-valued model/stable semantics, coincide for this class. We further investigate the concept of symmetric ADFs. In contrast to the case of AFs, we will see that properties as coherence and relatively-groundedness do not carry over and require further restrictions which leads us to the classes of acyclic support symmetric ADFs (ASSADFs) and support-free symmetric ADFs (SFSADFs). For both classes we show that they satisfy a weaker form of coherence. We also show that these two classes differ in the sense that odd-cycle free SFSADFs are coherent while odd-cycle free ASSADFs are not. As a second contribution, following the work of Dunne et al. (2015), we investigate the expressiveness of our ADF subclasses in terms of signatures, i.e. the set of possible outcomes which can be achieved by ADFs (of a particular class) under the different semantics. We thus complement here results which have been obtained for general (Pührer, 2015; Strass, 2015) and bipolar ADFs (Linsbichler et al., 2016) and also compare our ADF subclasses to abstract argumentation frameworks in terms of expressibility.

Our results lead to the following implications. Firstly, studying subclasses of ADFs provides us with a better understanding of which structures are required to reveal particular behaviors of the different semantics. We thus further advance the theory of ADFs. Secondly, since other generalizations of Dung AFs can be seen as special case of ADFs, results on ADFs

carry over to these special cases. We exemplify this aspect in the paper, by deriving new results for argumentation frameworks with collective attacks (SETAFs) (Nielsen and Parsons, 2006) which have received increasing interest recently (Dvořák et al., 2019; Flouris and Bikakis, 2019). To the best of our knowledge concepts like symmetric SETAFs have not been investigated yet, and we provide first results in this direction.

The chapter is structured as follows: In Section 8.2 we introduce subclasses of ADFs and we investigate whether these subclasses fulfill the same properties of the similar subclasses in AFs. We discuss how our results can be related to SETAFs and investigate some properties for symmetric SETAFs. Also the role of odd-length cycles is addressed. In Section 8.3, expressiveness of the subclasses of ADFs introduced in the current work is studied. In particular, we show that the expressiveness of SFSADFs, ASSADFs and bipolar ADFs is equal for some of the semantics, but different for admissibility-based semantics.

A preliminary version of this chapter appeared in (Diller et al., 2018). This extended version contains new technical results including investigations concerning coherence and relatively-groundedness for SFSADFs and symmetric SETAFs (Theorems 8.16 and 8.30); and results on the role of odd-cycles on coherence for subclasses of ADFs (Section 8.2.4). Also, the results on expressibility for SETAFs and SFSADFs (as well as for a further superclass of SFSADFs we introduce, that of SFADFs) in Section 8.3 are new.

## 8.2 Properties of ADF Subclasses

We start our investigation of ADF subclasses in terms of their semantics by first introducing the class of acyclic ADFs and showing that, just as is the case for acyclic AFs (Dung, 1995), the different semantics coincide on such ADFs. Then, we consider symmetric ADFs, where we will explore further restrictions that are needed in order to obtain results similar to the ones known for symmetric AFs. In Section 8.2.3 we discuss implications of our results for SETAFs. We conclude this section with a brief overview on semantic properties of odd-cycle free subclasses of ADFs.

### 8.2.1 Acyclic ADFs

In this section we show that, as has already been indicated and is the case for acyclic AFs, also for acyclic ADFs several semantics coincide (Theorem 8.3). We start by defining acyclic ADFs.

**Definition 8.1** *An ADF $D = (S, L, C)$ is acyclic if its corresponding directed graph $(S, L)$ is acyclic.*

In order to prove that the different semantics coincide on acyclic ADFs we need the concepts of level and maximum level of arguments. The *level* of an argument $s$ of an ADF $D$ is the number of links on the longest path from an initial argument to $s$ plus 1. The *maximum level* of an (acyclic) ADF $D$ then is the level of an argument of $D$ that is at least as large as the level of any other argument of $D$. It is clear that every acyclic ADF has a maximum level. This is a crucial observation needed for our proof of Proposition 8.2, which in turn provides the basis to show that most semantics defined for ADFs are indistinguishable when evaluating acyclic ADFs.

**Proposition 8.2** *In every acyclic ADF $D$ the $\leq_i$-least fixed point of $\Gamma_D$ is a model of $D$.*

**Proof** Let $D = (S, L, C)$ be an acyclic ADF and let $m$ be its maximum level. Moreover, let $v_0 := v_{\mathbf{u}}$ and $v_i := \Gamma_D(v_{i-1})$ for $1 \leq i \leq m$. We claim that for all $i$ with $1 \leq i \leq m$, and every argument $s_j$ with level $j \leq i$ it holds that either $v_i(s_j) = \mathbf{t}$ or $v_i(s_j) = \mathbf{f}$. We show this claim by induction on $i$:

- Base case: Suppose $s_1$ is an arbitrary argument of level one (an acyclic ADF always includes an initial argument). Since $s_1$ is an initial argument, either $\varphi_{s_1} = \top$ or $\varphi_{s_1} = \bot$. Hence $v_1(s_1) = \Gamma_D(v_0)(s_1)$ is either true or false.

- Inductive step: Assuming this property holds for all $k$ with $1 \leq k \leq i < m$, we show it holds for $i + 1$. We know that $\varphi_{s_j}^{v_i} = \varphi_{s_j}[s_k/\top : v_i(s_k) = \mathbf{t}][s_k/\bot : v_i(s_k) = \mathbf{f}]$. For all $s_k$ that occur in $\varphi_{s_j}$ it holds that $k < j \leq i + 1$, with $k$ being the level of $s_k$. Therefore, by the inductive hypothesis, for each $s_k$, either $v_i(s_k) = \mathbf{t}$ or $v_i(s_k) = \mathbf{f}$. Hence either $\varphi_{s_j}^{v_i} \equiv \top$ or $\varphi_{s_j}^{v_i} \equiv \bot$ and, consequently, either $v_{i+1}(s_j) = \mathbf{t}$ or $v_{i+1}(s_j) = \mathbf{f}$.

Since $m$ is the maximum level of any argument in $D$, we now get that $v_m(s)$ is either true or false for all $s \in S$, i.e. it is a two-valued interpretation. Moreover, it holds that $v_m = \Gamma_D(v_m)$, i.e. $v_m$ is a fixed point.

To show that $v_m$ is the least fixed point of $\Gamma_D$, assume, towards a contradiction, that there exists an interpretation $v <_i v_m$ such that

$v = \Gamma_D(v)$. Then there exists an argument $s$ such that either $v_m(s) = \mathbf{t}$ or $v_m(s) = \mathbf{f}$, but $v(s) = \mathbf{u}$. Assume $s$ has level $i$. Since $D$ is an acyclic ADF all arguments $s_k$ that occur in $\varphi_s$ have a level less than $i$. Therefore, there exists at least an argument $s_j$ of level $j < i$ in $\varphi_s$ such that $v(s_j) = \mathbf{u}$. By iterating this method after at most $i - 1$ times we reach an argument $s_1$ of level 1 for which $v(s_1) = \mathbf{u}$. This is a contradiction, since at level 1 all arguments are initial, it must be the case that either $\varphi_{s_1} = \top$ or $\varphi_{s_1} = \bot$, and therefore $\Gamma_D(v)(s_1) \neq \mathbf{u}$. Thus, interpretation $v_m$ is the least fixed point of $\Gamma_D$. □

**Theorem 8.3** *In every acyclic ADF $D$ the sets of grounded interpretations, complete interpretations, preferred interpretations, two-valued models, and stable models coincide.*

**Proof** First, the grounded interpretation $v$ of $D$ is also complete in $D$. Moreover, Proposition 8.2 implies that $v$ is a two-valued model of $D$. Since $w = w^{\mathbf{t}} = v^{\mathbf{t}}$ in which $w = grd(D^v)$, $v$ is a stable model. It remains to show that there is no further complete interpretation $v'$ of $D$. Since $v$ is two-valued it must hold that $v \not<_i v'$. However, since $v$ is grounded and therefore the least complete interpretation, such a $v'$ cannot exist. Therefore, $v$ is the unique complete interpretation of $D$ which is grounded, stable, two-valued, and preferred. □

An immediate consequence of Theorem 8.3 is that any acyclic ADF D possesses a non-trivial preferred interpretation, which is also a complete interpretation, grounded interpretation, stable interpretation, and a model. We conclude by noting that, on the other hand, if all semantics of an ADF coincide, there is no guarantee that the ADF in question is acyclic. This is shown via Example 8.4.

**Example 8.4** *Consider the ADF $D = (\{a, b, c\}, \{\varphi_a = \top, \varphi_b = \neg a \wedge \neg c, \varphi_c = \neg b\})$. This ADF possesses the unique complete interpretation $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$, which is also preferred, stable, grounded, as well as a model of $D$. That is, all semantics of $D$ coincide; however, $D$ is not acyclic.*

The ADF of Example 8.4 in fact represents an AF. Therefore, there is also no guarantee that an AF is acyclic, whenever all semantics yield the same extensions.

### 8.2.2 Symmetric ADFs

We turn now to our study of symmetric ADFs. We consider the properties of coherence (stable and preferred semantics coincide) and relatively-groundedness (grounded extension is the intersection of all preferred extensions) which have been shown to hold for symmetric AFs (Coste-Marquis et al., 2005). Since both the two-valued and stable model semantics for ADFs are proper generalisations of the stable semantics for AFs (Brewka et al., 2013), we consider further forms of coherence (weak coherence and semi-coherence; defined in Definition 8.6) that are possible in the realm of ADFs.

We will show that, contrary to symmetric AFs, symmetric ADFs do not satisfy any of the forms of coherence for ADFs we define, nor are they relatively-grounded (Theorem 8.8). We then define a further restricted form of symmetric ADFs, acyclic support symmetric ADFs, or ASSADFs for short (Definition 8.9), which we show do satisfy a weak form of coherence (each two-valued model is a stable model) (Theorem 8.12). Nevertheless, we conclude (Theorem 8.13) by showing that in ASSADFs it is still not the case that every preferred interpretation is a two-valued-model (semi-coherence). We also show that ASSADFs are not relatively-grounded (again, Theorem 8.13).

We start by giving the definition of symmetric ADFs.

**Definition 8.5** *An ADF $D = (S, L, C)$ is* symmetric *if $L$ is irreflexive and symmetric and $L$ does not contain any redundant links.*

The reason why we have to exclude redundant links is that otherwise we are able to add arbitrary links without changing the semantics of the ADF at hand: informally speaking, given an ADF $D = (S, L, C)$, take any link $(a, b) \in L$ such that $(b, a) \notin L$ and do the following: add $(b, a)$ to $L$ and change the acceptance condition $\varphi_a$ to $\varphi_a \wedge (\neg b \vee b)$. From the definition of the semantics, it follows that such a modification cannot change the set of $\sigma$-interpretations of ADF. Now, applying this modification exhaustively turns $L$ into a symmetric relation. The added links are clearly redundant since the newly introduced parent has no semantic effect on the altered acceptance condition.

Next we provide several notions of coherence which are possible for ADFs.

**Definition 8.6** *An ADF $D$ is called*

- coherent *if each preferred interpretation of $D$ is a stable model of $D$;*

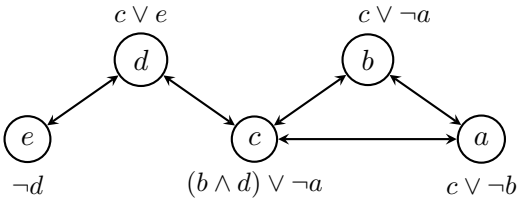


**Figure 8.1:** A symmetric ADF which is neither semi-coherent, weakly coherent nor relatively grounded.

- weakly coherent *if each two-valued model of $D$ is a stable model of $D$;*
- semi-coherent *if each preferred interpretation of $D$ is a two-valued model of $D$.*

We now turn to define the notion of relatively-groundedness for ADFs.

**Definition 8.7** *An ADF $D$ is called* relatively grounded *if $grd(D) = \bigsqcap_i prf(D)$.*

In what follows, we occasionally say that a class $\mathscr{C}$ of ADFs is coherent (resp. semi-coherent, weakly coherent, relatively grounded) if each of its elements satisfies the respective property.

It turns out that neither of the properties analogous to those holding for symmetric *AFs* hold for symmetric *ADFs*.

**Theorem 8.8** *The class of symmetric ADFs is neither semi-coherent, nor weakly coherent, nor relatively grounded.*

**Proof** Let $D$ be the symmetric ADF depicted in Figure 8.1. It holds that $prf(D) = \{v_1, v_2\}$ with $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$ and $v_2 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}, d \mapsto \mathbf{u}, e \mapsto \mathbf{u}\}$. Since $v_1$ is a two-valued model of $D$ which is not stable (since $D^v = D$ and $grd(D) = \{v_\mathbf{u}\}$), $D$ is not weakly coherent. Also, $D$ is not semi-coherent since $v_2$ is not two-valued. In addition, $\bigsqcap_i prf(D) = v_2 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}, e \mapsto \mathbf{u}\}$, but $grd(D) = \{v_\mathbf{u}\}$. Therefore, $D$ is not relatively grounded. □

Note that the ADF $D$ used as counter-example in the proof of Theorem 8.8 (Figure 8.1), is actually a symmetric BADF since $D$ is a symmetric ADF in which there are no dependent links and all links are either attacking or

supporting. This raises the question whether there is a particular subclass of symmetric BADFs which fulfills the properties considered in Theorem 8.8. One natural candidate is that of acyclic support symmetric ADFs, which we present in Definition 8.9.

**Definition 8.9** *Given an ADF $D = (S, L, C)$, let $L^+$ be the set of all supporting links in $D$. An ADF $D$ is an* acyclic support symmetric ADF *(ASSADF for short) if it is symmetric, bipolar and $(S, L^+)$ is acyclic.*

The method of determining whether a given ADF is an ASSADF is clarified in Example 8.10.

**Example 8.10** *Let $D = (S, L, C)$ be the symmetric BADF depicted in Figure 8.1. Since $(S, L^+)$ as shown in Figure 8.2 contains a cycle, namely the sequence $[d, c, d]$, $D$ is not an ASSADF.*

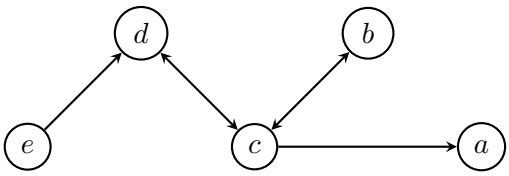


**Figure 8.2:** The support-links of the ADF $D$ from Theorem 8.8 (Figure 8.1).

We now show that ASSADFs are weakly coherent, using the following technical lemma.

**Lemma 8.11** *Let $D$ be an ADF, $v$ a two-valued model of $D$, and $s \in S$ an argument such that all parents of $s$ are attackers. $\varphi_s^v$ is irrefutable if and only if $\varphi_s[s_i/\bot : v(s_i) = \mathbf{f}]$ is irrefutable.*

**Theorem 8.12** *Every acyclic support symmetric ADF (ASSADF) is weakly coherent.*

**Proof** Let $D = (S, L, C)$ be an acyclic support symmetric ADF. We have to show that each two-valued model of $D$ is also a stable model of $D$. Let $v$ be a two-valued model of $D$, $D^v = (S^v, L^v, C^v)$ be the *stb*-reduct of $D$, $w$ be the unique grounded interpretation of $D^v$, and $\varphi'_s$ the acceptance condition of $s$ in $D^v$, i.e. $\varphi'_s = \varphi_s[s_i/\bot : v(s_i) = \mathbf{f}]$. We show that $v^{\mathbf{t}} = w^{\mathbf{t}}$. Suppose to the contrary that there exists an argument $s$, such that $v(s) = \mathbf{t}$ and $w(s) \neq \mathbf{t}$. That is, $\varphi'_s$ is not irrefutable. This means, by Lemma 8.11, that

$\varphi_s$ contains an argument $s_1$ supporting $s$ such that $v(s_1) = \mathbf{t}$, otherwise $\varphi_s$ cannot be irrefutable. Thus, $\varphi'_s$ also contains $s_1$. Since supports are acyclic in ASSADFs, for the same reasons $\varphi'_{s_1} = \varphi_{s_1}[s_i/\bot : v(s_i) = \mathbf{f}]$ contains an argument $s_2$ which is different from $s$ and $s_1$ and which supports $s_1$. Thus there exists an infinite sequence of arguments $s_1, s_2, \ldots$ such that $s_{i+1}$ supports $s_i$. This is a contradiction to $D$ being an ASSADF. □

We conclude this section by showing, on the other hand, that there are ASSADFs which are neither semi-coherent nor relatively grounded.

**Theorem 8.13** *The class of ASSADFs is neither semi-coherent nor relatively grounded.*

**Proof** Consider the ASSADF $D$ depicted in Figure 8.3. $D$ has 4 preferred interpretations, namely $v_1 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f},\ c \mapsto \mathbf{t},\ d \mapsto \mathbf{t},\ e \mapsto \mathbf{f}\}$, $v_2 = \{a \mapsto \mathbf{f},\ b \mapsto \mathbf{t},\ c \mapsto \mathbf{f},\ d \mapsto \mathbf{t},\ e \mapsto \mathbf{f}\}$, $v_3 = \{a \mapsto \mathbf{t},\ b \mapsto \mathbf{f},\ c \mapsto \mathbf{f},\ d \mapsto \mathbf{t},\ e \mapsto \mathbf{f}\}$, and $v_4 = \{a \mapsto \mathbf{u},\ b \mapsto \mathbf{u},\ c \mapsto \mathbf{u},\ d \mapsto \mathbf{f},\ e \mapsto \mathbf{t}\}$. As every two-valued interpretation of $D$ (that is $v_1, v_2$ and $v_3$) is also a stable model, $D$ is weakly coherent, confirming Theorem 8.12. However, $v_4$ is a preferred interpretation which is not a two-valued model. Hence, $D$ is not semi-coherent.

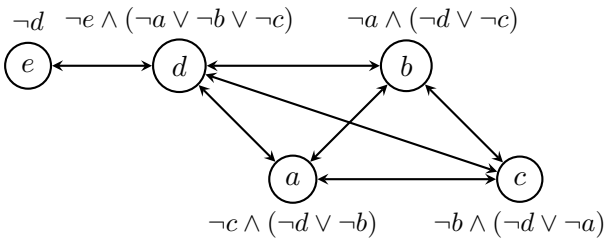


**Figure 8.3:** An ASSADF without supporting links that is not semi-coherent.

We show now that ASSADFs are not relatively grounded in general. Consider the ASSADF $D = (S, L, C)$, depicted in Figure 8.4. Here $D = (\{a, b, c\},\{\ \varphi_a : \neg b \wedge \neg c,\ \varphi_b : \neg a \wedge \neg c,$ and $\varphi_c : a \vee \neg b\})$. $D$ has the preferred interpretations $v_1 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$ and $v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}$. We obtain $v_1 \sqcap_i v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}$. However, the grounded interpretation of $D$ is the trivial interpretation $v_u$. That is, $D$ is not relatively grounded. □

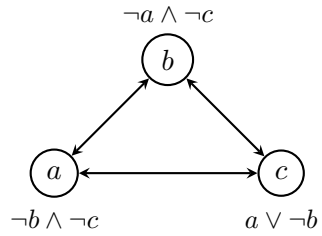


**Figure 8.4:** An ASSADF which is not relatively grounded.

### 8.2.3 Implications for SETAFs

The ASSADF used in the proof of Theorem 8.13 to show that ASSADFs are not semi-coherent does not have any supporting links. That is, even ASSADFs without supporting links are not semi-coherent. This leads us, in this section, to consider whether ASSADFs having only attacking links, which we call support-free symmetric ADFs or SFSADFs for short (Definition 8.14), satisfy the other properties considered in Section 8.2.2: being weakly coherent and relatively grounded. We show that SFSADFs are weakly coherent, but neither semi-coherent nor relatively grounded in Theorem 8.16.

Moreover, we derive from Theorem 8.16 that symmetric SETAFs are neither coherent nor relatively grounded (Theorem 8.30). The reason is that the SFSADFs we have used in the proof of Theorem 8.16 correspond to SETAFs. More concretely, the SETAFs in question correspond to a specific class of SFSADFs: those in which the acceptance condition of none of the arguments is unsatisfiable. On the way of proving Theorem 8.30 we show that, in fact, such SFSADFs exactly correspond to symmetric SETAFs (Theorem 8.18, Corollary 8.19, and Theorem 8.22; Lemmas 8.25 and 8.26). Thus, we obtain as a consequence of our investigations of semantic properties in the general settings of ADFs, results that complement those of (Nielsen and Parsons, 2006) for SETAFS, where the authors show that the standard semantics are indistinguishable on acyclic SETAFs (a result that is confirmed by our study in Section 8.2.1).

We start by defining SFSADFs:

**Definition 8.14** *Given an ADF $D = (S, L, C)$, let $L^-$ be the set of all attacking links in $D$. A bipolar ADF $D = (S, L, C)$ is a support free symmetric ADF (SFSADF for short) if it is symmetric and does not have any supporting links, that is, $L = L^-$.*

Note that since SFSADFs are BADFs by Definition 8.14, this means that SFSADFs do not have dependent links. Also, SFSADFs are symmetric by the same definition, which means that they do not have redundant links. Further, since SFSADFs do not have supporting links, they also do not have a support cycle. Thus, the class of SFSADFs is indeed a strict subclass of ASSADFs.

We next show that also SFSADFs, while being weakly coherent, are neither semi-coherent nor relatively grounded. Before doing so, we report a simple observation concerning the grounded interpretation.

**Lemma 8.15** *Let $D$ be an SFSADF with no isolated argument. The unique grounded interpretation of $D$ is the trivial interpretation, $v_{\mathbf{u}}$.*

**Proof** We show that for any SFSADF $D = (S, L, C)$ with no isolated argument, $\Gamma_D(v_{\mathbf{u}}) = v_{\mathbf{u}}$. Let $s$ be an argument. Let $v_1$ be an interpretation in which all parents of $s$ are assigned to $\mathbf{t}$ and let $v_2$ be an interpretation in which all $par(s)$ are assigned to $\mathbf{f}$. Since $D$ is an SFSADF, the former interpretation shows that $\varphi_s^{v_{\mathbf{u}}}$ is not irrefutable, since $\varphi_s^{v_1} \equiv \bot$ and the latter interpretation says that $\varphi_s^{v_{\mathbf{u}}}$ is not unsatisfiable, since $\varphi_s^{v_2} \equiv \top$. Therefore, for each argument $s$, $\Gamma_D(v_{\mathbf{u}})(s) = \mathbf{u}$. □

**Theorem 8.16** *The following properties hold for SFSADFs:*

- *every SFSADF is weakly coherent,*
- *the class of SFSADFs is not semi-coherent,*
- *the class of SFSADFs is not relatively grounded.*

**Proof** By Theorem 8.12, every ASSADF is weakly coherent, and since each SFSADF is an ASSADF, also SFSADFs are weakly coherent.

The ASSADF $D$, depicted in Figure 8.3 to show that the class of ASSADFs is not semi-coherent, does not have any supporting links. That is, $D$ is a SFSADF. Thus, the class of SFSADFs is not semi-coherent either.

It remains to show that the class of SFSADFs is not relatively grounded. Let $D$ be the ADF depicted in Figure 8.5. The unique preferred interpretation of $D$ is $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}\}$. However, since $D$ does not possess an isolated argument, Lemma 8.15 shows that the grounded interpretation of $D$ is the trivial interpretation. That is, the meet of the

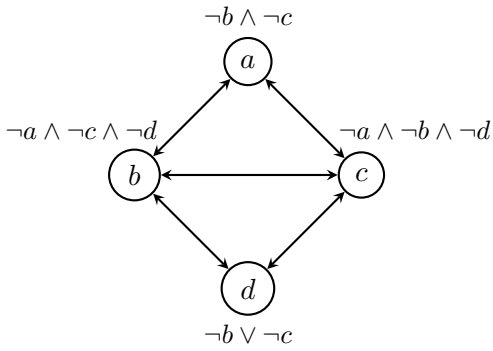


**Figure 8.5:** A SFSADF that is not relatively grounded.

preferred interpretations of $D$ is not equal to the grounded interpretation of $D$. Hence, $D$ is not relatively grounded. □

It is relatively easy to see that the ADFs used to show that SFSADFs are neither semi-coherent nor relatively grounded in the proof of Theorem 8.16 (ADFs from Figures 8.3 and 8.5) correspond to SETAFs (see Definitions 2.37 and 2.67). In fact, we proceed to show now that symmetric SETAFs are captured exactly by a subclass of SFSADFs: those in which the acceptance condition of none of the arguments is unsatisfiable. As already hinted at, apart from showing the link between SETAFs and SFSADFs this will allow us to also formally translate the content of Theorem 8.16 to the context of SETAFs.

We start by defining symmetric SETAFs.

**Definition 8.17** *A SETAF $F = (A, R)$, in which $R \subseteq (2^A \setminus \{\emptyset\} \times A)$, is a symmetric SETAF if the following properties hold:*

- *for all $(S, t) \in R$ and for all $s \in S$, there exists $(T, s) \in R$ such that $t \in T$,*
- *for each argument $s$ and for each $(S, s) \in R$, the set $S$ does not include $s$.*
- *for each $(S, s) \in R$ there is no $(S', s) \in R$ with $S' \subset S$.*

In Definition 8.17, the first item indicates that in the symmetric SETAFs all links are symmetric. The second item further means that there are also no reflexive links. Finally, the third item excludes redundant links.

From Definition 2.67 it follows that each SETAF can be represented as an ADF. Thus, also symmetric SETAFs can be encoded as ADFs. We now show that in fact symmetric SETAFs correspond to SFSADFs.

**Theorem 8.18** *The ADF associated to a given symmetric SETAF is a SFSADF.*

**Proof** Let $F = (A, R)$ be a symmetric SETAF. We show that the ADF $D_F = (S, L, C)$ associated to $F$ is a SFSADF. By Definition 2.67, $D_F$ does not contain any supporting link. It remains to show that $D_F$ is a symmetric ADF. It is clear that $L$ does not have any redundant links. We hence show that $L$ is symmetric and irreflexive. Towards a contradiction, thus assume that either $L$ is not symmetric or not irreflexive.

- Assume that $L$ is not symmetric. This means that there is an argument $s$ which is a parent of $t$ but not visa versa. That is, $s$ appears in the acceptance condition of $t$ but $t$ does not appear in the acceptance condition of $s$. Since $s$ is a parent of $t$, by Definition 2.67 there is $(S, t) \in R$ such that $s \in S$. Since $F$ is a symmetric SETAF, there is a $(T, s) \in R$ such that $t \in T$. Then, again via Definition 2.67, the argument $t$ appears in the acceptance condition of $s$. Thus, $t$ is a parent of $s$. This shows that the assumption that $L$ is not symmetric is false.

- Assume now that $L$ is not irreflexive. Therefore, there is an argument $s$ which is contained in $par(s)$. By Definition 2.67 there is $(S, s) \in R$ such that $s \in S$ which is in contradiction to $F$ being symmetric, cf. Definition 8.17.

Thus, if $F$ is a symmetric SETAF, then the associated ADF $D_F$ is a SFSADF. □

Let $F = (A, R)$ be a SETAF and let $a$ be an argument on which there is no attack in $F$, that is, there is no $(B, a) \in R$. By Definition 2.67, in the ADF corresponding to the SETAF F the acceptance condition of $a$ has the form $\varphi_a = \bigwedge_{(B,a)\in R} \bigvee_{a'\in B} \neg a' \equiv \top$. On the other hand, if there exists $(B, a) \in R$, it holds that $\varphi_a = \bigwedge_{(B,a)\in R} \bigvee_{a'\in B} \neg a' \not\equiv \bot$. These facts together with Theorem 8.18 lead to the following corollary.

**Corollary 8.19** *The SFSADF associated to a given symmetric SETAF does not contain any argument with an unsatisfiable acceptance condition.*

Next, we detail how the special group of SFSADFs that do not have any argument with an unsatisfiable acceptance condition can be represented as SETAFs. For this we make use of a fact from (Wallner, 2020), namely that ADFs for which the acceptance condition of each argument is either tautological or in CNF having only negative literals can be represented as SETAFs. Note that in symmetric ADFs each initial argument needs to be isolated and thus might have acceptance condition $\top$ or $\bot$; the latter is problematic in representing SFSADFs as symmetric AFs and thus needs special treatment.

**Lemma 8.20** *Let $D = (S, L, C)$ be a SFSADF and let $s \in S$ be an argument that is not isolated. Then the acceptance condition of the argument $s$ can be written in conjunctive normal form and having only negative literals.*

**Proof** Since the acceptance condition of each argument in an ADF is indicated by a propositional formula, it can be transformed to CNF. It remains to show that each of the resulting formulas can be written as a CNF consisting of only negative literals. Toward a contradiction, assume that $\varphi_s$ is the acceptance condition of an argument $s$ in CNF that is not isolated, but $\varphi_s$ cannot be written in CNF with only negative literals. That is, there is no $\varphi'_s$ such that $\varphi_s \equiv \varphi'_s$ and $\varphi'_s$ is in CNF with no negative literals, that is, $\varphi_s$ contains positive literals.

Assume that $t$ is the only argument that appears in $\varphi_s$ as a positive literal. The following method can be adapted for the case that $\varphi_s$ contains more than one positive literal. Let $\{c_t^i\}_{1 \leq i \leq n}$ be the set of all clauses $c_t^i$ in which $t$ occurs ($n \geq 1$). Also, let $v$ be a two-valued interpretation which assigns the truth value $\mathbf{f}$ to $t$; all other arguments in each $c_t^i$ are assigned to $\mathbf{t}$. Also $v$ assigns $\mathbf{f}$ to all other arguments of $par(s)$, which means that $v(\varphi_s) = \mathbf{f}$. However, $v|_{\mathbf{t}}^t(\varphi_s) = \mathbf{t}$. Thus, by the definition of attacking links, $(t, s)$ is not an attacking link in $D$. This is a contradiction with $D$ being a SFSADF. Note that if there exists $i$ with $1 \leq i \leq n$ such that $c_t^i = \neg t$, then $\varphi_s$ can be written in CNF with only negative literals. If in a $c_t^i$, the argument $t$ appears as a negative literal but it is not the only literal in $c_t^i$, the above method can be extended to show that $(t, s)$ is not an attacking link in $D$. □

We now provide the construction associating a SETAF to every SFSADF for which no argument has an unsatisfiable acceptance condition.

**Definition 8.21** *Let $D = (S, L, C)$ be a SFSADF in which there is no argument having an unsatisfiable acceptance condition. $D$ can be written*

*as a SETAF $F_D = (A, R)$ such that $A = S$ and $R$ is as follows. Let $\varphi_a$ be a CNF having only negative literals, let $c$ be a clause of $\varphi_a$ and let $R_c$ be the set of all arguments in the clause $c$. Then, $(R_c, a)$ represents a joint attack to $a$. The set $R_a = \{(R_c, a) \mid c \text{ is a clause in } \varphi_a\}$ is the set of all joint attacks to an argument $a$. Let $R = \bigcup_{a \in S} R_a$, be the set of all joint attacks in $F_D$. We call $F_D$ the SETAF associated to $D$.*

Analogously to Theorem 8.18, we show in Theorem 8.22 that SFSADFs in which none of the acceptance conditions of the arguments is unsatisfiable can be mapped to symmetric SETAFs.

**Theorem 8.22** *Let $D = (S, L, C)$ be a SFSADF in which the acceptance condition of none of the arguments is unsatisfiable. The SETAF $F_D$ associated to $D$ is a symmetric SETAF.*

**Proof** Assume that $D$ is a SFSADF in which there is no argument with an unsatisfiable acceptance condition. Thus, via Definition 8.21, $D$ can be written as a SETAF $F_D = (A, R)$. It remains to show that $F_D$ is a symmetric SETAF. Towards a contradiction, assume that $F_D$ is not symmetric. Then, there are two possibilities to consider:

- There exists $(S, t) \in R$ and there exists $s \in S$ such that for all $(T, s) \in R$, $T$ does not include $t$. Therefore, the acceptance condition of $t$ in $D$ includes $s$, however, the acceptance condition of $s$ in $D$ does not include $t$. That is, $L$ in $D$ is not symmetric.

- There exists $(S, t) \in R$ such that $t \in S$. That is, $t$ appears in the acceptance condition of $t$ in $D$. That is, $L$ in $D$ contains a reflexive link.

Both possibilities are in contradiction with the assumption that $D$ is a SFSADF. Therefore, the assumption that $F_D$ is not symmetric is not true. Then, the SETAF associated to a SFSADF is a symmetric SETAF. □

The SFSADF $D = (\{a, b, c, d\}, \{\varphi_a = \neg b \wedge \neg c, \varphi_b = \neg a \wedge \neg c \wedge \neg d, \varphi_c = \neg a \wedge \neg b \wedge \neg d, \varphi_d = \neg b \vee \neg c\})$, depicted in Figure 8.5, corresponds to the symmetric SETAF $F_D = (\{a, b, c, d\}, R)$ with $R = \{(\{a\}, b), (\{b\}, a), (\{a\}, c), (\{c\}, a), (\{b\}, c), (\{c\}, b), (\{b, c\}, d), (\{d\}, b), (\{d\}, c)\})$ depicted in Figure 8.6. As can be seen in the figure, in this SETAF there is a joint attack from $b$ and $c$ to $d$. This joint attack is symmetric in the sense of Definition 8.17, because of $(\{d\}, b) \in R$ and $(\{d\}, c) \in R$.

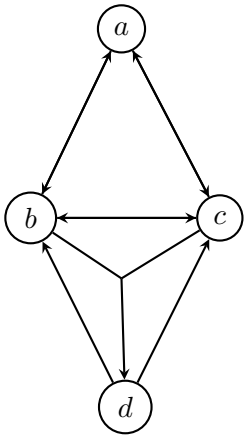


**Figure 8.6:** A symmetric SETAF that is not relatively grounded.

Now, in order to be able to relate SFSADFs and symmetric SETAFs on the semantic level, we first make precise the relation between extension-based semantics of AFs and interpretation-based semantics of ADFs. To do so, we start by introducing some terminology. Given a formalism $F$, the set of all extensions of $F$ are denoted by $\mathscr{E}$ and the set of all possible interpretations of $F$ are denoted by $\mathscr{V}$. The function $Ext2Int_F$, in Definition 8.23, is a modification of the function (associating labellings to extensions) given in Definition 5.1. of (Flouris and Bikakis, 2019).

**Definition 8.23** *Let $F = (A, R)$ be a SETAF, and let $e$ be an extension of $F$ ($e \in \mathscr{E}$). The truth value assigned to each argument $a \in A$ by the three-valued interpretation $v_e$ associated to $e$ is given by the function $Ext2Int_F : \mathscr{E} \to \mathscr{V}$ as follows.*

$$Ext2Int_F(e)(a) = \begin{cases} \mathbf{t} & a \in e, \\ \mathbf{f} & \exists B \in 2^A \text{ such that } (B, a) \in R \text{ and } \forall b \in B, b \in e, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

An ADF interpretation, on the other hand, can be represented as an extension via the following mapping.

**Definition 8.24** *Let $D = (S, L, C)$ be an ADF and $v$ an interpretation of $D$, that is, $v \in \mathscr{V}$. The associated extension $e_v$ of $v$ is obtained via application of the function $Int2Ext_D : \mathscr{V} \to \mathscr{E}$ on $v$, as follows:*

$$Int2Ext_D(v) = \{s \in S \mid s \mapsto \mathbf{t} \in v\}$$

The two subsequent lemmas are adopted from (Flouris and Bikakis, 2019).

**Lemma 8.25** *Let $F$ be a SETAF, and $\sigma \in \{prf, stb, mod, com, grd\}$. Moreover, let $D_F$ be the ADF associated to $F$. Then, $\sigma(D_F) = \{Ext2Int_F(e) \mid e \in \sigma(F)\}$.*

**Lemma 8.26** *Let $D$ be a SFSADF in which the acceptance condition of none of the arguments is unsatisfiable, $\sigma \in \{adm, prf, stb, mod, com, grd\}$ and $F_D$ the SETAF associated to $D$. Then, $\sigma(F_D) = \{Int2Ext_D(v) \mid v \in \sigma(D)\}$.*

Note that Lemma 8.25 does not mention admissible semantics, while Lemma 8.26 does. The reason is that for a given SETAF $F$ the associated three-valued interpretations obtained via $Ext2Int_F$ usually do not cover all admissible interpretations of the ADF $D_F$. This is illustrated next.

**Example 8.27** *Let $F = (\{a, b, c\}, \{\{a, b\}, c\})$ be a SETAF. By Definition 2.67, the associated ADF to $F$ is $D_F = (\{a, b, c\}, \{\varphi_a = \top, \varphi_b = \top, \varphi_c = \neg a \vee \neg b\})$. It is clear that $e = \{a, b\} \in adm(F)$. Applying $Ext2Int_F(e)$ of Definition 8.23 to $e$ leads to the three-valued interpretation $v_e = \{a \mapsto t, b \mapsto t, c \mapsto f\}$. However, $\{a \mapsto t, b \mapsto t, c \mapsto u\}$ is also an admissible interpretation of $D_F$ which is not obtained from any admissible extension $e$ of $F$ via $Ext2Int_F(e)$.*

One can overcome this problem by mapping each admissible set $e$ of a SETAF $F = (A, R)$ to several interpretations in a way that $Ext2Int_F(e)(a)$ yields either $\mathbf{u}$ or $\mathbf{f}$ in case there exists $(B, a) \in R$ such that $B \subseteq e$, and for all $(B', c)$ with $a \in B'$ and $c \in e$ there exists $(b \neq a) \in B'$ such that $Ext2Int_F(e)(b) = \mathbf{f}$. That is, $Ext2Int_F(e)(a)$ can be either $\mathbf{f}$ or $\mathbf{u}$, if $a$ is attacked by some arguments of $e$ but the truth value of $a$ does not play any role for the truth value of elements of $e$. However, since the forthcoming results do not involve admissible semantics, we leave a more formal investigation on this issue as topic for future work.

The next lemma rephrases Lemma 8.15 in terms of SETAFs.

**Lemma 8.28** *Let $F$ be a symmetric SETAF with no isolated argument. The unique grounded extension of $F$ is the empty set.*

**Proof** Let $F = (A, R)$ be a symmetric SETAF with no isolated argument, and let $D_F = (S, L, C)$ be the associated ADF of $F$. Toward a contradiction assume that the unique grounded extension $e$ (in $F$) is not the empty

set, that is, there exists argument $a$ such that $a \in e$. First, we show that $D_F$ does not contain any isolated argument. To this end, let $b$ be an argument. Since $F$ does not contain any isolated argument, there exists $B \subseteq A$ such that $(B, b) \in R$. The associated acceptance condition of $b$ in $D_F$, namely $\varphi_b = \bigwedge_{(B,b)\in R} \bigvee_{b' \in B} \neg b'$ shows that $par(b) \neq \{\}$ in $D_F$. Thus, the associated $D_F$ does not contain any isolated argument, as well. By Lemma 8.25 and Definition 8.23, $a$ is assigned to $\mathbf{t}$ in $grd(D_F)$. Further by Theorem 8.18, $D_F$ is an SFSADF. This is a contradiction by Lemma 8.15. Therefore, the unique grounded extension of $F$ is the empty set. □

The following corollary is a direct result of Lemma 8.28.

**Corollary 8.29** *A symmetric SETAF $F$ with no isolated argument is relatively grounded if and only if the intersection of all preferred extensions of $F$ is the empty set.*

We are now in position to derive that symmetric SETAFs are neither coherent nor relatively grounded from our proof of Theorem 8.16.

**Theorem 8.30** *The class of symmetric SETAFs is neither coherent nor relatively grounded.*

**Proof** The ADFs which are used in the proof of Theorem 8.16, to show that the class of SFSADFs is neither semi-coherent (and thus not coherent) nor relatively grounded, do not consist of any argument with an unsatisfiable acceptance condition. Then, by Theorem 8.22, the associated AFs to those SFSADFs are symmetric SETAFs.

Now, let $D$ be such an SFSADF that does not satisfy coherence. We show that the SETAF $F_D$ associated to $D$ cannot be coherent either. Let $w$ be a preferred interpretation of $D$ that is not a stable model of $D$. By Lemma 8.26, $prf(F_D) = \{Int2Ext_D(v) \mid v \in prf(D)\}$ and $stb(F_D) = \{Int2Ext_D(v) \mid v \in stb(D)\}$. Towards a contradiction, suppose $prf(F_D) = stb(F_D)$. It follows that there is an interpretation $u \in prf(D)$ with $u^{\mathbf{t}} = w^{\mathbf{t}}$, such that $u \in stb(D)$, and hence $u \neq w$. Hence, either $u^{\mathbf{f}} \subset w^{\mathbf{f}}$ or $w^{\mathbf{f}} \subset u^{\mathbf{f}}$. In the first case $u <_i w$ and hence $u$ cannot be preferred; in the second case $w <_i u$ and hence $w$ cannot be preferred. In both cases we have a contradiction.

For relatively groundedness, we already have provided a symmetric SETAF that violates this property in Figure 8.6. For that SETAF it can be checked that its complete sets are $\emptyset$ and $\{a, d\}$. Corollary 8.29 immediately implies that this SETAF cannot be relatively grounded. □

### 8.2.4 The Role of Odd-Length Cycles

In (Dunne and Bench-Capon, 2002) it is proven that if an AF is not coherent then it contains a cycle of odd length. This means, on the other hand, that if an AF does not contain any odd-length cycle it is coherent. It is easy to show that this property does not generalise to ADFs because of the possibility of support links. Indeed, consider the ADF $D = (\{a, b, c, d\}, \{\varphi_a = d, \varphi_b = a, \varphi_c = b, \varphi_d = c\})$. The interpretation $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, d \mapsto \mathbf{t}\}$ is a two-valued (and, hence, preferred) model of $D$, however it is not a stable model. That is, $D$ is not weakly coherent and, therefore, also not coherent.

In this section we study whether the property of having only odd-length-cycles implying coherence carries over to any subclasses of ADFs we introduced in our work so far. In Section 8.2.2 we had shown that ASSADFs are weakly coherent but not semi-coherent. We here show that also ASSADFs not containing any odd-length-cycle are not semi-coherent and, thus, that such ASSADFs are not coherent (Theorem 8.31). On the other hand, we are able to show that SFSADFs not containing any odd-length-cycle are coherent (Theorem 8.32). In fact we prove a more general result: bipolar ADFs without supporting links, which we dub SFADFs (Definition 8.33; note the difference with SFSADFs which have an "S" after the first "F"), that also do not have any odd-length-cycle are coherent (Corollary 8.34). Moreover, given that SETAFs correspond to a special class of SFSADFs (see Section 8.2.3), the result also applies to SETAFs.

**Theorem 8.31** *The subclass of ASSADFs not containing any odd-length cycle is not coherent.*

**Proof** Consider the ADF $D = (\{a, b\}, \{\varphi_a = b, \varphi_b = \neg a\})$. $D$ is an ASSADF in which there is no odd cycle. The unique preferred interpretation of $D$ is $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}\}$, however, $v$ is not a two-valued model. Then, $D$ is not semi-coherent and $D$ is not coherent either. □

Contrarily, we show next that the subclass of SFSADFs in which each ADF does not contain any odd cycle is coherent.

**Theorem 8.32** *The class of SFSADFs that do not contain any odd-length cycle is coherent.*

**Proof** Let $D = (S, L, C)$ be a SFSADF that does not contain any odd cycle. Since we showed in Theorem 8.16 that the class of SFSADFs is

weakly coherent, $D$ is weakly coherent, as well. To complete the proof of the theorem it is enough to show that $D$ is semi-coherent, i.e. $prf(D) = mod(D)$. Since each two-valued model is a preferred interpretation of ADFs, it is trivial that $prf(D) \supseteq mod(D)$. Thus, it is enough to show that $prf(D) \subseteq mod(D)$. Toward a contradiction, assume that there exist a preferred interpretation $v$ of $D$ that is not a two-valued model. That is, there must exist an argument $a$ such that $v(a) = \mathbf{u}$.

Let $S$ be the set of all arguments which are assigned the truth value $\mathbf{u}$ by $v$. Since $D$ is a SFSADF that does not contain any odd cycle, the arguments of $S$ have to be in even cycles in the associated graph. Assume that the associated graph of $D$ contains only one such even cycle. Note that the following method can be adapted for the case that the associated graph of $D$ contains more than one even cycle. Then one can construct a bipartite graph of nodes of this cycle with partitions $S_1$ and $S_2$. Assign all arguments in $S_1$ to $\mathbf{t}$, and all arguments of $S_2$ to $\mathbf{f}$. Construct the interpretation $v'$ as follows.

$$v'(a) = \begin{cases} v(a) & \text{if } \ v(a) = \mathbf{t}/\mathbf{f}, \\ \mathbf{t} & \text{if } \ a \in S_1, \\ \mathbf{f} & \text{if } \ b \in S_2. \end{cases}$$

It is clear that $v <_i v'$. We now show that $v'$ is a two-valued model. First, there is no argument in $v'$ assigned to $\mathbf{u}$. To show that $v'$ is a two-valued model it remains to show that $\Gamma_D(v') = v'$. Assume that $a$ is assigned to $\mathbf{t}$ in $v'$, we show that $\Gamma_D(v')(a) = \mathbf{t}$. (The method for proving the case that $a$ is assigned to $\mathbf{f}$ is analogous). If $a \mapsto \mathbf{t}$ in $v'$ either $v(a) = \mathbf{t}$ or $a \in S_1$.

- If $v(a) = \mathbf{t}$, since $v$ is a preferred interpretation, $\Gamma_D(v)(a) = \mathbf{t}$. In addition, the characteristic operator is a monotonic operator, that is, $\Gamma_D(v')(a) = \mathbf{t}$.
- If $a \in S_1$, then $par(a) \neq \{\}$. Let $\varphi_a$ be in CNF having only negative literals, this is possible by Lemma 8.20.
  - If $\Gamma_D(v')(a) = \mathbf{f}$, since $\varphi_a$ contains only negative literals, then there exists a clause $c$ in $\varphi_a$ all arguments of which are assigned to $\mathbf{t}$ in $v'$. By the construction of $S_1$, $par(a) \not\subseteq S_1$. Therefore, all arguments of $c$ are assigned to $\mathbf{t}$ in $v$, since $c$ has only negative literals, (note that all arguments in $S_2$ are assigned to $\mathbf{f}$ by the definition of $v'$). That is, $\Gamma_D(v)(a) = \mathbf{f}$, which is a contradiction with the assumption that $v(a) = \mathbf{u}$. Thus, $\Gamma_D(v')(a) \neq \mathbf{f}$.

– If $\Gamma_D(v')(a) = \mathbf{u}$, then there exists a parent of $a$ the truth value of which is not indicated in $v'$. This is also a contradiction, since all parents of $a$ are either in $S_2$ that are assigned to $\mathbf{f}$ in $v'$ or assigned to $\mathbf{t}/\mathbf{f}$ by $v$. Therefore, $\Gamma_D(v')(a) \neq \mathbf{u}$.

Thus, if $a \in S_1$, then $\Gamma_D(v')(a) = \mathbf{t}$.

Therefore, $v'$ is a two-valued model of $D$ and hence a preferred interpretation of $D$. Moreover $v <_i v'$, which is a contradiction to the assumption that $v$ is a preferred interpretation. Therefore, each preferred interpretation of a SFSADF that does not contain any odd cycle is a two-valued model. Thus, if a SFSADF does not contain any odd cycle, then it is coherent. □

As it turns out, the proof of Theorem 8.32 is independent from the notion of symmetry. Hence, we obtain as a final observation (Corollary 8.34) in this section that the general class of support-free ADFs which we define next is coherent.

**Definition 8.33** *Given an ADF $D = (S, L, C)$, let $L^-$ be the set of all attacking links in $D$. A bipolar ADF $D = (S, L, C)$ is called a support-free ADF (SFADF) if it does not have any supporting links, that is, $L = L^-$.*

**Corollary 8.34** *SFADFs and SETAFs without odd-length cycles are coherent.*

## 8.3 Expressiveness of ADF Subclasses

Following the work of (Dunne et al., 2015) in this section we consider the expressiveness, i.e. the set of possible outcomes which can be achieved under the different semantics, of ASSADFs and SFSADFs. We thus complement here results that have been obtained for general (Strass, 2015; Pührer, 2015) and bipolar ADFs (Linsbichler et al., 2016) and also compare the ADF subclasses we introduced in this work to abstract argumentation frameworks in terms of their expressivity.

Formally, the study of expressivity of a formalism w.r.t. a semantics can be done by considering the outcomes that can be realised by the formalism under the semantics of interest.

**Definition 8.35** *Let $\mathscr{F}$ be a formalism (e.g. AFs or (subclasses of) ADFs), i.e. the set of structures available in $\mathscr{F}$ (e.g. all possible AFs and ADFs) and $\sigma$ a semantics for $\mathscr{F}$. Moreover, let $\mathbb{V}$ be an interpretation-set*

*or extension-set.* $\mathbb{V}$ *is said to be* $\sigma$*-realizable in* $\mathscr{F}$*, if there exists an element* $kb$ *("knowledge base") of* $\mathscr{F}$ *such that* $\sigma(kb) = \mathbb{V}$*.*

The signature of a formalism w.r.t. a semantics is then the set of possible outcomes that can be realised by the formalism under the semantics, this is encoded in Definition 8.36.

**Definition 8.36** *The signature* $\Sigma^{\sigma}_{\mathscr{F}}$ *of a formalism* $\mathscr{F}$ *w.r.t. a semantics* $\sigma$ *is defined as:*

$$\Sigma^{\sigma}_{\mathscr{F}} = \{\sigma(kb) \mid kb \in \mathscr{F}\}.$$

*Formalisms can now be compared for expressivity by considering their signatures. Specifically, given two formalisms* $\mathscr{F}_1$*,* $\mathscr{F}_2$ *as well as a semantics* $\sigma$*, we say that* $\mathscr{F}_1$ *is strictly more expressive than* $\mathscr{F}_2$ *for* $\sigma$*, whenever* $\Sigma^{\sigma}_{\mathscr{F}_2} \subsetneq \Sigma^{\sigma}_{\mathscr{F}_1}$*.* $\mathscr{F}_1$ *and* $\mathscr{F}_2$ *are, on the other hand, incomparable under a semantics* $\sigma$ *if neither* $\Sigma^{\sigma}_{\mathscr{F}_1} \subseteq \Sigma^{\sigma}_{\mathscr{F}_2}$ *nor* $\Sigma^{\sigma}_{\mathscr{F}_2} \subseteq \Sigma^{\sigma}_{\mathscr{F}_1}$*. This is denoted as* $\Sigma^{\sigma}_{\mathscr{F}_1} \bowtie \Sigma^{\sigma}_{\mathscr{F}_2}$*.*

In what follows we concentrate on studying ASSADFs and SFSADFs from the perspective of realisability. We compare these novel subclasses of ADFs to that of AFs, BADFs, and general ADFs. We build on studies comparing the expressivity of AFs, BADFs and (general) ADFs reported on in (Strass, 2015; Linsbichler et al., 2016).

We begin by showing that BADFs are strictly more expressive than ASSADFs for the admissible, preferred, complete, and model semantics.

**Theorem 8.37** *For* $\sigma \in \{adm, prf, com, mod\}$ *it holds that* $\Sigma^{\sigma}_{ASSADF} \subsetneq \Sigma^{\sigma}_{BADF}$*.*

**Proof** Since every ASSADF is, by definition, a BADF, $\Sigma^{\sigma}_{\mathrm{ASSADF}} \subseteq \Sigma^{\sigma}_{\mathrm{BADF}}$ is clear. To show that $\Sigma^{\sigma}_{\mathrm{BADF}}$ is a strict superset of $\Sigma^{\sigma}_{\mathrm{ASSADF}}$ it is enough to find an interpretation-set $\mathbb{V}$ which is $\sigma$-realizable in BADFs, but not $\sigma$-realizable in ASSADFs, for $\sigma \in \{adm, prf, com, mod\}$.

For $\sigma \in \{prf, mod\}$, let $\mathbb{V} = \{\{a \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{f}\}\}$, and for $\sigma' \in \{com, adm\}$, let $\mathbb{V}' = \{\{a \mapsto \mathbf{u}\}, \{a \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{f}\}\}$. The BADF $D = (S, L, C)$ with $S = \{a\}$ and $\varphi_a = a$ realizes $\mathbb{V}$ under $\sigma$ and $\mathbb{V}'$ under $\sigma'$. On the other hand, it is easy to check that there is no ASSADF with one argument which realizes $\mathbb{V}$ under $\sigma$, and respectively, $\mathbb{V}'$ under $\sigma'$. To complete the proof toward a contradiction assume that there exists an ASSADF $D' = (S', L', C')$ such that $\sigma(D') = \mathbb{V}$ and $\sigma'(D') = \mathbb{V}'$. Since $\mathbb{V}$ contains an assignment to only one argument, $D'$ has to have just one

argument. Since $D'$ is an ASSADF, the argument $a$ is an isolated argument in $D'$. That is, either $\varphi_a = \top$ or $\varphi_a = \bot$. Thus, it can realize neither $\mathbb{V}$ under $\sigma$ nor $\mathbb{V}'$ under $\sigma'$. □

Next, we show that ASSADFs are strictly more expressive than SFSADFs for the admissible, preferred, and complete semantics. In a certain sense, this complements our observation in Section 8.2.4 that ASSADFs and SFSADFs differ in terms of coherence on odd-cycle free frameworks.

**Theorem 8.38** *For $\sigma \in \{adm, prf, com\}$, it holds that $\Sigma^{\sigma}_{SFSADF} \subsetneq \Sigma^{\sigma}_{ASSADF}$.*

**Proof** Since each support-free symmetric ADF is an acyclic support symmetric ADF, it is clear that $\Sigma^{\sigma}_{\text{SFSADF}} \subseteq \Sigma^{\sigma}_{\text{ASSADF}}$. To show that $\Sigma^{\sigma}_{\text{ASSADF}}$ is a strict superset of $\Sigma^{\sigma}_{\text{SFSADF}}$, for $\sigma \in \{adm, prf, com\}$, we give an interpretation-set $\mathbb{V}$, which is $\sigma$-realizable in ASSADFs, but not $\sigma$-realizable in SFSADFs.

Let $\mathbb{V} = \{\{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}\}\}$. The ASSADF $D = (S, L, C)$ with $S = \{a, b\}$, $\varphi_a = b$, and $\varphi_b = \neg a$ realizes $\mathbb{V}$ under $\sigma \in \{adm, prf, com\}$. However, it is easy to check that there is no SFSADF that can realize $\mathbb{V}$. Assume to the contrary that there exists a SFSADF $D' = (S', L', C')$ that can realize $\mathbb{V}$ under $\sigma$, for $\sigma \in \{adm, prf, com\}$. The set of arguments of $D'$ is $\{a, b\}$ as these are the only ones appearing in the $\sigma$ interpretations of $D'$.

- If any of the arguments of $S'$ is an isolated argument, then its acceptance condition is either equivalent to $\top$ or $\bot$. That is, $\sigma(D') \neq \mathbb{V}$. Thus, none of the arguments could be an isolated argument in $D'$.

- If there is a symmetric attack relation between $a$ and $b$, then the interpretation $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}\}$ is a preferred interpretation of $D'$ and therefore it is a complete and admissible interpretation of $D'$. Therefore, $\sigma(D') \neq \mathbb{V}$, for $\sigma \in \{adm, prf, com\}$.

Thus, $\mathbb{V}$ is not $\sigma$-realizable in SFSADFs under $\sigma \in \{adm, prf, com\}$. □

The forthcoming result shows why, on the other hand, AFs and SFSADFs are incomparable, and also AFs and ASSADFs are incomparable, in terms of their expressivity for the admissible, preferred, and complete semantics.

**Theorem 8.39** *For $\sigma \in \{adm, prf, com\}$, it holds that $\Sigma^{\sigma}_{AF} \bowtie \Sigma^{\sigma}_{SFSADF}$ and $\Sigma^{\sigma}_{AF} \bowtie \Sigma^{\sigma}_{ASSADF}$.*

**Proof** To obtain our theorem we show that $\Sigma^{\sigma}_{\text{AF}} \not\subseteq \Sigma^{\sigma}_{\text{ASSADF}}$ and $\Sigma^{\sigma}_{\text{SFSADF}} \not\subseteq \Sigma^{\sigma}_{\text{AF}}$. Since, via Theorem 8.38, $\Sigma^{\sigma}_{\text{ASSADF}}$ is a strict superset of $\Sigma^{\sigma}_{\text{SFSADF}}$ for $\sigma \in \{adm, prf, com\}$, we can then conclude that $\Sigma^{\sigma}_{\text{AF}} \not\subseteq \Sigma^{\sigma}_{\text{SFSADF}}$ and $\Sigma^{\sigma}_{\text{ASSADF}} \not\subseteq \Sigma^{\sigma}_{\text{AF}}$.

- To show $\Sigma^{\sigma}_{\text{AF}} \not\subseteq \Sigma^{\sigma}_{\text{ASSADF}}$ consider $\mathbb{V} = \{\{a \mapsto \mathbf{u}\}\}$. A witness of $\sigma$-realizability in AFs is $F = (\{a\}, \{(a, a)\})$. However, there is no ASSADF to realize $\mathbb{V}$ under $\sigma$.

- To verify that $\Sigma^{\sigma}_{\text{SFSADF}} \not\subseteq \Sigma^{\sigma}_{\text{AF}}$ for $\sigma \in \{adm, prf, com\}$ we first show that $\Sigma^{prf}_{\text{SFSADF}} \not\subseteq \Sigma^{prf}_{\text{AF}}$. Let $\mathbb{V} = \{v_1, v_2, v_3\}$ with $v_1 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, e \mapsto \mathbf{t}\}$, $v_2 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$, and $v_3 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, e \mapsto \mathbf{f}\}$. A witness of *prf*-realizability of $\mathbb{V}$ in SFSADFs is $D = (S, L, C)$ with $S = \{a, b, c, e\}$, $\varphi_a = \neg e \wedge (\neg b \vee \neg c)$, $\varphi_b = \neg a \vee \neg c$, $\varphi_c = \neg a \vee \neg b$, and $\varphi_e = \neg a$. However, there is no AF with $\mathbb{V}$ as its preferred interpretations. If there is an AF $F'$ such that $\sigma(F') = \mathbb{V}$ then the structure of $v_1, v_2$ and $v_3$ implies that there is no attack between $a, b$ and $c$ in $F'$. Thus, if there is an attack from any of $a, b$ and $c$ to $e$ then $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$ is a preferred interpretation of $F'$. If there is no attack from any of $a, b$ and $c$ to $e$ then $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, e \mapsto \mathbf{t}\}$ is a preferred interpretation of $F'$. In both cases $\sigma(F') \neq \mathbb{V}$. For $\sigma = com$ let $\mathbb{V}' = \mathbb{V} \cup \{\{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, e \mapsto \mathbf{u}\}\}$. It is easy to check that $\mathbb{V}'$ is *com*-realizable by the SFSADF $D$ defined above. If there is an AF $F'$ that realizes $\mathbb{V}'$ under *com* then each of the elements of $\mathbb{V}$ would be a preferred interpretation of $F'$. Thus, $prf(F') = \mathbb{V}$ would be the case, which it is easy to see is actually false. Finally, we get $\Sigma^{adm}_{\text{SFSADF}} \not\subseteq \Sigma^{adm}_{\text{AF}}$ by observing that from $adm(D)$ being realizable under *adm* in AFs it would follow that $prf(D)$ is realizable under *prf* in AFs. But we already showed that the latter is not the case.

□

Our final result on the admissibility-based semantics concerns the class of support-free ADFs (SFADFs). Recall that support-free symmetric ADFs (SFSADFs) have been defined in Section 8.2.3 in order to investigate subclasses of symmetric ADFs that satisfy certain properties.

**Theorem 8.40** *For $\sigma \in \{prf, adm, com\}$, the following hold:*

- $\Sigma^{\sigma}_{SFSADF} \subsetneq \Sigma^{\sigma}_{SFADF}$,

- $\Sigma^{\sigma}_{SFADF} \bowtie \Sigma^{\sigma}_{ASSADF}$,
- $\Sigma^{\sigma}_{AF} \subsetneq \Sigma^{\sigma}_{SFADF}$,
- $\Sigma^{\sigma}_{SFADF} \subsetneq \Sigma^{\sigma}_{BADF}$.

**Proof** We separately prove the four relations provided in the theorem.

- Each SFSADF is a SFADF, that is, $\Sigma^{\sigma}_{\text{SFSADF}} \subseteq \Sigma^{\sigma}_{\text{SFADF}}$. To show that $\Sigma^{\sigma}_{\text{SFADF}} \not\subseteq \Sigma^{\sigma}_{\text{SFSADF}}$, let $\mathbb{V} = \{\{a \mapsto \mathbf{u}\}\}$. A witness of $\sigma$-realizability of $\mathbb{V}$ in SFADFs is $D = (\{a\}, \{\varphi_a = \neg a\}\})$, for $\sigma \in \{prf, adm, com\}$. In contrast $\mathbb{V} \notin \Sigma^{\sigma}_{\text{SFSADF}}$.
- To show $\Sigma^{\sigma}_{\text{SFADF}} \not\subseteq \Sigma^{\sigma}_{\text{ASSADF}}$, let $\mathbb{V} = \{\{a \mapsto \mathbf{u}\}\}$, which is $\sigma$-realizable in SFADFs but not in ASSADFs. To show $\Sigma^{\sigma}_{\text{ASSADF}} \not\subseteq \Sigma^{\sigma}_{\text{SFADF}}$, let $\mathbb{V} = \{\{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}\}\}$. $\mathbb{V}$ can be realized by $D = (\{a, b\}, \{\varphi_a = \neg b, \varphi_b = a\})$ in ASSADF under $\sigma$ but not in SFADF.
- $\Sigma^{\sigma}_{AF} \subset \Sigma^{\sigma}_{SFADF}$ is clear. To show that $\Sigma^{\sigma}_{SFADF} \not\subseteq \Sigma^{\sigma}_{AF}$, let $\mathbb{V} = \{\{a \mapsto \mathbf{f}\}\}$ for $\sigma \in \{prf, com\}$, and respectively let $\mathbb{V}' = \{\{a \mapsto \mathbf{u}\}, \{a \mapsto \mathbf{f}\}\}$ for $\sigma = adm$. Both interpretation-sets are realizable by $D = (\{a\}, \{\varphi_a = \bot\})$ in SFADF under $\sigma$. In contrast $\mathbb{V} \notin \Sigma^{\sigma}_{\text{AF}}$.
- Each SFADF does not contain any dependent link; hence it is a BADF. The interpretations that are given in the proof of Theorem 8.37 work here to show that $\Sigma^{\sigma}_{\text{BADF}} \not\subseteq \Sigma^{\sigma}_{SFADF}$.

□

Our results comparing the expressivity of ASSADFs, SFSADFs, SFADFs, AFs, BADFs, and general ADFs w.r.t. the admissible, complete, and preferred semantics are summarised in Figure 8.7. To complete the picture here we also incorporate results about the relative expressivity of AFs, BADFs, and general ADFs from (Linsbichler et al., 2016).

The general picture for the stable semantics, which we proceed to investigate now, deviates somewhat from that of the admissibility based semantics. To start we remind the reader that stable models $v, w$ of any ADF are always incomparable, i.e. $w^{\mathbf{t}} \subseteq v^{\mathbf{t}}$ implies $w^{\mathbf{t}} = v^{\mathbf{t}}$, see (Strass, 2013b). Now, in order to complete previous results from (Strass, 2015) comparing AFs, BADFs and general ADFs in terms of their expressivity w.r.t. the stable semantics, we make use of this fact and the forthcoming lemma, in order to prove that $\Sigma^{stb}_{\text{BADF}} \subseteq \Sigma^{stb}_{\text{SFSADF}}$.

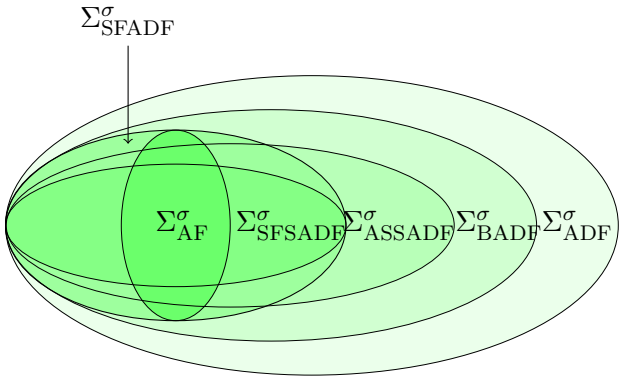


**Figure 8.7:** Expressiveness of subclasses of ADFs for $\sigma \in \{adm, prf, com\}$.

**Lemma 8.41** *Any incomparable set of two-valued interpretations* $\mathbb{V}$ *is stb-realizable in SFSADFs.*

**Proof** (sketch) To show that any incomparable set of two-valued interpretations is *stb*-realizable by some SFSADF, let $\mathbb{V}$ be an incomparable set over arguments $S$ (that is, each $v \in \mathbb{V}$ assigns $\mathbf{t}$ or $\mathbf{f}$ to every $s \in S$) and consider a SFSADF $D = (S, L, C)$ with the following acceptance conditions for $s \in S$:[2]

- If $v(s) = \mathbf{t}$ for every $v \in \mathbb{V}$ then $\varphi_s \equiv \top$.
- If $v(s) = \mathbf{f}$ for every $v \in \mathbb{V}$ then $\varphi_s \equiv \bot$.
- Otherwise, $\varphi_s = \bigvee_{v \in \mathbb{V}, v(s)=\mathbf{t}} \quad \bigwedge_{v(t)=\mathbf{f} \wedge \exists w \in \mathbb{V}:(w(s)=\mathbf{f} \wedge w(t)=\mathbf{t})} \neg t$.

We show that $D$ is a SFSADF and $stb(D) = \mathbb{V}$.

- By the definition of the acceptance condition of argument $s$ in $D$, $s \notin par(s)$. Therefore, $L$ is irreflexive. Further, $t \in par(s)$ iff $s \in par(t)$, that is, $L$ is symmetric. In addition, all links in $D$ are attacking. Thus, $D$ is a SFSADF.
- To prove that $stb(D) = \mathbb{V}$, we show that $stb(D) \subseteq \mathbb{V}$ and also $\mathbb{V} \subseteq stb(D)$.
  - To show $\mathbb{V} \subseteq stb(D)$, let $v \in \mathbb{V}$. We show that $v \in mod(D)$ and since SFSADFs are weakly coherent by Theorem 8.16,

[2]This construction is a slight adaption of a result from (Strass, 2015).

$v \in stb(D)$. Let $v(s) = \mathbf{t}$, we show that $s$ is acceptable in $v$. (The proof for the case that $v(s) = \mathbf{f}$ is analogous). If $s$ is assigned $\mathbf{t}$ by each element of $\mathbb{V}$ there is nothing to prove. Otherwise, there exists a $w \in \mathbb{V}$ s.t. $w(s) = \mathbf{f}$. Since $\mathbb{V}$ is incomparable, there exists $t$ such that $v(t) = \mathbf{f}$ and $w(t) = \mathbf{t}$.Therefore, $t \in par(s)$. The set of all arguments like $t$ make a conjunctive clause of $\varphi_s$ which guarantees that $s$ is accepted in $v$.

– To show that $stb(D) \subseteq \mathbb{V}$, toward a contradiction, assume that there exists $v \in stb(D)$ such that $v \notin \mathbb{V}$. Since $stb(D)$ is incomparable, there exists $s \in S$ such that $v(s) = \mathbf{t}$ and $\varphi_s \not\equiv \top$. Further, the acceptance condition of $s$ is not unsatisfiable, otherwise $s$ has to be assigned to $\mathbf{f}$ in $v$. Thus, there exists $v' \in \mathbb{V}$ in which $s$ is assigned to $\mathbf{t}$. Let $K$ be the set of all $v'$ in which $s$ is assigned to $\mathbf{t}$. Since $v \notin \mathbb{V}$ and $stb(D)$ is incomparable, in each $v' \in K$ there exists $t$ such that $v(t) = \mathbf{t}$ and $v'(t) = \mathbf{f}$. Let $T$ be a set of all arguments like $t$. It can be shown that either each conjunctive clause of $\varphi_s$ contains a $t \in T$, or there exists a $t \in T$ such that each conjunctive clause of $\varphi_t$ consists of an argument of $T$. The former means that $s$ is deniable with respect to $v$, and the latter means that $t$ is deniable in $v$. That is, $v$ is not a two-valued model of $D$ and since SFSADFs are weakly coherent, $v$ is not a stable model of $D$. This is in contradiction with our assumption that there exists $v \in stb(D)$ such that $v \notin \mathbb{V}$. Therefore, $stb(D) \subseteq \mathbb{V}$.

□

Note that the incomparability condition for the set of two valued interpretations $\mathbb{V}$ in the statement of Lemma 8.41 is necessary. For instance, $\mathbb{V} = \{\{a \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{f}\}\}$ is a set of two-valued interpretations, which are not incomparable, that is not *stb*-realizable in SFSADFs.

**Theorem 8.42** $\Sigma^{stb}_{AF} \subsetneq \Sigma^{stb}_{SFSADF} = \Sigma^{stb}_{SFADF} = \Sigma^{stb}_{ASSADF} = \Sigma^{stb}_{BADF}$.

**Proof** $\Sigma^{stb}_{\mathrm{AF}} \subsetneq \Sigma^{stb}_{\mathrm{BADF}}$ is shown in (Strass, 2015), and $\Sigma^{\sigma}_{\mathrm{SFSADF}} \subseteq \Sigma^{\sigma}_{\mathrm{ASSADF}} \subseteq \Sigma^{\sigma}_{\mathrm{BADF}}$ and $\Sigma^{\sigma}_{\mathrm{SFSADF}} \subseteq \Sigma^{\sigma}_{SFADF} \subseteq \Sigma^{\sigma}_{\mathrm{BADF}}$, for each semantics $\sigma$, are clear. If we can show $\Sigma^{stb}_{\mathrm{BADF}} \subseteq \Sigma^{stb}_{\mathrm{SFSADF}}$ we are thus done. Let $\mathbb{V} \in \Sigma^{stb}_{\mathrm{BADF}}$. Since $\mathbb{V}$ is a set of incomparable two-valued interpretations, by Lemma 8.41, $\mathbb{V}$ is *stb*-realizable in SFSADFs. □

In Example 8.43 we give an example of an interpretation-set that is *stb*-realizable in SFSADFs but not *stb*-realizable in AFs.

**Example 8.43** *Let* $\mathbb{V} = \{\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}, \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}\}$ *be an interpretation-set. A witness of stb-realizability of* $\mathbb{V}$ *in SFSADFs is* $D = (S, L, C)$ *for which* $S = \{a, b, c\}$ *and the acceptance conditions are* $\varphi_a = \neg b \vee \neg c$, $\varphi_b = \neg a \vee \neg c$ *and* $\varphi_c = \neg a \vee \neg b$. *We show that* $\mathbb{V}$ *is not stb-realizable in AFs. Towards a contradiction assume that there exists an AF* $F = (A, R)$ *such that* $stb(F) = \mathbb{V}$. *The set of arguments of* $F$ *is* $S$ *as these are the arguments occurring in* $stb(F)$. *The interpretations in* $\mathbb{V}$ *imply that there is no link between* $a$, $b$ *and* $c$ *in* $F$. *Then, by the definition of the stable semantics for AFs, the interpretation* $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$ *is also a stable interpretation of* $F$. *Therefore,* $stb(F) \neq \mathbb{V}$. *Thus, the interpretation-set* $\mathbb{V}$ *is not stb-realizable in AFs and* $\Sigma^{stb}_{SFSADF} \not\subseteq \Sigma^{stb}_{AF}$

The final semantics to investigate is the model-semantics. We only need one technical lemma.

**Lemma 8.44** *For any SFADF* $D = (S, L, C)$, $mod(D)$ *is incomparable.*

**Proof** Toward a contradiction assume that there are $v, w \in mod(D)$ such that $v^{\mathbf{t}} \subset w^{\mathbf{t}}$. Let $B \subset S$ be the set of argument which are assigned to $\mathbf{t}$ in $w$, but are assigned to $\mathbf{f}$ by $v$. Moreover, let $a$ be an argument which is denied in $v$ and accepted in $w$. At least a parent of $a$ has to be in $B$, otherwise, $\varphi^v_a = \varphi^w_a$. Assume $par(a) \cap B = \{b_1, \ldots, b_n\}$. Since $D$ is a SFADF, all links are attacking. Then, by the definition of attacking links $v|^{b_1}_{\mathbf{t}}(\varphi_a) = \mathbf{f}$. That is, $a$ is denied with respect to $v_1 = v|^{b_1}_{\mathbf{t}}$. Following the same method construct $v_i = v|^{b_i}_{\mathbf{t}}$, for $1 \leq i \leq n$. It is obvious that $a$ is denied in each $v_i$, in particular, $a$ is denied in $v_n$, which is equal to $w$. This is a contradiction to the assumption that $a$ is accepted in $w$. Therefore, $mod(D)$ is a set of incomparable interpretations. □

**Theorem 8.45** $\Sigma^{mod}_{AF} \subsetneq \Sigma^{mod}_{SFSADF} = \Sigma^{mod}_{SFADF} = \Sigma^{mod}_{ASSADF} \subsetneq \Sigma^{mod}_{BADF}$.

**Proof** We first argue that $\Sigma^{mod}_{\mathscr{F}} = \Sigma^{stb}_{\mathscr{F}}$ for $\mathscr{F} \in \{\text{SFSADF}, \text{SFADF}, \text{ASSADF}\}$. This follows immediately for $\mathscr{F} \in \{\text{SFSADF}, \text{ASSADF}\}$ since both ASSADFs and SFSADFs are weakly coherent (cf. Theorem 8.12 and Theorem 8.16). To show the relation for $\mathscr{F} = \text{SFADF}$ we need to show $\Sigma^{mod}_{\text{SFADF}} \subseteq \Sigma^{stb}_{\text{SFADF}}$ (the other inclusion follows by standard properties of semantics). Let $\mathbb{V} \in \Sigma^{mod}_{\text{SFADF}}$. By Lemma 8.44, $\mathbb{V}$ is incomparable and by Lemma 8.41, any incomparable set of two-valued interpretations is

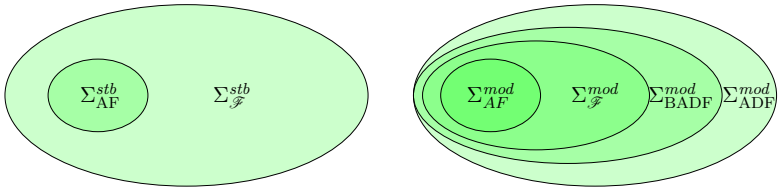


**Figure 8.8:** Expressiveness of AFs, SFSADFs, SFADFs, ASSADFs, BADFs, ADFs for $\sigma \in \{stb, mod\}$.

*stb*-realizable in SFSADF. Therefore, there is a SFSADF $D'$ such that $stb(D') = \mathbb{V}$. Since each SFSADF is also an SFADF, $\mathbb{V} \in \Sigma^{stb}_{\mathrm{SFADF}}$.

Using these relations and observing that *stb* and *mod* are equivalent for AFs, it follows from Theorem 8.42 that $\Sigma^{mod}_{\mathrm{AF}} \subsetneq \Sigma^{mod}_{\mathrm{SFSADF}} = \Sigma^{mod}_{\mathrm{SFADF}} = \Sigma^{mod}_{\mathrm{ASSADF}}$ holds. Finally, $\Sigma^{mod}_{\mathrm{ASSADF}} \subsetneq \Sigma^{mod}_{\mathrm{BADF}}$ is by Theorem 8.37. □

Figure 8.8 summarises the results regarding expressivity w.r.t. the model and stable semantics expressed in Theorem 8.42 and Theorem 8.45. Again, to complete the picture we make use of results from (Strass, 2015) ($\Sigma^{stb}_{\mathrm{BADF}} = \Sigma^{stb}_{\mathrm{ADF}}$ and $\Sigma^{mod}_{\mathrm{BADF}} \subset \Sigma^{mod}_{\mathrm{ADF}}$).

We conclude this section by pointing out that the realisability relationships depicted in Figure 8.8 change when we restrict the cardinality of the interpretation sets. As it turns out, any set of interpretations of size 2 obtained from an ADF when evaluated using the stable semantics is also realizable in AFs.

**Proposition 8.46** *Suppose that* $|\mathbb{V}| = 2$ *and* $\mathbb{V}$ *is stb-realizable in ADFs. Then* $\mathbb{V}$ *is stb-realizable in AFs.*

**Proof** Let $\mathbb{V} = \{v_1, v_2\}$ be a set of interpretations that is *stb*-realizable in ADFs, i.e. there exists an ADF $D = (S, L, C)$ such that $stb(D) = \mathbb{V}$. Construct an AF $F = (A, R)$ by setting $A = S$ and $R = \{(a, b) \mid v_i(a) = \mathbf{t}, v_i(b) = \mathbf{f}, v_j(a) = \mathbf{f}, 1 \leq i \neq j \leq 2\}$. To prove that $stb(F) = \mathbb{V}$, take $v_i \in \mathbb{V}$. First, there is no attack between arguments with $v_i(a) = \mathbf{t}$. Moreover, if $v_i(b) = \mathbf{f}$ then, since neither $v_1 \leq_i v_2$ nor $v_2 \leq_i v_1$, there must be some $a \in A$ with $v_i(a) = \mathbf{t}$ and $v_j(a) = \mathbf{f}$, hence $(a, b) \in R$. Therefore $v_i$ is a stable interpretation of $F$. That is, $\mathbb{V}$ is *stb*-realizable in AFs. □

## 8.4 Conclusion

Motivated by related results on the semantic properties of acyclic (Dung, 1995), symmetric (Coste-Marquis et al., 2005), and odd-length-cycle-free (Dunne and Bench-Capon, 2002) AFs, in this chapter we investigated analogous classes for ADFs and their properties. We showed that for acyclic ADFs, just as is the case for acyclic AFs, the different semantics coincide. On the other hand, we demonstrated that the properties of coherence and relatively-groundedness that hold for symmetric AFs do not carry over to symmetric ADFs. The latter impelled us to go on a quest for an appropriate subclass of symmetric ADFs for which some form of coherence holds. In the process we defined several subclasses, in particular acyclic support symmetric ADFs (ASSADFs) and support-free symmetric ADFs (SFSADFs) which we show satisfy a weaker form of coherence.

Also odd-length-cycle-free ADFs do not satisfy coherence (which is the case for AFs), but here we were able to show that this property does hold for SFSADFs. In fact this is the case for a superclass of SFSADFs (which also contains AFs with collective attacks), which we dubbed SFADFs. This property also allowed us to distinguish ASSADFs from SFSADFs in that odd-length-cycle-free ASSADFs are not coherent in general.

The motivation behind this line of investigation lies in the fact that different semantics show different complexities (Strass and Wallner, 2015). It is thus valuable to know under which circumstances higher complexities can be avoided. Acyclicity is a positive example since the coincidence with grounded semantics shows that, for instance, the more complex preferred semantics becomes easier for this class of frameworks. The practical implication is as follows: an ADF solver should check for acyclicity before computing preferred interpretations, since in case the ADF to be treated is acyclic, the easier procedure for grounded semantics suffices to do the job. As we have shown, such a strategy does not carry over to symmetric ADFs. This is in contrast to symmetric AFs where coherence holds, i.e. the (more complex) preferred semantics coincides with the (easier) stable semantics. Nonetheless, there is still a chance that symmetric ADFs are of practical help. In contrast to acyclic ADFs where the complexity drop is immediate, our results underline that dedicated complexity analyses for symmetric ADFs should be considered as future work.

As a further contribution and also following in the footsteps of work for AFs (Dunne et al., 2015), we considered the subclasses of ADFs we introduced (ASSADFs, SFADFs, and SFSADFs) in terms of their expressivity as can be gleaned from their signatures. Here SFSADFs

are a strict subset of ASSADFs for the admissibility-based semantics we considered, while SFADFs are incomparable w.r.t. ASSADFs (and a strict superset of SFSADFs). On the other hand the signatures of ASSADFs, SFADFs, and SFSADFs coincide for the model and stable semantics.

We also complemented previous work on expressivity of AFs and ADFs (Strass, 2015; Pührer, 2015; Linsbichler et al., 2016) by comparing the expressivity of ASSADFs, SFADFs, and SFSADFs with that of AFs, bipolar ADFs (BADFs), and ADFs. Here ASSADFs and SFSADFs are incomparable with AFs for the admissibility semantics, while SFADFs are strictly more expressive. ASSADFs, SFADFs, and SFSADFs are strictly more expressive for the model and stable semantics. On the other hand, they are strictly less expressive than BADFs for the model and admissibility based semantics, while they coincide in expressivity with BADFs and general ADFs for the stable semantics.

This work is an elaboration on the more theoretical aspects of our work presented in (Diller et al., 2018). There we had also included results on an empirical evaluation of some of the main systems for ADFs (`QADF` (Diller et al., 2014), `YADF` (Brewka et al., 2017a), and `goDIAMOND` (Strass and Ellmauthaler, 2017)) on acyclic vs. non-acyclic ADFs. These show usually only a slight improvement of these systems (which do not detect subclasses of ADFs) on the acyclic instances despite the fact that the results we present in this work indicate that even the most difficult of reasoning problems become tractable for acyclic ADFs. Thus, our work offers further guidelines for designing more efficient systems for ADFs. As a notable example and as we suggested in the introduction, backdoor approaches that utilize the distance to subclasses with easier complexity, such as the systems `cegartix` (Dvořák et al., 2014) for AFs as well as the very recent `k++ADF` (Linsbichler et al., 2018) for ADFs, can be devised on the basis of our results.

Finally, also from a theoretical perspective our work leaves open several further avenues to explore. In particular, we consider to extend our studies to other ADF semantics (Gaggl et al., 2015; Polberg, 2016) as well as generalizations of ADFs (Brewka et al., 2018b).

# Chapter 9

# Expressiveness of SETAFs and Support-Free ADFs under 3-Valued Semantics

Generalizing the attack structure in argumentation frameworks (AFs) has been studied in different ways. Most prominently, the binary attack relation of Dung's frameworks has been extended to the notion of collective attacks. The resulting formalism is often termed SETAFs. Another approach is provided via abstract dialectical frameworks (ADFs), where acceptance conditions specify the relation between arguments; restricting these conditions naturally allows for so-called support-free ADFs. The aim of this chapter is to shed light on the relation between these two different approaches. To this end, we investigate and compare the expressiveness of SETAFs and support-free ADFs under the lens of 3-valued semantics. Our results show that it is only the presence of unsatisfiable acceptance conditions in support-free ADFs that discriminates the two approaches.

## 9.1 Introduction

Abstract argumentation frameworks (AFs) as introduced by Dung (1995) are a core formalism in formal argumentation. A popular line of research investigates extensions of Dung AFs that allow for a richer syntax (see, e.g. (Brewka et al., 2014)). In this chapter we investigate two generalisations of Dung AFs that allow for a more flexible attack structure (but do not consider support between arguments).

The first formalism we consider are SETAFs as introduced by Nielsen

and Parsons (2006). SETAFs extend Dung AFs by allowing for collective attacks such that a set of arguments $B$ attacks another argument $a$ but no proper subset of $B$ attacks $a$. Argumentation frameworks with collective attacks have received increasing interest in the last years. For instance, semi-stable, stage, ideal, and eager semantics have been adapted to SETAFs in (Dvořák et al., 2019; Flouris and Bikakis, 2019); translations between SETAFs and other abstract argumentation formalisms are studied in (Polberg, 2017); (Yun et al., 2018) observed that for particular instantiations, SETAFs provide a more convenient target formalism than Dung AFs. The expressiveness of SETAFs with two-valued semantics has been investigated in (Dvořák et al., 2019) in terms of signatures. Signatures have been introduced in (Dunne et al., 2015) for AFs. In general terms, a signature for a formalism and a semantics captures all possible outcomes that can be obtained by the instances of the formalism under the considered semantics. Besides that, signatures are recognized as crucial for operators in dynamics of argumentation (cf. (Baumann and Brewka, 2019)).

The second formalism we consider are support-free abstract dialectical frameworks (SFADFs), a subclass of abstract dialectical frameworks (ADFs) (Brewka et al., 2018a) which are known as an advanced abstract formalism for argumentation, that is able to cover several generalizations of AFs (Brewka et al., 2014; Polberg, 2017). This is accomplished by acceptance conditions which specify, for each argument, its relation to its neighbour arguments via propositional formulas. These conditions determine the links between the arguments which can be, in particular, attacking or supporting. SFADFs are ADFs where each link between arguments is attacking; they have been introduced in a recent study on different sub-classes of ADFs (Diller et al., 2020).

For comparison of the two formalisms, we need to focus on 3-valued (labelling) semantics (Verheij, 1996; Caminada and Gabbay, 2009), which are integral for ADF semantics (Brewka et al., 2018a). In terms of SETAFs, we can rely on the recently introduced labelling semantics in (Flouris and Bikakis, 2019). We first define a new class of ADFs (SETADFs) where the acceptance conditions strictly follow the nature of collective attacks in SETAFs and show that SETAFs and SETADFs coincide for the main semantics, i.e. the $\sigma$-labellings of a SETAF are equal to the $\sigma$-interpretations of the corresponding SETADF. We then provide exact characterisations of the 3-valued signatures for SETAFs (and thus for SETADFs) for most of the semantics under consideration. While SETADFs are a syntactically defined subclass of ADFs, the second formalism we study can be understood

as semantical subclass of ADFs. In fact, for SFADFs it is not the syntactic structure of acceptance conditions that is restricted but their semantic behavior, in the sense that all links need to be attacking. The second main contribution of this chapter is to determine the exact difference in expressiveness between SETADFs and SFADFs.

We briefly discuss related work. The expressiveness of SETAFs has first been investigated in (Linsbichler et al., 2016) where different sub-classes of ADFs, i.e. AFs, SETAFs and Bipolar ADFs, are related w.r.t. their signatures of 3-valued semantics. Moreover, they provide an algorithm to decide realizability in one of the formalisms under different semantics. However, no explicit characterisations of the signatures are given. Recently, Pührer (2020b) presented explicit characterisations of the signatures of general ADFs (but not for the sub-classes discussed above). In contrast, (Dvořák et al., 2019) provides explicit characterisations of the two-valued signatures of SETAFs and shows that SETAFs are more expressive than AFs. In both works all arguments are relevant for the signature, while in (Flouris and Bikakis, 2019) it is shown that when allowing to add extra arguments to an AF which are not relevant for the signature, i.e. the extensions/labellings are projected on common arguments, then SETAFs and AFs are of equivalent expressiveness. Other recent work (Wallner, 2020) already implicitly showed that SFADFs with satisfiable acceptance conditions can be equivalently represented as SETAFs. This provides a sufficient condition for rewriting an ADF as SETAF and raises the question whether it is also a necessary condition. In fact, we will show that a SFADF has an equivalent SETAF if and only if all acceptance conditions are satisfiable. Different sub-classes of ADFs (including SFADFs) have been compared in (Diller et al., 2020), but no exact characterisations of signatures as we provide here are given in that work.

To summarize, the main contributions of this chapter are as follows:

- We embed SETAFs under 3-valued labeling based semantics (Flouris and Bikakis, 2019) in the more general framework of ADFs. That is, we show 3-valued labeling based SETAF semantics to be equivalent to the corresponding ADF semantics. As a side result, this also shows the equivalence of the 3-valued SETAF semantics in (Linsbichler et al., 2016) and (Flouris and Bikakis, 2019).

- We investigate the expressiveness of SETAFs under 3-valued semantics by providing exact characterizations of the signatures for preferred, stable, grounded and conflict-free semantics, thus com-

plementing the investigations on expressiveness of SETAFs (Dvořák et al., 2019) in terms of extension-based semantics.

- We study the relations between SETAFs and support-free ADFs (SFADFs). In particular we give the exact difference in expressiveness between SETAFs and SFADFs under conflict-free, admissible, preferred, grounded, complete, stable and two-valued model semantics.

## 9.2 Embedding SETAFs in ADFs

As observed by Polberg (2016) and Linsbichler et.al (2016), the notion of collective attacks can also be represented in ADFs by using the right acceptance conditions. We next introduce the class SETADFs of ADFs for this purpose.

**Definition 9.1** *An ADF $D = (S, L, C)$ is called SETAF-like (SETADF) if each of the acceptance conditions in $C$ is given by a formula (with $\mathscr{C}$ a set of non-empty clauses)*

$$\bigwedge_{cl \in \mathscr{C}} \bigvee_{a \in cl} \neg a.$$

That is, in a SETADF each acceptance condition is either $\top$ (if $\mathscr{C}$ is empty) or a proper CNF formula over negative literals. SETADFs and SETAFs can be embedded in each other as follows.

**Definition 9.2** *Let $F = (A, R)$ be a SETAF. The ADF associated to $F$ is a tuple $D_F = (S, L, C)$ in which $S = A$, $L = \{(a, b) \mid (B, b) \in R, a \in B\}$ and $C = \{\varphi_a\}_{a \in S}$ is the collection of acceptance conditions defined, for each $a \in S$, as*

$$\varphi_a = \bigwedge_{(B,a) \in R} \bigvee_{a' \in B} \neg a'.$$

*Let $D = (S, L, C)$ be a SETADF. We construct the SETAF $F_D = (A, R)$ in which, $A = S$, and $R$ is constructed as follows. For each argument $s \in S$ with acceptance formula $\bigwedge_{cl \in \mathscr{C}} \bigvee_{a \in cl} \neg a$ we add the attacks $\{(cl, s) \mid cl \in \mathscr{C}\}$ to $R$.*

Clearly the ADF $D_F$ associated to a SETAF $F$ is a SETADF and $D$ is the ADF associated to the constructed SETAF $F_D$. We next deal with the fact that SETAF semantics are defined as three-valued labellings while

semantics for ADFs are defined as three valued interpretations. In order to compare these semantics we associate the *in* label with $t$, the *out* label with $f$, and the *undec* label with $u$.

**Theorem 9.3** *For $\sigma \in \{cf, adm, com, prf, grd, stb\}$, a SETAF $F$ and its associated SETADF $D$, we have that $\sigma_{\mathscr{L}}(F)$ and $\sigma(D)$ are in one-to-one correspondence with each labelling $\mathbb{L} \in \sigma_{\mathscr{L}}(F)$ corresponding to an interpretation $v \in \sigma(D)$ such that $v(s) = \mathbf{t}$ iff $\lambda(s) = \texttt{in}$, $v(s) = \mathbf{f}$ iff $\lambda(s) = \texttt{out}$, and $v(s) = \mathbf{u}$ iff $\lambda(s) = \texttt{undec}$.*

Notice that by the above theorem we have that the 3-valued SETAF semantics introduced in (Linsbichler et al., 2016) coincide with the 3-valued labelling based SETAF semantics of (Flouris and Bikakis, 2019) and the model semantics of (Linsbichler et al., 2016) corresponds to the stable semantics of (Flouris and Bikakis, 2019).

## 9.3 3-valued Signatures of SETAFs

We adapt the concept of signatures (Dunne et al., 2015) towards our needs first.

**Definition 9.4** *The signature of SETAFs under a labelling-based semantics $\sigma_{\mathscr{L}}$ is defined as $\Sigma^{\sigma_{\mathscr{L}}}_{SETAF} = \{\sigma_{\mathscr{L}}(F) | F \in SETAF\}$. The signature of an ADF-subclass $\mathscr{C}$ under a semantics $\sigma$ is defined as $\Sigma^{\sigma}_{\mathscr{C}} = \{\sigma(D) \mid D \in \mathscr{C}\}$.*

By Theorem 9.3 we can use labellings of SETAFs and interpretations of the SETADF class of ADFs interchangeably, yielding that $\Sigma^{\sigma_{\mathscr{L}}}_{SETAF} \equiv \Sigma^{\sigma}_{SETADF}$, i.e. the 3-valued signatures of SETAFs and SETADFs only differ in the naming of the labels. For convenience, we will use the SETAF terminology in this section.

**Proposition 9.5** *The signature $\Sigma^{stb_{\mathscr{L}}}_{SETAF}$ is given by all sets $\mathbb{L}$ of labellings such that*

1. *all $\lambda \in \mathbb{L}$ have the same domain $\textsc{Args}_{\mathbb{L}}$; $\lambda(s) \neq \texttt{undec}$ for all $\lambda \in \mathbb{L}$, $s \in \textsc{Args}_{\mathbb{L}}$.*
2. *If $\lambda \in \mathbb{L}$ assigns one argument to* `out` *then it also assigns an argument to* `in`*.*

3. *For arbitrary* $\lambda_1, \lambda_2 \in \mathbb{L}$ *with* $\lambda_1 \neq \lambda_2$ *there is an argument* $a$ *such that* $\lambda_1(a) = \texttt{in}$ *and* $\lambda_2(a) = \texttt{out}$.

**Proof** We first show that for each SETAF $F$ the set $stb_{\mathscr{L}}(F)$ satisfies the conditions of the proposition. First clearly all $\lambda \in stb_{\mathscr{L}}(F)$ have the same domain and by the definition of stable semantics do not assign `undec` to any argument. That is the first condition is satisfied. For Condition (2), towards a contradiction assume that the domain is non-empty and $\lambda \in stb_{\mathscr{L}}(F)$ assigns all arguments to `out`. Consider an arbitrary argument $a$. By definition of stable semantics $a$ is only labeled `out` if there is an attack $(B, a)$ such that all arguments in $B$ are labeled in `in`, a contradiction. Thus we obtain that there is at least one argument $a$ with $\lambda(a) = \texttt{in}$. For Condition (3), towards a contradiction assume that for all arguments $a$ with $\lambda_1(a) = \texttt{in}$ also $\lambda_2(a) = \texttt{in}$ holds. As $\lambda_1 \neq \lambda_2$ there is an $a$ with $\lambda_2(a) = \texttt{in}$ and $\lambda_1(a) = \texttt{out}$. That is, there is an attack $(B, a)$ such that $\lambda_1(b) = \texttt{in}$ for all $b \in B$. But then also $\lambda_2(b) = \texttt{in}$ for all $b \in B$ and by $\lambda_2(a) = \texttt{in}$ we obtain that $\lambda_2 \notin cf_{\mathscr{L}}(F)$, a contradiction.

Now assume that $\mathbb{L}$ satisfies all the conditions. We give a SETAF $F_{\mathbb{L}} = (A_{\mathbb{L}}, R_{\mathbb{L}})$ with $A_{\mathbb{L}} = \textsc{Args}_{\mathbb{L}}$ and $R_{\mathbb{L}} = \{(\lambda_{\texttt{in}}, a) \mid \lambda \in \mathbb{L}, \lambda(a) = \texttt{out}\}$. We show that $stb_{\mathscr{L}}(F_{\mathbb{L}}) = \mathbb{L}$.

To this end we first show $stb_{\mathscr{L}}(F_{\mathbb{L}}) \supseteq \mathbb{L}$. Consider an arbitrary $\lambda \in \mathbb{L}$: By Condition (1) there is no $a \in \textsc{Args}_{\mathbb{L}}$ with $\lambda(a) = \texttt{undec}$ and it only remains to show $\lambda \in cf_{\mathscr{L}}(F_{\mathbb{L}})$. First, if $\lambda(a) = \texttt{out}$ for some argument $a$ then by construction of $R_{\mathbb{L}}$ and Condition (2) we have an attack $(\lambda_{\texttt{in}}, a)$ and thus $a$ is legally labeled `out`. Now towards a contradiction assume there is a conflict $(B, a)$ such that $B \cup \{a\} \subseteq \lambda_{\texttt{in}}$. Then, by construction of $R_{\mathbb{L}}$ there is a $\lambda' \in \mathbb{L}$ with $\lambda'_{\texttt{in}} = B$ and $\lambda_{\texttt{in}} \neq B$ (as $a \in \lambda_{\texttt{in}}$). That is, $\lambda'_{\texttt{in}} \subset \lambda_{\texttt{in}}$, a contradiction to Condition (3). Thus, $\lambda \in cf_{\mathscr{L}}(F_{\mathbb{L}})$ and therefore $\lambda \in stb_{\mathscr{L}}(F_{\mathbb{L}})$.

To show $stb_{\mathscr{L}}(F_{\mathbb{L}}) \subseteq \mathbb{L}$, consider $\lambda \in stb_{\mathscr{L}}(F_{\mathbb{L}})$. If $\lambda$ maps all arguments to `in` then there is no attack in $R_{\mathbb{L}}$ which means that $\mathbb{L}$ contains only the labelling $\lambda$. Thus, we assume that there is $a$ with $\lambda(a) = \texttt{out}$ and there is $(B, a) \in R_{\mathbb{L}}$ with $B \subseteq \lambda_{\texttt{in}}$. By construction there is $\lambda' \in \mathbb{L}$ such that $\lambda'_{\texttt{in}} = B$. Then by construction we have $(B, c) \in R_{\mathbb{L}}$ for all $c \notin B$ and thus $\lambda'_{\texttt{in}} = B = \lambda_{\texttt{in}}$ and moreover $\lambda'_{\texttt{out}} = \lambda_{\texttt{out}}$ and thus $\lambda = \lambda'$. □

We now turn to the signature for preferred semantics. Compared to the conditions for stable semantics, labelling may now assign `undec` to arguments. Note that stable is the only semantics allowing for an empty labelling set.

**Proposition 9.6** *The signature $\Sigma^{prf_{\mathscr{L}}}_{SETAF}$ is given by all non-empty sets $\mathbb{L}$ of labellings s.t.*

1. *all labellings $\lambda \in \mathbb{L}$ have the same domain $\textsc{Args}_{\mathbb{L}}$.*
2. *If $\lambda \in \mathbb{L}$ assigns one argument to* `out` *then it also assigns an argument to* `in`*.*
3. *For arbitrary $\lambda_1, \lambda_2 \in \mathbb{L}$ with $\lambda_1 \neq \lambda_2$ there is an argument $a$ such $\lambda_1(a) =$* `in` *and $\lambda_2(a) =$* `out`*.*

**Proof** [Proof sketch] We first show that for each SETAF $F$ the set $prf_{\mathscr{L}}(F)$ satisfies the conditions of the proposition. The first condition is satisfied as all $\lambda \in prf_{\mathscr{L}}(F)$ have the same domain. The second condition is satisfied by the definition of conflict-free labellings. Condition (3) is by the $\subseteq$-maximality of $\lambda_{\texttt{in}}$ which implies that there is a conflict between each two preferred extensions.

Now assume that $\mathbb{L}$ satisfies all the conditions. We give a SETAF $F_{\mathbb{L}} = (A_{\mathbb{L}}, R_{\mathbb{L}})$ with $A_{\mathbb{L}} = \textsc{Args}_{\mathbb{L}}$ and $R_{\mathbb{L}} = \{(\lambda_{\texttt{in}}, a) \mid \lambda \in \mathbb{L}, \lambda(a) = \texttt{out}\} \cup \{(\lambda_{\texttt{in}} \cup \{a\}, a) \mid \lambda \in \mathbb{L}, \lambda(a) = \texttt{undec}\}$. It remains to show that $prf_{\mathscr{L}}(F_{\mathbb{L}}) = \mathbb{L}$. To show $prf_{\mathscr{L}}(F_{\mathbb{L}}) \supseteq \mathbb{L}$, consider an arbitrary $\lambda \in \mathbb{L}$. $\lambda \in cf_{\mathscr{L}}(F_{\mathbb{L}})$ can be seen by construction, and $\lambda \in adm_{\mathscr{L}}(F_{\mathbb{L}})$ since argument labelled out is attacked by $\lambda$; finally $\lambda \in prf_{\mathscr{L}}(F_{\mathbb{L}})$ is guaranteed since the arguments $a$ with $\lambda(a) =$ `undec` are involved in self-attacks. To show $prf_{\mathscr{L}}(F_{\mathbb{L}}) \subseteq \mathbb{L}$ consider $\lambda \in prf_{\mathscr{L}}(F_{\mathbb{L}})$. It can be checked that $\lambda$ satisfies all the conditions of the proposition. □

**Proposition 9.7** *The signature $\Sigma^{cf_{\mathscr{L}}}_{SETAF}$ is given by all non-empty sets $\mathbb{L}$ of labellings s.t.*

1. *all $\lambda \in \mathbb{L}$ have the same domain $\textsc{Args}_{\mathbb{L}}$.*
2. *If $\lambda \in \mathbb{L}$ assigns one argument to* `out` *then it also assigns an argument to* `in`*.*
3. *For $\lambda \in \mathbb{L}$ and $C \subseteq \lambda_{\texttt{in}}$ also $(C, \emptyset, \textsc{Args}_{\mathbb{L}} \setminus C) \in \mathbb{L}$.*
4. *For $\lambda \in \mathbb{L}$ and $C \subseteq \lambda_{\texttt{out}}$ also $(\lambda_{\texttt{in}}, \lambda_{\texttt{out}} \setminus C, \lambda_{\texttt{undec}} \cup C) \in \mathbb{L}$.*
5. *For $\lambda, \lambda' \in \mathbb{L}$ with $\lambda_{\texttt{in}} \subseteq \lambda'_{\texttt{in}}$ also $(\lambda'_{\texttt{in}}, \lambda_{\texttt{out}} \cup \lambda'_{\texttt{out}}, \lambda_{\texttt{undec}} \cap \lambda'_{\texttt{undec}}) \in \mathbb{L}$.*
6. *For $\lambda, \lambda' \in \mathbb{L}$ and $C \subseteq \lambda_{\texttt{out}}$ (s.t. $C \neq \emptyset$) we have $\lambda_{\texttt{in}} \cup C \not\subseteq \lambda'_{\texttt{in}}$.*

**Proof** [Proof sketch] Let $F$ be an arbitrary SETAF we show that $cf_{\mathscr{L}}(F)$ satisfies the conditions of the proposition. The first two conditions are

clearly satisfied by the definition of conflict-free labelling. For Condition (3), towards a contradiction assume that $(C, \emptyset, \textsc{Args}_{\mathbb{L}} \setminus C)$ is not conflict-free. Then there is an attack $(B, a)$ such that $B \cup \{a\} \subseteq C \subseteq \lambda_{\texttt{in}}$, and thus $\lambda \notin cf_{\mathscr{L}}(F)$, a contradiction. Condition (4) is satisfied as in the definition of conflict-free labellings there are no conditions for labeling an argument `undec`. Further, the conditions that allow to label an argument `out` solely depend on the `in` labeled arguments. For Condition (5), consider $\lambda, \lambda' \in cf_{\mathscr{L}}(F)$ with $\lambda_{\texttt{in}} \subseteq \lambda'_{\texttt{in}}$ and $\lambda^* = (\lambda'_{\texttt{in}}, \lambda_{\texttt{out}} \cup \lambda'_{\texttt{out}}, \lambda_{\texttt{undec}} \cap \lambda'_{\texttt{undec}})$. Since $\lambda, \lambda' \in \mathbb{L}$, it is easy to check that $\lambda^*$ is a well-founded labelling and $\lambda^* \in cf_{\mathscr{L}}(F)$. For Condition (6), consider $\lambda, \lambda' \in cf_{\mathscr{L}}(F)$ and a set $C \subseteq \lambda_{\texttt{out}}$ containing an argument $a$ such that $\lambda(a) = \texttt{out}$. That is, there is an attack $(B, a)$ with $B \subseteq \lambda_{\texttt{in}}$ and thus $\lambda_{\texttt{in}} \cup C \not\subseteq \lambda'_{\texttt{in}}$. That is, Condition (6) is satisfied.

Now assume that $\mathbb{L}$ satisfies all the conditions. We give a SETAF $F_{\mathbb{L}} = (A_{\mathbb{L}}, R_{\mathbb{L}})$ with $A_{\mathbb{L}} = \textsc{Args}_{\mathbb{L}}$ and $R_{\mathbb{L}} = \{(\lambda_{\texttt{in}}, a) \mid \lambda \in \mathbb{L}, \lambda(a) = \texttt{out}\} \cup \{(B, b) \mid b \in B, \nexists \lambda \in \mathbb{L} : \lambda_{\texttt{in}} = B\}$. To complete the proof it remains to show that $cf_{\mathscr{L}}(F_{\mathbb{L}}) = \mathbb{L}$. □

Finally, we give an exact characterisation of the signature of grounded semantics.

**Proposition 9.8** *The signature $\Sigma^{grd_{\mathscr{L}}}_{SETAF}$ is given by sets $\mathbb{L}$ of labellings such that $|\mathbb{L}| = 1$, and if $\lambda \in \mathbb{L}$ assigns one argument to* `out` *then $\lambda_{\texttt{in}} \neq \emptyset$.*

Notice that Proposition 9.8 basically exploits that grounded semantics is a unique status semantics based on admissibility. The result thus immediately extends to other semantics satisfying these two properties, e.g. to ideal or eager semantics (Flouris and Bikakis, 2019).

So far, we have provided characterisations for the signatures $\Sigma^{stb_{\mathscr{L}}}_{\text{SETAF}}$, $\Sigma^{prf_{\mathscr{L}}}_{\text{SETAF}}$, $\Sigma^{cf_{\mathscr{L}}}_{\text{SETAF}}$, $\Sigma^{grd_{\mathscr{L}}}_{\text{SETAF}}$. By Theorem 9.3 we get analogous characterizations of $\Sigma^{\sigma}_{\text{SETADF}}$ for the corresponding ADF semantics.

We have not yet touched admissible and complete semantics. Here, the exact characterisations seem to be more cumbersome and are left for future work. However, for admissible semantics the following proposition provides necessary conditions for an labelling-set to be *adm*-realizable, but it remains open whether they are also sufficient.

**Proposition 9.9** *For each $\mathbb{L} \in \Sigma^{adm_{\mathscr{L}}}_{SETAF}$ we have:*

1. *all $\lambda \in \mathbb{L}$ have the same domain $\textsc{Args}_{\mathbb{L}}$.*

2. *If* $\lambda \in \mathbb{L}$ *assigns one argument to* $\mathtt{out}$ *then it also assigns an argument to* $\mathtt{in}$.
3. *For* $\lambda, \lambda' \in \mathbb{L}$ *and* $C \subseteq \lambda_{\mathtt{out}}$ *(s.t.* $C \neq \emptyset$*) we have* $\lambda_{\mathtt{in}} \cup C \not\subseteq \lambda'_{\mathtt{in}}$.
4. *For arbitrary* $\lambda, \lambda' \in \mathbb{L}$ *either (a)* $(\lambda_{\mathtt{in}} \cup \lambda'_{\mathtt{in}}, \lambda_{\mathtt{out}} \cup \lambda'_{\mathtt{out}}, \lambda_{\mathtt{undec}} \cap \lambda'_{\mathtt{undec}}) \in \mathbb{L}$ *or (b) there is an argument* $a$ *such* $\lambda(a) = \mathtt{in}$ *and* $\lambda'(a) = \mathtt{out}$.
5. *For* $\lambda, \lambda' \in \mathbb{L}$ *with* $\lambda_{\mathtt{out}} \subseteq \lambda'_{\mathtt{out}}$*, and* $C \subseteq \lambda_{\mathtt{in}} \setminus \bigcup_{\lambda^* \in \mathbb{L}:\ \lambda^*_{\mathtt{in}} = \lambda'_{\mathtt{in}}} \lambda^*_{\mathtt{out}}$ *we have* $(\lambda'_{\mathtt{in}} \cup C, \lambda'_{\mathtt{out}}, \lambda'_{\mathtt{undec}} \setminus C) \in \mathbb{L}$.
6. *For* $\lambda, \lambda' \in \mathbb{L}$ *with* $\lambda_{\mathtt{in}} \subseteq \lambda'_{\mathtt{in}}$*, and* $C \subseteq \lambda_{\mathtt{out}}$ *we have* $(\lambda'_{\mathtt{in}}, \lambda'_{\mathtt{out}} \cup C, \lambda'_{\mathtt{undec}} \setminus C) \in \mathbb{L}$.
7. *For* $\lambda, \lambda' \in \mathbb{L}$ *with* $\lambda_{\mathtt{in}} \subseteq \lambda'_{\mathtt{in}}$ *and* $\lambda_{\mathtt{out}} \supseteq \lambda'_{\mathtt{out}}$ *we have* $(\lambda_{\mathtt{in}}, \lambda'_{\mathtt{out}}, \textsc{Args}_{\mathbb{L}} \setminus (\lambda_{\mathtt{in}} \cup \lambda'_{\mathtt{out}})) \in \mathbb{L}$.
8. $(\emptyset, \emptyset, \textsc{Args}_{\mathbb{L}}) \in \mathbb{L}$.

**Proof** We show that for each SETAF $F$ the set $adm_{\mathscr{L}}(F)$ satisfies the conditions of the proposition. Conditions (1)–(3) are by the fact that $adm_{\mathscr{L}}(F) \subseteq cf_{\mathscr{L}}(F)$. For Condition (4), let $\lambda, \lambda' \in adm_{\mathscr{L}}(F)$ with $\lambda_{\mathtt{in}} \cap \lambda'_{out} = \{\}$ (since each admissible labelling defends itself, $\lambda'_{\mathtt{in}} \cap \lambda_{out} = \{\}$). Thus, $\lambda^* = (\lambda_{\mathtt{in}} \cup \lambda'_{\mathtt{in}}, \lambda_{\mathtt{out}} \cup \lambda'_{\mathtt{out}}, \lambda_{\mathtt{undec}} \cap \lambda'_{\mathtt{undec}})$ is a well-defined labelling. Further, since $\lambda, \lambda' \in adm_{\mathscr{L}}(F)$ it is easy to check that $\lambda^* \in adm_{\mathscr{L}}(F)$.

For Condition (5), let $\lambda^* = (\lambda'_{\mathtt{in}} \cup C, \lambda'_{\mathtt{out}}, \lambda'_{\mathtt{undec}} \setminus C)$. First, $\lambda^*$ is a well-defined labelling. Notice that the set $C$ contains arguments defended by $\lambda$ and not attacked by $\lambda'_{\mathtt{in}}$. Now, it is easy to check that $\lambda^*$ meets the condition for being an admissible labelling. For Condition (6), let $\lambda^* = (\lambda'_{\mathtt{in}}, \lambda'_{\mathtt{out}} \cup C, \lambda'_{\mathtt{undec}} \setminus C)$. Notice that the set $C$ contains only arguments attacked by $\lambda_{\mathtt{in}}$ and thus are also attacked by $\lambda'_{\mathtt{in}}$. Thus, starting from the admissible labelling $\lambda'$ we can relabel arguments in $C$ to $\mathtt{out}$ and obtain that $\lambda^*$ is also an admissible labelling.

For Condition (7), let $\lambda^* = (\lambda_{\mathtt{in}}, \lambda'_{\mathtt{out}}, \textsc{Args}_{\mathbb{L}} \setminus (\lambda_{\mathtt{in}} \cup \lambda'_{\mathtt{out}}))$. First, $\lambda^*$ is a well-defined labelling. We have that setting $\lambda'_{\mathtt{out}}$ to $\mathtt{out}$ is sufficient to make all the $\mathtt{in}$ labels for arguments in $\lambda'_{\mathtt{in}}$ valid and thus are also sufficient to make the $\mathtt{in}$ labels for arguments $\lambda_{\mathtt{in}} \subseteq \lambda'_{\mathtt{in}}$ valid. Moreover, as $\lambda_{\mathtt{out}} \supseteq \lambda'_{\mathtt{out}}$ also labelling arguments $\lambda_{\mathtt{in}}$ with $\mathtt{in}$ is sufficient to make the $\mathtt{out}$ labels for $\lambda'_{\mathtt{out}}$ valid. Hence, $\lambda^*$ is admissible.

For Condition (8), the conditions of admissible labelling for arguments labelled $\mathtt{in}$ or $\mathtt{out}$ in $(\emptyset, \emptyset, \textsc{Args}_{\mathbb{L}})$ are clearly met, since there are no such arguments.

□

## 9.4 On the Relation between SETAFs and Support-Free ADFs

In order to compare SETAFs with SFADFs, we can rely on SETADFs (recall Theorem 9.3). In particular, we will compare the signatures $\Sigma^{\sigma}_{SETADF}$ and $\Sigma^{\sigma}_{SFADF}$, cf. Definition 9.4. We start with the observation that each SETADF can be rewritten as an equivalent SETADF that is also a SFADF.[1]

**Lemma 9.10** *For each SETADF $D = (S, L, C)$ there is an equivalent SETADF $D' = (S, L', C')$ that is also a SFADF, i.e. for each $s \in S$, $\varphi_s \in C$, $\varphi'_s \in C'$ we have $\varphi_s \equiv \varphi'_s$.*

**Proof** Given a SETADF $D$, by Definition 9.1, each acceptance condition is a CNF over negative literals and thus does not have any support link which is not redundant. We can thus obtain $L'$ by removing the redundant links from $L$ and $C'$ by, in each acceptance condition, deleting the clauses that are super-sets of other clauses. □

By the above we have that $\Sigma^{\sigma}_{\text{SETADF}} \subseteq \Sigma^{\sigma}_{\text{SFADF}}$. Now consider the interpretation $v = \{a \mapsto \mathbf{f}\}$. We have that for all considered semantics $\sigma$, $v$ is a $\sigma$-interpretation of the SFADF $D = (\{a\}, \{\varphi_a = \bot\})$ but there is no SETADF with $v$ being a $\sigma$-interpretation. We thus obtain $\Sigma^{\sigma}_{\text{SETADF}} \subsetneq \Sigma^{\sigma}_{SFADF}$.

**Theorem 9.11** *$\Sigma^{\sigma}_{SETADF} \subsetneq \Sigma^{\sigma}_{SFADF}$, for $\sigma \in \{cf, adm, stb, mod, com, prf, grd\}$.*

In the remainder of this section we aim to characterise the difference between $\Sigma^{\sigma}_{\text{SETADF}}$ and $\Sigma^{\sigma}_{SFADF}$. To this end we first recall a characterisation of the acceptance conditions of SFADF that can be rewritten as collective attacks.

**Lemma 9.12** *(Wallner, 2020) Let $D = (S, L, C)$ be a SFADF. If $s \in S$ has at least one incoming link then the acceptance condition $\varphi_s$ can be written in CNF containing only negative literals.*

It remains to consider those arguments in an SFADF with no incoming links. Such arguments allow for only two acceptance conditions $\top$ and $\bot$. While condition $\top$ is unproblematic (it refers to an initial argument in a SETAF),

[1] As discussed in (Polberg, 2017), in general, SETAFs translate to bipolar ADFs that contain attacking and redundant links. However, when we first remove redundant attacks from the SETAF we obtain a SFADF.

an argument with unsatisfiable acceptance condition cannot be modeled in a SETADF. In fact, the different expressiveness of SETADFs and SFADFs is solely rooted in the capability of SFADFs to set an argument to **f** via a $\bot$ acceptance condition.

We next give a generic characterisations of the difference between $\Sigma^{\sigma}_{\text{SETADF}}$ and $\Sigma^{\sigma}_{SFADF}$.

**Theorem 9.13** *For $\sigma \in \{cf, adm, stb, mod, com, prf, grd\}$, we have $\Delta_\sigma = \Sigma^{\sigma}_{SFADF} \setminus \Sigma^{\sigma}_{SETADF}$ with*

$$\Delta_\sigma = \{\mathbb{V} \in \Sigma^{\sigma}_{SFADF} \mid \exists v \in \mathbb{V} \text{ s.t. } \forall a : v(a) \in \{\mathbf{f}, \mathbf{u}\} \wedge \exists a : v(a) = \mathbf{f}\}.$$

**Proof** [Proof sketch] If a SFADF has a $\sigma$-interpretation $v$ that assigns some arguments to **f** without assigning an argument to **t** then we have that the arguments assigned to **f** are exactly the arguments with acceptance condition $\bot$. For *stb* and *mod* semantics this means all arguments have acceptance condition $\bot$ and the result follows. Each preferred interpretation assigns arguments with acceptance condition $\bot$ to **f** and thus the existence of another preferred interpretation would violate the $\leq_i$-maximality of $v$. □

In other words each interpretation-set which is $\sigma$-realizable in SFADFs and contains at least two interpretations can be realized in SETADFs, for $\sigma \in \{stb, prf, mod\}$. We close this section with an example illustrating that the above characterisation thus not hold for *cf*, *adm*, and *com*.

**Example 9.14** *Let $D = (\{a, b, c\}, \{\varphi_a = \bot, \varphi_b = \neg c, \varphi_c = \neg b\})$. We have $com(D) = \{\{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}, \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}, \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}\}$. By Theorem 9.13, $com(D)$ cannot be realized as SETADF. Moreover, as $com(D) \subseteq adm(D) \subseteq cf(D)$ for every ADF $D$, we have that, despite all three contain more than one interpretation, none of them can be realized via a SETADF.*

## 9.5 Conclusion

In this chapter, we have characterised the expressiveness of SETAFs under 3-valued signatures. The more fine-grained notion of 3-valued signatures reveals subtle differences of the expressiveness of stable and preferred semantics which are not present in the 2-valued setting (Dvořák et al., 2019) and enabled us to compare the expressive power of SETAFs and SFADFs, a subclass of ADFs that allows only for attacking links. In particular,

we have exactly characterized the difference for conflict-free, admissible, complete, stable, preferred, and grounded semantics; this difference is rooted in the capability of SFADFs to set an initial argument to false. Together with our exact characterisations on signatures of SETAFs for stable, preferred, grounded, and conflict-free semantics, this also yields the corresponding results for SFADFs. Exact characterisations for admissible and complete semantics can be investigated as future work. Another aspect to be investigated is to which extent our insights on labelling-based semantics for SETAFs and SFADFs can help improve the performance of reasoning systems.

# Chapter 10

# Embedding Probabilities, Utilities and Decisions in a Generalization of ADFs

Life is made up of a long list of decisions. In each of them, there exists quite a number of choices and most decisions are affected by uncertainties and preferences, from choosing a healthy lunch and nice clothes to choosing a profession and a field of study. Uncertainties can be modeled by probabilities and preferences can be modeled by utilities. A rational decision maker prefers to make a decision with the least regret or the most satisfaction. The principle of maximum expected utility can be helpful in this issue. Expected utility deals with problems in which agents make a decision under conditions in which probabilities of states play a role in the choice, as well as the utilities of outcomes.

Argumentation formalisms could be an option to model these problems and to pick one or several alternatives. In this chapter, a new argument-based framework, numerical abstract dialectical frameworks (nADFs for short), is introduced to do so. This framework is a generalization of abstract dialectical frameworks (ADFs for short). First, the semantics based on many-valued interpretations are introduced, including preferred, grounded, complete, and model-based semantics. Second, it is shown how nADFs are expressive enough to formalize standard decision problems. It is shown that the different types of semantics of an nADF that is associated with a decision problem all coincide and have the standard meaning. In this way, it is shown how the nADF semantics can be used to choose the best set of decisions.

## 10.1 Introduction

During life, people are faced with a long series of decisions. A good decision may lead to a cure for a disease, to an investment in a proper project by a business person, to a judgment in a crime case, and to a fair debate. Definitely, different decisions that are made by an agent yield different consequences. At the moment of decision making, we are usually not certain of what is the consequence of our decision, but we may know the set of possible consequences that our decision can lead to. That is, we usually make decisions under uncertainty. The uncertainty mostly arises because of external factors that are out of control of agents, which are called *state*, such as needing to undergo emergency surgery.

Assume that Maryam wants to travel abroad. She wants to decide whether or not to buy an international health insurance by spending 100 euros. The decision depends on some factors. Here, the external factor is the probability of having to undergo emergency surgery abroad. For example, if Maryam had a heart attack recently, the need for health insurance abroad is higher than for healthy people. Maryam's decision leads to either: 1) buying an international insurance for 100 euros and needing it when she is abroad; 2) losing 100 euros because of buying an international health insurance and not needing it; 3) needing an emergency surgery without any insurance, that is, she has to spend at least $10,000$ euros; 4) not buying international health insurance and not needing it, that is, spending nothing. Another factor with crucial importance in making decisions is the preferences that Maryam has on different consequences, which are called *outcomes*. Maryam prefers not to spend any money for an insurance and not to undergo emergency surgery to other outcomes, however, in the case that she needs emergency surgery abroad, she prefers to spend 100 euros rather than at least $10,000$ euros.

Maryam can choose among actions (buying an international health insurance or not), but she does not have any control over the states (having to undergo emergency surgery abroad or not). However, the probability of occurrence of each state has an effect on her decision. Actually, if a state of the world can be affected by an agent, it is not a state in the sense of decision theory. An agent has control but not belief over actions, however, over states she/he has belief but no control.

A theory concerned with making the best decision under uncertainty is called expected utility theory (Von Neumann and Morgenstern, 1947; von Neumann and Morgenstern, 2007; Savage, 1954; Briggs, 2019; Gilboa, 2009; Mongin, 1998; Russell and Norvig, 2009). The expected utility of

each decision or action is the weighted average of utilities of the possible outcomes, where *utility* is a numerical measure of preference of outcomes from an agent point of view, representing the agent's desire. These utilities are weighted by the probability of the state that leads to that outcome for a specific action.

Although there exist many formalisms, solvers and automated methods in decision theory such as influence diagrams (Howard and Matheson, 2005; Olmsted, 1985; Shachter, 1986), because of the importance of decision making in human life and the wide variety of decision problems, new approaches of modeling and evaluating them are required.

Argumentation is a reasoning model that can help to select one or several alternative actions, or explain an already adopted decision. Several efforts have been put into the study and definition of argumentation formalisms within which the values or preferences of agents are of crucial importance for everyday reasoning (Amgoud and Prade, 2009; Atkinson and Bench-Capon, 2007, 2018; Bench-Capon, 2003; Dung and Thang, 2010; Hunter and Thimm, 2014; Verheij, 2016a; Vlek et al., 2016). One might wonder whether an argumentation formalism can be considered for modeling and solving decision problems. Motivated by this question, we will here introduce an argumentation formalism to represent problems in which both the probability of states and utility over outcomes play a role in making decisions. Then in future work, abstract dialectical framework (ADF) solvers can be generalized to numerical abstract dialectical frameworks (nADFs) to make a decision automatically.

The main goal of this chapter is to investigate how an argumentation formalism can accommodate a decision problem. We model scenarios with utility, using a formalism of argumentation that will allow us to compute the maximum expected utility of a problem with the help of semantics of that argumentation framework. To this end, we introduce *numerical Abstract Dialectical Frameworks* (nADFs for short), which are a generalization of abstract dialectical frameworks, introduced first in (Brewka and Woltran, 2010) and then revised in (Brewka et al., 2013, 2018a), as a generalization of Dung's argumentation frameworks (AFs) (Dung, 1995). An nADF shows how the structure of arguments can be constructed from a given knowledge base and how arguments interact with each other. In argumentation formalisms like AFs and ADFs, the area that deals with evaluating arguments is called semantics. Semantics are criteria used to select subsets of available arguments that satisfy desirable properties. We follow the same way in our work to choose the best action in the nADF

that is constructed based on a decision problem. We do not claim that our results will make decision theory computationally more efficient. The reasons why we combined decision theory with ADFs are as follows:

- Argumentation theory can shed light on the process of decision making, from modeling to evaluating a problem. ADFs are expressive formalisms in that area.
- Decision theory uses the well-known tools of probabilities and utilities, of which the relation with argumentation theory are still to be well-understood.

In nADFs as well as ADFs, each argument is associated with an acceptance condition. However, in contrast with ADFs, the language used to define acceptance conditions of nADFs is a variation of propositional logic allowing numerical calculation.

This chapter is organized as follows. In Section 10.2, we summarize the relevant background. In particular, we provide a short reminder on decision problems, expected utility theory and ADFs. In Section 10.3, the structures of numerical abstract argumentation frameworks, which are generalization of ADFs, are introduced. Semantics of nADFs are defined based on many-valued interpretation on rational numbers of the unit interval. In Section 10.4, we investigate how nADFs can be used to model decision problems, that is, how an nADF can be constructed from a given decision problem. Then, we show that in the constructed nADF all semantics collapse to the same set of interpretations. Moreover, it is shown how this unique set of interpretations can be used to choose the best action. Finally, in Section 10.5 we will summarize and conclude the presented results and refer to the open questions we would like to address next. Moreover, we compare nADFs with ADFs and their generalization called weighted ADFs (Brewka et al., 2018b).

## 10.2 Background

In this section, we summarize decision problems, expected utility theory, and abstract dialectical frameworks.

### 10.2.1 Decision Problems

Decision making under uncertainty infuses the life of every decision maker, which can be an individual, an organization or a society. To say that a

| state $\rightarrow$ | $s_1$ | $s_2$ | $\cdots$ | $s_n$ |
|---|---|---|---|---|
| act $\downarrow$ | | | | |
| $a_1$ | $o_{11}$ | $o_{12}$ | $\cdots$ | $o_{1n}$ |
| $\vdots$ | | | | |
| $a_m$ | $o_{m1}$ | $o_{m2}$ | $\cdots$ | $o_{mn}$ |

**Figure 10.1:** The table of a decision-making problem

decision is made by a decision maker, called an *agent*, means that an action among the set of *actions* $A$ is chosen to be done. Uncertainty in decision making means that an available action may lead to the set of *outcomes* $O$. The outcome of each decision is also influenced by some external factors which are called *states* $S$. Following the example introduced in the introduction, Maryam can choose whether to buy a health insurance. The consequences of her decision depend on whether she gets emergency surgery abroad. That is, Maryam's decision depends on the probability of getting an emergency surgery. Beyond the probability of states, Maryam's decision depends on her preferences on the consequences. For instance, she prefers not to buy a health insurance and not to get a surgery to other consequences. However, she prefers to spend 100 euros to buy a health insurance rather than to spend at least $10,000$ euros to get emergency surgery. The basic model of decision under uncertainty is a table or matrix in which the columns are labeled with states and the rows are labeled with actions and the consequence of picking an action in each state is an outcome, as depicted in Figure 10.1.

The notation $o_1 \succ_p o_2$ means an agent strictly prefers $o_1$ to $o_2$, $o_1 \sim_p o_2$ means $o_1$ and $o_2$ are equally preferred by an agent or an agent is indifferent between $o_1$ and $o_2$, and $o_1 \succeq_p o_2$ means $o_1$ is preferred at least as much as is $o_2$. The preference relation $\succeq_p$ over the set of outcomes is called *rational* iff it is transitive and complete. The technical name for the value of a possible outcome is *utility*. In (Bentham, 1961; Sidgwick, 1981), utility is interpreted as a measure of pleasure or happiness. Contemporary decision theorists typically interpret utility as a measure of preference (Von Neumann and Morgenstern, 1947; von Neumann and Morgenstern, 2007; Sen, 1977). That is, it is not the case that an agent prefers outcome $o_1$ over $o_2$ because $o_1$ generates a higher utility than $o_2$. But for an agent, $o_1$ has a higher utility than $o_2$ because she/he prefers $o_1$ to $o_2$.

**Definition 10.1** *Given $\succeq_p$, a rational order over the finite set of outcomes $O$. A function $u : O \to \mathbb{R}$ is called a utility function that represents $\succeq_p$ if, for every two outcomes $o_1$ and $o_2$, $u(o_1) \geq u(o_2)$ iff $o_1 \succeq_p o_2$.*

In Cantor's result characterizing dense order, dating from around 1895, it is shown that a binary relation $\succeq_p$ over a finite set can be represented by a real-valued function $u$ if and only if $\succeq_p$ is a rational order. Note that in the current chapter, utility functions are defined over $\mathbb{Q}$, in which $\mathbb{Q}$ denotes the set of rational numbers. A decision problem is formally defined in Definition 10.2.

**Definition 10.2** *A decision problem is a tuple $(A, S, O, p, u)$ where:*

- *$A$ is a finite set of actions that can be chosen by an agent;*
- *$S$ is a finite set of states;*
- *$O$ is a finite set of outcomes;*
- *$p$ is a probability function on states, namely, $p : S \to [0, 1]$ such that $\Sigma_{s \in S}\, p(s) = 1$;*
- *$u$ is a utility function on outcomes, namely, $u : O \to \mathbb{Q}$.*

The criterion that deals with the analysis of situations where individuals must make a decision without knowing which outcomes may result from that decision (act) is called expected utility, which was first introduced by Daniel Bernoulli in his work on a paradox of probability (Arrow, 1974). Expected utility theory (EUT) states that a decision maker chooses among actions $A$ under uncertainty by comparing the expected utility (Gilboa, 2009; Mongin, 1998; von Neumann and Morgenstern, 2007) of each action computed as the sum of the utilities of outcomes which are weighted by states respective probability. EUT is a standard theory of individual choice under uncertainty. Expected utility theory says that the higher the expected utility of an action is, the better it is to be chosen. The expected utility of each action $a \in A$ depends on two features of the problem: The value of each outcome $o \in O$ forms an agent's standpoint, the utility of an outcome, $u(o)$; and the probability of each outcome conditional on $a$, represented by $p_a(o)$.

In expected utility theory, probability can be interpreted as a subjective estimate by the individual or as objectively obtained from relevant (past) data. The former is a measure of individual degrees of belief as described in (Ramsey, 2016; Savage, 1954). However, probability can be

interpreted as an objective chance as in (von Neumann and Morgenstern, 2007; Von Neumann and Morgenstern, 1947). In the current work, the interpretation of probability introduced by (Savage, 1954) is used, in which $p_a(o)$ is calculated by summing the probabilities of states that, when combined by the action $a$, lead to the outcome $o$. To present $p_a(o)$ formally, let $\chi_{a,s}(o)$ be a function on $O$ defined as follows:

$$\chi_{a,s}(o) = \begin{cases} 1 & \text{if } o \text{ results from performing action of } a \text{ in state } s, \\ 0 & \text{otherwise} \end{cases}$$

Then $p_a(o) = \Sigma_{s \in S}\, p(s)\chi_{a,s}(o)$, where $p(s)$ is the probability of occurring of state $s$.

**Definition 10.3** *Let $A$ be a set of actions that could be chosen by an agent, $S$ a set of states, and $O$ a set of outcomes. The expected utility of $a \in A$ is defined as:*

$$EU(a) = \Sigma_{o \in O}\, p_a(o)u(o)$$

The principle of maximum expected utility (MEU) says that a rational agent should choose the action that belongs to the set of actions with maximum expected utility. An action $a$ belongs to the set of *maximum expected utility* if for each $a' \in A$, $EU(a) \geq EU(a')$.

## 10.3 Numerical Abstract Dialectical Frameworks

In many argumentation situations, it is natural to assume $n$-valued acceptance degrees of arguments, for $n > 3$. For instance, if one wants to investigate in a given semantics (preferred, complete, ...) whether the probability of a state is below or above some threshold in an interpretation.

In this section, we introduce a modification of ADFs called numerical abstract dialectical frameworks (nADFs). nADFs enhance ADFs by allowing numerical acceptance conditions of arguments and arithmetical computations among them. The logic used to define the acceptance conditions of arguments in nADFs is a variation of propositional logic, defined in Definition 10.4.

**Definition 10.4** *This logic contains:*

- *a countably infinite number of propositional variables: $x_1, x_2, \ldots$;*

- *a countably infinite number of constants which are called propositional atoms:* $a, b, s, \ldots$;
- *truth constants:* $\top$ *and* $\bot$;
- *the connectives of propositional logic:* $\wedge, \vee, \neg$;
- *binary function symbols:* $\oplus$ *and* $\otimes$;
- *a binary predicate symbol* $\succeq$ *that takes entities in the domain of discourse as input while outputs are either* 1 *(True),* 0 *(False), or* $\mathbf{u}$ *(unknown or undecided).*
- *the truth-functional operator of* $\bar{\bigwedge}$ *with the output of either* 1 *(True) or* 0 *(False).*

*The set of terms* $\{t_1, t_2, \ldots\}$ *is inductively defined by the following rules:*

- *any variable and any propositional atom is a term;*
- *applying of each binary function of the language on two terms* $t_1$ *and* $t_2$ *also results in a term, for instance,* $t_1 \otimes t_2$;
- *nothing else is a term.*

*The set of formulas is inductively defined by the following rules:*

- *any formula of propositional logic is also a formula;*
- *for arbitrary terms* $t_1$ *and* $t_2$, $t_1 \succeq t_2$ *is a formula;*
- *nothing else is a formula.*

Note that interpretations of the connectives and function symbols of Definition 10.4 are given below in Section 10.3.1. An nADF is introduced in Definition 10.5.

**Definition 10.5** *Let* $V$ *be* $\mathbb{Q}$. *An nADF is a tuple* $U = (N, L, C, i)$ *in which the following hold:*

- $N$ *is a finite set of nodes;*
- $L \subseteq N \times N$ *is a set of links;*
- $C = \{C_n\}_{n \in N}$ *is a collection of total functions called acceptance conditions over* $V$, *that is,* $C_n : (par(n) \to V) \to V$, *where* $par(n) = \{a \mid (a, n) \in L\}$;

- *$i$ is a function called input function, namely, $i : N' \to V$ where $N' \subseteq N$.*

Note that this definition is a generalization of Definition 2.40 of ADFs. In the current work, the $C_n$ correspond to formulas of the language introduced in Definition 10.4 indicated by $\varphi_n$. Note that the set of links $L$ is also implicitly determined by the acceptance conditions.

An nADF, just like an ADF, is a directed graph in which nodes indicate arguments or statements and links represent relations between statements. Each node $n$ has an attached formula, denoted by $\varphi_n$, of the logical language introduced in Definition 10.4, which is a language of propositional logic with new binary functions: $\oplus$ used for the plus function and $\otimes$ used for the times function, plus a binary relation $\succeq$ used for the preference relation.

In Definition 10.5, $i$ is a partial function on nodes; however, $i(n)$ does not appear in the acceptance conditions. It is used to indicate the input value of $n$ and $i(n)$ is called input value of $n$. Input function $i$ is used in the computation of semantics of nADFs. In our setting, $i(n)$ will be used to represent the probabilities of states and the utilities of outcomes. In general, if $i(n)$ is defined in an nADF, this does not mean that the degree of acceptance of $\varphi_n$ is $i(n)$ or the initial value of $n$ is $i(n)$, but the input value that is considered for $n$ is $i(n)$. For instance, an atom $n$ can be used to represent the number of heart-beats per minute and $i(n)$ indicates the normal number of heart-beats per minute. The input value of normal heart-beats $i(n)$ can be compared with a person's number of heart-beats $n$ or can be used in an equation to decide whether a person's heart beats normally, but it does not mean that $n$ is assigned the value $i(n)$.

Example 10.6 is an abstract nADF with three arguments. Their values in an interpretation are computed in Example 10.15. In Section 10.4, we give a concrete example in terms of decision making.

**Example 10.6** *Let $U = (\{a, b, c\}, L, \{\varphi_a, \varphi_b, \varphi_c\}, \{i(b) = 1/5, i(c) = 4/5\})$ be an nADF in which $\varphi_a = a$, $\varphi_b = b \vee a$, $\varphi_c = (a \otimes c) \succeq b$, depicted in Figure 10.2. In this nADF, function $i$ is defined on $b$ and $c$ and this means that the input value of $b$ is $1/5$ and the input value of $c$ is $4/5$. The acceptance condition of $a$ says that the degree of acceptance of $a$ depends only on $a$. The acceptance condition of $b$ says the degree of acceptance of $b$ depends on the degree of acceptance of $b$ and $a$. The acceptance condition of $c$ is composed from the predicate $\succeq$ on the terms $a \otimes c$ and $b$.*

nADFs are also used to answer queries, for instance, in Example 10.6, an nADF can be used to clarify for which amount of $a$ the acceptance

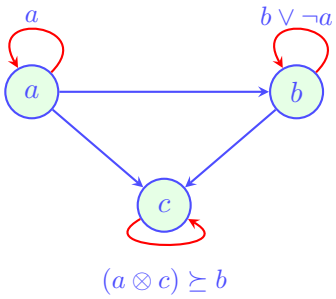


**Figure 10.2:** nADF of Example10.6

condition of $c$ has truth value 1 (true). The computation of acceptance degrees of nodes is introduced in Section 10.3.1.

### 10.3.1 Semantics of nADFs

Semantics of nADFs indicate the degree of acceptance of each argument and they are introduced based on many-valued interpretations given below.

**Definition 10.7** *A many-valued interpretation $v$ for an nADF $U$ is a function mapping each argument to a rational number or to undecided* ($\mathbf{u}$)*, namely, $v : N \to V_{\mathbf{u}}$, where $V_{\mathbf{u}} = \mathbb{Q} \cup \{\mathbf{u}\}$.*

The definition is a generalization of three-valued interpretations of ADFs (Definition 2.1). That is, an interpretation assigns a rational number or $\mathbf{u}$ to the nodes of an nADF. The intuition of $\mathbf{u}$ is that an argument is unknown (undecided). Any rational number assigned to an argument shows the degree of acceptance. Interpretations can be extended to assign a degree of acceptance to each acceptance condition. The evaluation of the acceptance condition of each argument $n$ under a given interpretation $v$ is a partial evaluation of $\varphi_n$ under $i$-correction $\mathbf{v}$ introduced in Definition 10.8.

**Definition 10.8** *Let $U = (N, L, C, i)$ be an nADF and let $v$ be a many-valued interpretation. The $i$-correction of $v$ under $U$ denoted by $\mathbf{v}$ is defined as $\mathbf{v}(n) = i(n)$ if $i$ is defined on $n \in N$ in $U$ and $\mathbf{v}(n) = v(n)$ otherwise.*

The evaluation of non-standard connectives, functions and the predicate $\succeq$ under the $i$-correction of a given interpretation $v$ in a given nADF $U$ is as follows.

**Definition 10.9** *Given an nADF $U = (N, L, C, i)$, let $v$ be a many-valued interpretation. The partial evaluation of acceptance conditions under $v$ is*

*defined inductively as follows, in which* $\mathbf{v}$ *is the* $i$*-correction of* $v$ *under* $U$*, lowercase* $a$*,* $b$ *are propositional atoms, uppercase* $A$*,* $B$ *are formulas, and* $t_1$*,* $t_2$*,* $t_i$ *are terms and* $I$ *is a finite set of natural numbers.*

$$\mathbf{v}(A \wedge B) := min\{\mathbf{v}(A), \mathbf{v}(B)\},$$
$$\mathbf{v}(A \vee B) := max\{\mathbf{v}(A), \mathbf{v}(B)\},$$
$$\mathbf{v}(a \otimes b) := \mathbf{v}(a) \times \mathbf{v}(b),$$
$$\mathbf{v}(a \oplus b) := \mathbf{v}(a) + \mathbf{v}(b),$$
$$\mathbf{v}(\bar{\bigwedge}_{i \in I} t_i) := \begin{cases} 1 & \text{if for each } i\ \mathbf{v}(t_i) = 1, \\ 0 & \text{otherwise.} \end{cases}$$
$$\mathbf{v}(t_1 \succeq t_2) := \begin{cases} 1 & \text{if } \mathbf{v}(t_1), \mathbf{v}(t_2) \in \mathbb{Q} \text{ and } \mathbf{v}(t_1) \geq \mathbf{v}(t_2), \\ 0 & \text{if } \mathbf{v}(t_1), \mathbf{v}(t_2) \in \mathbb{Q} \text{ and } \mathbf{v}(t_1) < \mathbf{v}(t_2), \\ \mathbf{u} & \text{if either } \mathbf{v}(t_1) \text{ or } \mathbf{v}(t_2) \text{ is undecided.} \end{cases}$$

Here, multiplication $\times$ on rational numbers is the standard multiplication. Moreover, $\mathbf{u} \times 0 = 0 \times \mathbf{u} = 0$ and $\mathbf{u} \times n = n \times \mathbf{u} = \mathbf{u}$ for $n \neq 0$. Also, $+$ and $-$ on rational numbers are the standard addition and subtraction, respectively, such that $n - \mathbf{u} = \mathbf{u} - n = \mathbf{u} + n = n + \mathbf{u} = \mathbf{u}$ for $n \in \mathbb{Q}$. Finally, $\mathbf{v}(A \vee B)$ and $\mathbf{v}(A \wedge B)$ are $\mathbf{u}$ if either $\mathbf{v}(A)$ or $\mathbf{v}(B)$ is $\mathbf{u}$.

The set of all many-valued interpretations over $N$ is denoted by $\mathscr{V}$, i.e., $\mathscr{V} = \{v \mid v : N \to V_{\mathbf{u}}\}$. Interpretations can be ordered by the ordering $<_i$ which assigns a greater value to the rational numbers than to $\mathbf{u}$, that is, $\mathbf{u} <_i x$ for $x \in \mathbb{Q}$. The reflexive closure of $<_i$ is $\leq_i$, i.e., $\mathbf{u} \leq_i \mathbf{u}$ and $x \leq_i x$ for each $x \in \mathbb{Q}$. Now we can define:

$$v_1 \leq_i v_2 \text{ iff for each } n \in N,\ v_1(n) \leq_i v_2(n).$$

Note that in the current work, we assume that all rational numbers are incomparable via $\leq_i$. That is, for each $x, y \in \mathbb{Q}$, if $x \neq y$, then neither $x <_i y$ nor $y <_i x$,

**Definition 10.10** *Let* $\mathscr{V}$ *be the set of all many-valued interpretations and let* $v_1$ *and* $v_2$ *be two interpretations of* $\mathscr{V}$*. Interpretation* $v_2$ *is called an extension of* $v_1$ *if* $v_1 \leq_i v_2$*. Interpretations* $v_1$ *and* $v_2$ *are called incomparable, denoted by* $v_1 \bowtie v_2$*, if neither* $v_1 \leq_i v_2$ *nor* $v_2 \leq_i v_1$*.*

The least interpretation, which is called *trivial interpretation*, is the one that maps all arguments to undecided, which is denoted by $v_u : N \to \{\mathbf{u}\}$.

**Example 10.11** *Let* $v = \{a \mapsto \mathbf{u}, s \mapsto \mathbf{u}, o \mapsto 1/3\}, v_1 = \{a \mapsto \mathbf{u}, s \mapsto 1/10, o \mapsto 1/3\}$ *and* $v_2 = \{a \mapsto \mathbf{u}, s \mapsto 1/10, o \mapsto 1/2\}$ *be three interpretations of* $\mathscr{V}$*. Since* $v$ *and* $v_1$ *are equivalent on an argument which is assigned*

*to a rational number by $v$ and $v(s) <_i v_1(s)$, $v_1$ is an extension of $v$. However, $v_2$ and $v$ are incomparable $v \bowtie v_2$, because neither $v(o) \leq_i v_2(o)$ nor $v_2(o) \leq_i v(o)$.*

**Definition 10.12** *Let $v \in \mathscr{V}$ be a many-valued interpretation. Then an extension $w$ of $v$ is called total if, for each $n \in N$, it holds that $w(n) \in \mathbb{Q}$. The set of total extensions of $v$ is denoted by $[v]_c$.*

The semantics of nADFs, similarly to the semantics of ADFs defined in Section 2.5.1, are defined based on a *characteristic operator* $\Gamma_U$ on many-valued interpretations that are ordered by ordering $\leq_i$. This shows that nADFs form an appropriate generalization of ADFs. The meet operator $\sqcap$ for nADFs[1] is a generalization of the meet operator for ADFs, presented in Section 2.2, defined on rational numbers plus $\mathbf{u}$ such that for each $x, y \in \mathbb{Q} \cup \{\mathbf{u}\}$, $x \sqcap y = x$ if $x = y$, and it returns $\mathbf{u}$, otherwise. The meet of two interpretations $v$ and $w$ is then defined as $(v \sqcap w)(n) = v(n) \sqcap w(n)$ for $n \in N$. The operator $\Gamma_U$ transforms interpretations of nADFs into others. Specifically, $\Gamma_U : \mathscr{V} \to \mathscr{V}$; the operator takes a many-valued interpretation $v$ as an input and returns a many-valued interpretation $\Gamma_U(v)$. For a given nADF $U = (N, L, C, i)$, the characteristic operator $\Gamma_U$ on an argument $n$ for the given interpretation $v$ is the meet of all total extensions of $v$ on $n$, as defined below.

**Definition 10.13** *Let $U = (N, L, C, i)$ be an nADF, let $v$ be an interpretation, and let $\varphi_n$ be an acceptance condition of $n$. The operator $\Gamma_U(v)$ yields a new interpretation:*

$$\Gamma_U(v) : N \to V \quad \textit{with} \quad n \mapsto \bigsqcap \{w(\varphi_n) \mid w \in [v]_c\}.$$

Some of the different types of semantics of nADFs are given below. These are the same as the types of semantics of standard ADFs when interpretations are three-valued and input function $i$ is not defined on any argument. The intuition of defining semantics of nADFs is the same as the intuition of semantics of ADFs, presented in Section 2.5.1.

**Definition 10.14** *Let $U = (N, L, C, i)$ be an nADF, let $v$ be an interpretation and let $\mathbf{v}$ be the $i$-correction of $v$ under $U$. An interpretation $v$ is:*

- *admissible in $U$ iff $v \leq_i \Gamma_U(\mathbf{v})$;*

[1] Note that the meet operator in this chapter has a different meaning (taking input from $\mathbb{Q} \cup \{\mathbf{u}\}$) than the one presented in Section 2.2 (taking three values as input).

- *complete in $U$ iff $v = \Gamma_U(\mathbf{v})$;*
- *grounded in $U$ iff $v$ is the $\leq_i$-least fixed point of $\Gamma_U$;*
- *preferred in $U$ iff $v$ is $\leq_i$-maximal admissible;*
- *model in $U$ iff $v = \Gamma_U(\mathbf{v})$ and $\forall n \in N,\ v(n) \neq \mathbf{u}$;*

*Note that in this definition, $\Gamma_U$ is applied to $\mathbf{v}$, the i-correction of $v$ under $U$. The sets of adm($U$), com($U$), grd($U$), prf($U$) and mod($U$) denote the sets of all admissible interpretations, complete interpretations, the unique grounded interpretation, preferred interpretations and models of $U$, respectively.*

**Example 10.15** *Continuing Example 10.6, let $v = \{a \mapsto 0, b \mapsto u, c \mapsto u\}$. The i-correction of $v$ under $U$ is $\mathbf{v} = \{a \mapsto 0, b \mapsto 1/5, c \mapsto 4/5\}$ since $i$ is defined on $b$ and $c$. Since none of the arguments of $\mathbf{v}$ assign to $\mathbf{u}$, $[\mathbf{v}]_c = \{\mathbf{v}\}$. Therefore, $\mathbf{v}(\varphi_b) = \mathbf{v}(b \vee a) = max\{\mathbf{v}(b), \mathbf{v}(a)\} = max\{i(b), \mathbf{v}(a)\} = max\{1/5, 0\} = 1/5$. That is, $\Gamma_U(\mathbf{v})(b) = 1/5$. In the same way, since $\mathbf{v}(a \otimes c) = \mathbf{v}(a) \times i(c) = 0$ and $\mathbf{v}(b) = 1/5$, $\mathbf{v}(a \otimes c) < \mathbf{v}(b)$ and $\Gamma_U(\mathbf{v})(c) = \mathbf{v}(\varphi_c) = 0$. Since $\Gamma_U(\mathbf{v}) = \{a \mapsto 0, b \mapsto 1/5, c \mapsto 0\}$ and $v \leq_i \Gamma_U(\mathbf{v})$, $v$ is an admissible interpretation of $U$. In addition, $\Gamma_U(\mathbf{v})$ is a preferred interpretation, a complete interpretation, and a model of $U$.*

*To compute the grounded interpretation of $U$, we evaluate the $\leq_i$-least fixed point of $\Gamma_U$. The i-correction of the trivial interpretation $v_\mathbf{u}$, that assigns all arguments to $\mathbf{u}$, is $\mathbf{v}' = \{a \mapsto \mathbf{u}, b \mapsto 1/5, c \mapsto 4/5\}$. Since $a$ is assigned to $\mathbf{u}$ in $\mathbf{v}'$, it holds that $[\mathbf{v}']_c$ has infinitely many elements. For instance, $w, w' \in [\mathbf{v}']$, where $w = \{a \mapsto 0, b \mapsto 1/5, c \mapsto 4/5\}$ and $w' = \{a \mapsto 1, b \mapsto 1/5, c \mapsto 4/5\}$. Since $w(\varphi_b) = w(b \vee a) = max\{w(b), w(a)\} = max\{i(b), w(a)\} = max\{1/5, 0\} = 1/5$ and $w'(\varphi_b) = w'(b \vee a) = max\{w'(b), w'(a)\} = max\{i(b), w'(a)\} = max\{1/5, 1\} = 1$, it holds that $w(\varphi_b) \sqcap w'(\varphi_b) = 1/5 \sqcap 1 = \mathbf{u}$. Thus, $\Gamma_U(\mathbf{v}')(b) = \mathbf{u}$. By the same method we find that $\Gamma_U(\mathbf{v}')(a) = \mathbf{u}$ and $\Gamma_U(\mathbf{v}')(c) = \mathbf{u}$. Thus, $\Gamma_U(\mathbf{v}') = v_\mathbf{u}$. That is, the revision of the trivial interpretation $v_u$ under $\Gamma_U$ is $v_u$, that is, $\Gamma_U(\mathbf{v}') = v_\mathbf{u}$, which is the unique grounded interpretation and a complete interpretation of $U$ as well but not a preferred interpretation.*

## 10.4 Embedding of Decision Problems in nADFs

In this section, we investigate how the standard decision problems introduced in Definition 10.2 can be embedded in nADFs. Since there are three

main different types of arguments, namely action, state and outcome, in a decision problem, different symbols are used for distinct types of them. Circles are used to represent action nodes; diamonds represent state nodes; and boxes represent outcome statements, depicted in Figure 10.3.

**Definition 10.16** *A decision problem* $D = (A, S, O, p, u)$*, where* $A = \{a_1, \ldots, a_n\}$*,* $S = \{s_1, \ldots, s_m\}$*, and* $O = \{o_{11}, \ldots, o_{nm}\}$ *can be modeled by nADF* $U_D = (N, L, C, i)$ *as follows:*

- $N = A \cup S \cup O$*;*
- $\varphi_s = s$ *for* $s \in S$*;*
  $\varphi_o = o$ *for* $o \in O$*;*
  $\varphi_{a_i} = \bar{\bigwedge}_{k \neq i, k \leq n}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj}))$ *for* $a_i \in A$*;*
- $i(s) = p(s)$ *for* $s \in S$ *and* $i(o) = u(o)$ *for* $o \in O$*.*

Let us explain some of the elements of the above definition. Self-loops in a graph of a decision problem can be utilized as a guess whether or not to accept an argument or to which extent to accept an argument. For instance, self-loops on state nodes are used to show that the degree of acceptance of each state node depends on the probability of occurrence of that state. This notion leads to name this type of links as self-dependent links; the set of self-dependent links is denoted by $R_d$. Self-dependent links are reflexive relations that can be defined on $N$. In nADFs, any other link which is not self-dependent is called an event link; the set of event links is denoted by $R_e$. In the nADF depicted in Figure 10.3, $(s_1, s_1) \in R_d$ and $(s_1, a_1) \in R_e$.

Since both the probability of occurrence of states and the utility of outcomes play a role in choosing actions, there are event links from states and outcomes to actions. The degree of acceptance of states only depends on the probability of occurrence of that state. Therefore, the acceptance condition of each state node $s \in S$ is $\varphi_s = s$. Similarly, the degree of the acceptance of outcomes only depends on the utility of that outcome from an agent's point of view, that is, $\varphi_o = o$ for each $o \in O$. Thus, in each nADF there exists a self-dependent link on each state and outcome node. The acceptance condition of each action is defined in a way that the best action can be chosen via semantics. That is, for $a_i \in A$ we have:

$$\varphi_{a_i} = \bar{\bigwedge}_{k \neq i, k \leq n}\Big(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj})\Big).$$

This formula uses a comparison of the expected utilities of actions, in order to express that action $a_i$ has maximal expected utility. To model decision

problems by nADFs, function $i$ on each state node $s$ is defined to be $p(s)$ and on each outcome node $o$ is $u(o)$. In the current work, we assume that in a decision problem, an agent is aware of the probability of states and her/his utility of each outcome. That is, the values of these functions are part of an input of the decision problem.

**Example 10.17** *Continuing the example introduced in Section 10.1 in which Maryam wants to decide whether to buy the international health insurance, the following propositional atoms are used to model this knowledge base.*

> $a_1$*: Maryam buys the international health insurance.*
>
> $a_2$*: Maryam does not buy the international health insurance.*
>
> $s_1$*: Maryam gets emergency surgery when she is abroad.*
>
> $s_2$*: Maryam does not get emergency surgery when she is abroad.*
>
> $o_{11}$*: Maryam gets emergency surgery when she is abroad and it is paid by the health insurance company.*
>
> $o_{12}$*: Maryam buys the international health insurance but she does not use it.*
>
> $o_{21}$*: Maryam gets emergency surgery when she is abroad and she has to pay by herself.*
>
> $o_{22}$*: Maryam does not buy the international health insurance and she does not need it.*

*We assume that p is a probability function on states, that is,* $p(s_1)$ *shows the probability of Maryam getting emergency surgery when she is abroad and* $p(s_2)$ *indicates the probability of Maryam does not get emergency surgery when she is abroad. Assume that* $p(s_1) = 1/10$ *and therefore* $p(s_2) = 9/10$*. Maryam's preference order on outcomes is as follow:* $o_{22} \succ_p o_{12} \succ_p o_{11} \succ_p o_{21}$*. The utility function u which keeps the same order, from Maryam's point of view, is:* $u(o_{22}) = 7/8, u(o_{12}) = 5/8, u(o_{11}) = 1/2, u(o_{21}) = 3/8$*. Therefore, this problem is modeled by decision problem* $D = (\{a_1, a_2\}, \{s_1, s_2\}, \{o_{11}, o_{12}, o_{21}, o_{22}\}, \{p(s_1), p(s_2)\}, \{u(o_{11}), u(o_{12}), u(o_{21})\})$.
*The corresponding nADF of D, depicted in Figure 10.3, is* $U_D = (N, L, C, i)$ *where:*

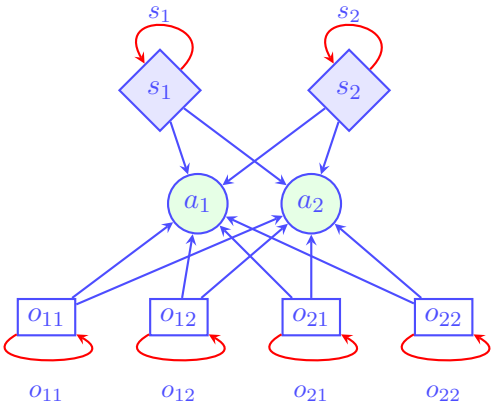


**Figure 10.3:** nADF of whether to buy international health insurance, used in Example 10.17

- $N = \{a_1, a_2, s_1, s_2, o_{11}, o_{12}, o_{21}, o_{22}\}$;
- $\varphi_{a_1} = \bigoplus_{j\in\{1,2\}}(s_j \otimes o_{1j}) \succeq \bigoplus_{j\in\{1,2\}}(s_j \otimes o_{2j})$;
- $\varphi_{a_2} = \bigoplus_{j\in\{1,2\}}(s_j \otimes o_{2j}) \succeq \bigoplus_{j\in\{1,2\}}(s_j \otimes o_{1j})$;
- $\varphi_{s_1} = s_1$;
- $\varphi_{s_2} = s_2$;
- $\varphi_{o_{11}} = o_{11}$;
- $\varphi_{o_{12}} = o_{12}$;
- $\varphi_{o_{21}} = o_{21}$;
- $\varphi_{o_{22}} = o_{22}$;
- $i(s_1) = p(s_1),\ i(s_2) = p(s_2)$;
- $i(o_{11}) = u(o_{11}),\ i(o_{12}) = u(o_{12}),\ i(o_{21}) = u(o_{21}),\ i(o_{22}) = u(o_{22})$.

The notation $X_x^v$ is used to show the set of arguments of $X$ which are assigned to $x$ by $v$ such that $x \in \mathbb{Q} \cup \{\mathbf{u}\}$ and $X$ can be either $A$, $S$ or $O$. For instance, in Example 10.11 that $v = \{a \mapsto \mathbf{u}, s \mapsto \mathbf{u}, o \mapsto 1/3\}$, $A_1^v = \{\}$ and $O_{1/3}^v = \{o\}$.

**Example 10.18** *Continuing Example 10.17, let* $v = \{a_1 \mapsto 1, a_2 \mapsto \mathbf{u}, s_1 \mapsto \mathbf{u}, s_2 \mapsto 1/5, o_{11} = \mathbf{u}, o_{12} = 5/8, o_{21} = \mathbf{u}, o_{22} = \mathbf{u}\}$. *Intuitively, interpretation* $v$ *wants to investigate whether it is rational for an agent to pick action* $a_1$ *when she/he only knows the probability of* $s_2$ *occurring and utility of output* $o_{12}$*, as input values. Particularly, in the current example, Maryam wants to decide whether it is rational for her to buy the international health insurance when she assumed that the probability of not getting emergency surgery is* $1/5$ *and her utility of buying the international health insurance and not using it is* $5/8$*. To do so, we compute the revise of* $\mathbf{v}$ *by* $\Gamma_U$*,* $v_1 = \Gamma_U(\mathbf{v}) = \{a_1 \mapsto 0, a_2 \mapsto 1, s_1 \mapsto 1/10, s_2 \mapsto 9/10, o_{11} = 1/2, o_{12} = 5/8, o_{21} = 3/8, o_{22} = 7/8\}$.

*Since* $v$ *and* $v_1$ *are incomparable on* $a_1$ *and* $s_2$*, we have that* $v \not\leq_i v_1$*. That is,* $v$ *is not an admissible interpretation of* $U_D$*. That is, based on this piece of information that is presented in* $v$*, it is not reasonable that Maryam pick* $a_1$*. However,* $v_1$ *is the unique complete interpretation of* $U_D$*. That is, if the information of Maryam increases to* $v_1$ *about the probabilities of states and utilities of outcome, then choosing* $a_2$ *is a feasible choice for Maryam.*

The proof of the uniqueness of the model in an nADF which is constructed based on a decision problem is given in Proposition 10.19.

**Proposition 10.19** *Let* $D = (A, S, O, p, u)$ *be a decision problem and let* $U_D = (N, L, C, i)$ *be the corresponding nADF. Let* $v$ *be an arbitrary interpretation of* $U_D$ *and let* $\mathbf{v}$ *be the* $i$*-correction of* $v$ *under* $U_D$*. The least fixed point of* $\Gamma_{U_D}$ *on* $\mathbf{v}$ *is a model of* $U_D$*.*

**Proof** Let $v$ be an arbitrary interpretation on $U_D$ and let $\mathbf{v}$ be the $i$-correction of $v$ under $U_D$. By the definition of acceptance conditions of states and outcome nodes and the definition of $\Gamma_{U_D}$, $v_1 = \Gamma_{U_D}(\mathbf{v})$ assigns each state node to its probability, each outcome node to its utility, and after computation each action to either 1 or 0, because of the notion of the truth operator $\bar{\bigwedge}$, presented in Definition 10.9. Moreover, the value of actions, states and outcome nodes do not change by iteration of this operator on $v_1$. That is, $v_1$ is the least fixed point of $\Gamma_{U_D}$. If $v \leq_i v_1$, then $v_1$ is the least fixed point of $\Gamma_{U_D}$ and $v$ is an admissible interpretation. However, if $v$ and $v_1$ are incomparable, then $v_1$ is the least fixed point of $\Gamma_{U_D}$ and $v$ is not an admissible interpretation. Therefore, in all cases $v_1$ is the least fixed point of $\Gamma_{U_D}$, and we have $v_1 = \Gamma_{U_D}(v_1)$. Since $v_1(n) \neq \mathbf{u}$ for each $n \in N$, $v_1$ is a model of $U_D$. □

**Corollary 10.20** *Assume that a decision problem $D = (A, S, O, p, u)$ is modeled by nADF $U_D = (N, L, C, i)$. Then all types of semantics of $U_D$ coincide.*

**Proof** Let $v$ be an arbitrary interpretation and let $\mathbf{v}$ be the $i$-correction of $v$ under $U_D$. By Proposition 10.19 the least fixed point of $\Gamma_{U_D}$ on $\mathbf{v}$ is a model of $U_D$ and by the Definition 10.14 it is a preferred interpretation. By the Definition 10.14, each grounded interpretation is a complete interpretation. It is enough to show that this complete interpretation is unique. Thus, all semantics of $U_D$ are equivalent. Toward a contradiction, assume that $|com(U_D)| > 1$, then by the definition $\sqcap com(U_D)$ is the least fixed point of $\Gamma_{U_D}$ that cannot be a model of $U_D$. This is a contradiction by the assumption that the least fixed point of $\Gamma_{U_D}$ is a model of $U_D$. $\square$

Theorem 10.21 investigates how semantics of nADFs can be used to choose the set of the best actions of decision problems of an agent.

**Theorem 10.21** *Let $D = (A, S, O, p, u)$ be a decision problem, where $A = \{a_1, \ldots, a_n\}$, $S = \{s_1, \ldots, s_m\}$, and $O = \{o_{11}, \ldots, o_{nm}\}$. Now let $U_D = (N, L, C, i)$ be the corresponding nADF, and let $v$ be the grounded interpretation of $U_D$, which is also the unique preferred interpretation, complete interpretation, and model of $U_D$. Then the set $A_1^v$ of actions evaluated as 1 in the grounded interpretation $v$ equals the set of actions with maximal expected utility in the decision problem $D$.*

**Proof** Let $M$ be the set of actions with maximal expected utility in the decision problem $D$. We show that $A_1^v = M$. To this end, we show that $A_1^v \subseteq M$ and $M \subseteq A_1^v$. Let $a_i \in A_1^v$. Since $a_i \in A$, it holds that $\varphi_{a_i} = \bar{\bigwedge}_{k \neq i, k \leq n}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj}))$. Furthermore, $a_i \in A_1^v$ means that $\Gamma_U(v)(a_i) = 1$, that is, for each $k$ it holds that $\mathbf{v}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj})) = 1$. That is, for each $k$, it holds that $\mathbf{v}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij})) \geq \mathbf{v}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj}))$. That is, for each $k$ the expected utility of $a_i$ is not less than the expected utility of $a_k$. Hence, it holds that $a_i \in M$.

Assume that $a_i \in M$. Thus, $a_i$ is an action with the acceptance condition $\bar{\bigwedge}_{k \neq i, k \leq n}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj}))$ in $U_D$. Since $a_i \in M$, it holds that the expected utility of $a_i$ is greater than or equal with the excepted utility of each $a_k$. That is, for each $k$ it holds that $\mathbf{v}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj})) = 1$. Thus, $\mathbf{v}(\bar{\bigwedge}_{i \neq k}(\bigoplus_{1 \leq j \leq m}(s_j \otimes o_{ij}) \succeq \bigoplus_{1 \leq j \leq m}(s_j \otimes o_{kj}))) = 1$. Thus, $\Gamma_U(\mathbf{v})(a_i) = 1$. Hence, it holds that $a_i \in A_1^v$.

□

## 10.5 Conclusion

In the chapter, argumentation is formally connected to decision making, by developing a formal connection between argumentation formalisms and EUT. This is significant for argumentation since the general issue how argumentation relates to the standard setting of EUT is not fully understood. This chapter provides a step in that understanding. The result is significant for decision making since an argumentation perspective provides insight in how to defend different positions, which remains unaddressed in theories of decision making. Generally, it is relevant to study the bridging of qualitative and quantitative theories (here: ADF as a theory of argumentation and EUT as a theory of decision making).

The chapter proposed an argumentation formalism, numerical abstract dialectical frameworks (nADFs), that can model standard decision problems. In (Bondarenko et al., 1997; Dung et al., 2009; Fan and Toni, 2011; Verheij, 2016b), other formalisms for modeling a decision problem are presented. The ability of doing arithmetical calculation makes nADFs an applicable formalism in decision-making problems, for example, in the medical domain. Our proposal specifically generalizes abstract dialectical frameworks ADFs to allow the modeling of standard decision problems. ADFs are special cases of nADFs in which formulas are limited to the standard language of propositional logic, $i$ is empty, and the semantics is defined based on three-valued interpretations.

Semantics of nADFs are defined based on many-valued interpretations, similarly to weighted abstract dialectical frameworks wADFs (Brewka et al., 2018b), which are also generalizations of ADFs.

A weighted ADF is a tuple $(N, L, C, V, \leq_i)$ in which $V$ indicates the set of truth values of arguments and $\leq_i$ is an ordering on $V$. That is, semantics of wADFs are also defined based on many-valued interpretation. To do calculation in nADFs, the set of truth values is fixed to the rational numbers plus $\mathbf{u}$ and the information ordering is a generalization of the standard information ordering defined in ADFs. The language which is used in the acceptance conditions of nADFs is a variation of the language of propositional logic, with two new function symbols $\otimes$ and $\oplus$ and and a truth operator $\bar{\bigwedge}$ and a predicate $\succeq$, and the partial function $i$ in nADFs. These additions empower the formalism to represent arithmetical calculations.

In general, nADFs are not a special case of wADFs. However, if in an nADF the formulas are restricted to propositional logic and the input function is empty, then it can also be viewed as a wADF in which the set of truth values $V$ is $\mathbb{Q} \cup \{\mathbf{u}\}$ and $\leq_i$ is a standard generalization of information ordering in ADFs.

It is constructively proven in (Diller et al., 2018) that in each acyclic ADF, all semantics coincide. In the current work, it is shown that in each nADF that formalizes a decision problem, all semantics coincide, as well. In Section 10.4 it is shown how an nADF can be constructed for a decision problem for a single-agent system to choose the best action. As to future work, it can be investigated whether nADFs can be used for modeling decision problems in multi-agent systems. In addition, it would be interesting to investigate whether nADFs are powerful enough to answer queries, for instance, "for which probabilities of needing an emergency surgery Maryam will decide to buy an insurance?" where the answer can be an interval of probabilities. Moreover, the computational complexity of decision problems in nADFs can be studied. Finally, it can be interesting to study simulation experiments that show the effectiveness of nADFs modeling decision problems.

# Part V

# Discussion and Conclusion

# Chapter 11

# Discussion and Conclusion

In this final chapter, we recapitulate our main contributions, give an overview of related work, and refer to possible future research directions suggested by our work.

## 11.1 Summary

ADFs form one of the most comprehensive formalisms for abstract argumentation, capturing several of the most important relations beyond that of simple attacks underlying Dung's AFs. We have in this thesis studied ADFs from several different perspectives.

1. In Part II, we focused on the semantical evaluation of ADFs, presenting two novel semantics.
    - In Chapter 3, we developed strong admissibility semantics for ADFs, based on the corresponding semantics for AFs.
    - In Chapter 4, we analyzed the complexity of the reasoning tasks of ADFs under the strong admissibility semantics.
    - In Chapter 5, we introduced semi-stable semantics for ADFs. This semantics approximates the stable semantics of ADFs in the situations in which an ADF does not have any stable extension.
2. In Part III, we considered reasoning for ADFs, presenting discussion games that decide the credulous acceptance problem.
    - In Chapter 6, we presented grounded discussion games for ADFs.

- In Chapter 7, we presented preferred discussion games for ADFs.

3. In Part IV, we studied variations of ADFs, in particular in order to further clarify the expressiveness of ADFs.

    - In Chapter 8, we introduced subclasses of ADFs analogous to known important subclasses of AFs and investigated to what extent properties that these subclasses satisfy for AFs also hold for ADFs.
    - In Chapter 9, we considered another well-known generalization of AFs, namely SETAFs, to clarify the relation between SETAFs and (a subclass of) ADFs.
    - In Chapter 10, we combine ADFs and decision theory, proposing a generalization of ADFs to model expected utility problems.

We now give a more detailed summary of the above mentioned main contributions of our work.

**Part II: Semantics** As already indicated, in Part II we first presented and then investigated key properties of two novel semantics for ADFs.

**Chapter 3: Strong Admissibility** We first generalised the strong admissibility semantics for AFs to ADFs. The strong admissibility semantics is a refinement of the grounded semantics, which is often referred to as the most skeptical of semantics in that it only assigns truth values to arguments which are shared among all admissible and, hence, complete interpretations.

The strongly admissible interpretations of an ADF form a lattice with the trivial interpretation (assigning the truth value undecided to all statements) being the unique minimal element and the grounded interpretation being the unique maximal element. Strongly admissible interpretations assign truth values that correspond to the grounded semantics, but may leave some arguments as undecided.

Apart from showing that our definition of strongly admissible semantics for ADFs is a proper generalization of the strongly admissible semantics of AFs, in Chapter 3 we also present algorithms, first of all for the verification problem for the strongly admissible semantics for ADFs. This is the problem of deciding whether a given interpretation is strongly admissible. Also, we present an algorithm to decide whether an argument is strongly justifiable in a given interpretation of an ADF. Here the notion of strong

justifiability is at the basis of the strongly admissible semantics in the sense that all arguments that are given a truth value under the strongly admissible semantics are strongly justifiable.

**Chapter 4: Complexity of Strong Admissibility** Studying the complexity of reasoning tasks for a formalism sheds light on the expressivity of the formalism and serves to guide implementation efforts. With this motivation in mind, in Chapter 4, we investigated the computational properties of the strong admissibility semantics of ADFs that we defined in Chapter 3. We have shown that decision problems for ADFs have higher computational complexity under the strong admissibility semantics when compared to AFs. This is also the case for several of the other semantics that ADFs inherit from AFs.

Specifically, when comparing the complexity of grounded and strong admissibility semantics, we find that for AFs the verification problems can be (log-space) reduced to one another, while, in the case of ADFs, there is a gap between the coNP-complete $Ver_{sadm}$ problem and the DP-complete $Ver_{grd}$ problem. Table 11.1 shows our results regarding the complexity of strong admissibility semantics of ADFs (highlighted in green) in the context of the complexity of the other semantics for ADFs. Here $\mathscr{C}$-c denotes completeness for the complexity class $\mathscr{C}$.

In Chapter 4 we elaborate on the complexity of the strong admissibility semantics by first of all considering the complexity of the strong justification problem, i.e., the problem whether an argument of interest is strongly justified in a given interpretation. The interest in this problem results from the fact that an argument can be strongly justified in an interpretation that is not strongly admissible. We show that the strong justification problem is coNP-complete.

We then considered the problem of finding a smallest witness for strong justification of an argument, that is, the problem of determining whether there exists a strongly admissible interpretation that assigns a minimum number of arguments to $\mathbf{t}/\mathbf{f}$ and satisfies an argument of interest. We showed that this problem is $\Sigma_2^{\mathsf{P}}$-complete for ADFs, while the same reasoning task for AFs is NP-complete.

**Chapter 5: Semi-Stable Semantics** The second semantics for ADFs that we introduce in this thesis is the semi-stable semantics, again generalizing the homonymous semantics for AFs.

Note first of all that for ADFs there are two generalizations of the

| $\sigma$ | $Cred_\sigma$ | $Skept_\sigma$ | $Ver_\sigma$ |
|---|---|---|---|
| *cf* | NP-c | trivial | NP-c |
| *adm* | $\Sigma_2^P$-c | trivial | coNP-c |
| *prf* | $\Sigma_2^P$-c | $\Pi_3^P$-c | $\Pi_2^P$-c |
| *com* | $\Sigma_2^P$-c | coNP-c | DP-c |
| *grd* | coNP-c | coNP-c | DP-c |
| *stb* | $\Sigma_2^P$-c | $\Pi_2^P$-c | coNP-c |
| *mod* | NP-c | coNP-c | in P |
| *sadm* | coNP-c | trivial | coNP-c |

**Table 11.1:** Complexity of reasoning tasks with ADFs

stable semantics for AFs, the first being the two-valued model semantics and the second the stable semantics. In fact, every stable interpretation is also a two-valued model, but stable interpretations for ADFs also rule out circular dependencies in the support relation of ADFs. The latter is not an issue in AFs, because there is no support relation in AFs. Thus in fact there may be no stable interpretation for an ADF because either there is no two-valued model or each two-valued model contains a support cycle.

Our semi-stable semantics for ADFs keeps the mechanisms for detecting circular supports among arguments, yet it builds on an alternative semi-two-valued model rather than two-valued-model semantics for ADFs which in turn is based on the semi-stable semantics for AFs.

We show that the notions of semi-two-valued semantics and semi-stable semantics of ADFs presented in this thesis satisfy the following properties that make them proper generalizations of the semi-stable semantics for AFs:

1. If $v$ is a semi-two-valued model or a semi-stable model of $D$, then $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal among all complete interpretations of $D$.
2. Each semi-stable model and each semi-two-valued model is a preferred interpretation.
3. Each stable model of an ADF is a semi-stable model and a semi-two-valued model of that ADF.
4. Each ADF has at least one semi-two-valued model. However, there is no guarantee that a semi-stable model exists.
5. If an ADF has a two-valued model, then the notions of semi-two-valued semantics and two-valued semantics coincide for that ADF.

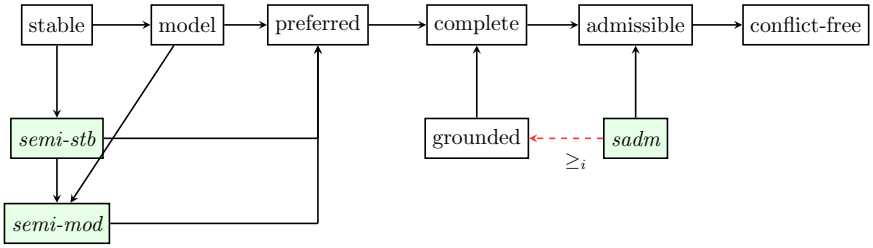


**Figure 11.1:** Relations among semantics of ADFs. The black arrow from semantics $\sigma$ to $\gamma$ indicates that $\sigma(D) \subseteq \gamma(D)$ for each ADF $D$. Here *sadm*, *semi-mod*, and *semi-stb* represent strong admissibility, semi-two-valued, and semi-stable semantics, respectively (with green boxes). The red dashed arrow indicates that each strongly admissible interpretation has at most an amount of information equal to the grounded interpretation, w.r.t. $\leq_i$-ordering.

6. If an ADF has a stable model, then the sets of stable models and semi-stable models coincide.
7. Semi-stable semantics and semi-two-valued semantics of ADFs are proper generalizations of semi-stable semantics of AFs.
8. Semi-stable semantics and semi-two-valued semantics coincide in the ADF associated to an AF, as is to be expected since in AFs there are no support cycles.

The relations between the novel semantics presented in this thesis and the existing semantics of ADFs are depicted in Figure 11.1. This figure thus extends Figure 2.11 which shows relations between semantics defined previous to our work.

In Figure 11.1, the novel semantics presented in this work, i.e., strong admissibility, semi-two-valued, and semi-stable semantics are represented as *sadm*, *semi-mod*, and *semi-stb*, respectively (with green boxes). The red-dash line between the strong admissibility semantics and the grounded semantics indicates that each strongly admissible interpretation has at most an amount of information equal to the grounded interpretation, with respect to $\leq_i$-ordering.

**Part III: Discussion Games** In Part III we defined discussion games to decide the credulous acceptance problems for the grounded and preferred semantics of ADFs. The credulous acceptance problem is that of deciding whether there exists an interpretation (grounded or preferred, respectively) in which an argument of interest is true. Discussion games not only decide

such problems, but also give a dialectical explanation of why an argument is true in terms of a strategy that a proponent of the argument can use to defend the argument against possible counter-arguments.

**Chapter 6: Grounded Discussion Games** In Chapter 6 we present a grounded discussion game (GDG for short) between two agents, i.e., proponent and opponent, to answer the credulous decision problem of ADFs under the grounded semantics. We show that such discussion games mostly allow for avoiding having to construct the whole grounded interpretation for deciding credulous acceptance of an argument of interest. We show soundness and completeness of our discussion game for the grounded semantics.

In Section 6.4 we study the relation between the notion of strong admissibility semantics of ADFs and the grounded discussion games of ADFs. Basically, the interpretations constructed at each step of a grounded discussion game are strongly admissible. On the other hand, our discussion games do not guarantee finding the least (with respect to the information order) strong admissible interpretation in which the argument of interest holds.

**Chapter 7: Preferred Discussion Games** Preferred interpretations are maximally informative admissible interpretations. In Chapter 7 we define discussion games for this semantics for ADFs. As before, discussion games for the preferred semantics allow to decide credulous acceptance of an argument without having to enumerate all preferred interpretations of an ADF. We show soundness and completeness of the discussion game that we propose for the preferred semantics.

**Part IV: Variations** In Part IV of the thesis we study several subclasses of ADFs and also present a generalization of ADFs.

**Chapter 8: Investigating Subclasses of ADFs** In Chapter 8, we defined several subclasses of ADFs based on specific subclasses of AFs and investigated how the restrictions that we consider influence the semantic evaluation of such ADFs. We list the main results of Chapter 8:

- First, we introduced acyclic ADFs (of which the link-structure forms an acyclic graph) and showed that, analogous to well-founded AFs (Dung, 1995), the main types of semantics, namely grounded,

complete, preferred, and two-valued model and stable semantics, coincide for this class of ADFs.

- We also introduced and studied the concept of symmetric ADFs. While the class of symmetric AFs is coherent and relatively-grounded, these properties do not carry over to the class of symmetric ADFs. This led us to propose subclasses of symmetric ADFs with further restrictions.

- We introduced acyclic support symmetric ADFs (ASSADFs) and support-free symmetric ADFs (SFSADFs). We showed that both classes satisfy a weaker form of coherence called weak coherence, where each two-valued model is a stable model, yet none of them satisfies the relatively-grounded property.

- It has been proved in (Dunne and Bench-Capon, 2002) that the reason that an AF is not coherent is that it contains an odd-cycle, i.e., odd-cycle free AFs are coherent. We showed that odd-cycle free SFSADFs are coherent while odd-cycle free ASSADFs are not. Thus, these two classes differ in the aspect of being odd-cycle free.

- Subsequently, we discussed the implication of our results for a generalization of AFs (captured by ADFs), namely SETAFs. To this end, we introduced symmetric SETAFs and we showed that this class is captured by a subclass of SFSADFs, namely, those in which the acceptance condition of none of the arguments is unsatisfiable.

- Finally, we studied the relation between the class of symmetric SETAFs and SFSADFs. Furthermore, we showed that the class of symmetric SETAFs, unlike symmetric AFs, is neither coherent nor relatively-grounded.

Another contribution of Chapter 8 is that we studied the expressiveness of subclasses of ADFs in terms of signatures, i.e., the sets of possible outcomes that can be achieved by ADFs (of a particular class) under the different semantics, following the work of Dunne et al. (2015). We compared our ADF subclasses to AFs in terms of expressiveness. Completing existing results regarding the relative expressivity of AFs and ADFs (Strass, 2015; Pührer, 2015; Linsbichler et al., 2016), we compared the expressiveness of the novel subclasses of ADFs, namely ASSADFs, SFADFs, and SFSADFs, with that of AFs, bipolar ADFs (BADFs), and ADFs. We showed the following results, depicted in Figures 8.7 and 8.8:

- ASSADFs and SFSADFs are incomparable with AFs for the admissibility semantics, while SFADFs are strictly more expressive.
- ASSADFs, SFADFs, and SFSADFs are strictly more expressive than AFs for the model and stable semantics.
- ASSADFs, SFADFs, and SFSADFs are strictly less expressive than BADFs for the model and admissibility based semantics.
- These classes, i.e., ASSADFs, SFADFs, and SFSADFs, coincide in expressiveness with BADFs and general ADFs for the stable semantics.

**Chapter 9: Expressiveness of SETAFs and Support-Free ADFs** After studying the relation between symmetric SETAFs and SFSADFs, as a subclass of ADFs in Chapter 8, we investigated further relations between SETAFs and subclasses of ADFs in Chapter 9. Specifically, we considerd two subclasses of ADFs, namely, SETADFs and SFADFs. In the first subclass, the acceptance conditions of ADFs are restricted so that they encode collective attacks as in SETAFs. In the second subclass, SFADFs, links between arguments are restricted semantically to be attacking.

To compare the SETAFs, SETADFs, and SFADFs we focused on 3-valued (labelling) semantics (Verheij, 1996; Caminada and Gabbay, 2009). In particular for SETAFs, we relied on the labelling semantics, introduced in (Flouris and Bikakis, 2019). Wallner (2020) already implicitly showed that SFADFs with satisfiable acceptance conditions can be equivalently represented as SETAFs. This provides a sufficient condition for encoding ADFs as SETAFs and raises the question whether it is also a necessary condition. In Chapter 9, we provided the following results:

- We showed that SETAFs and SETADFs coincide for the main semantics, i.e., the $\sigma$-labellings of a SETAF are equal to the $\sigma$-interpretations of the corresponding SETADF. That is, we showed 3-valued labeling based SETAF semantics to be equivalent to the corresponding ADF semantics.
- We provided exact characterisations of the 3-valued signatures for SETAFs (and thus for SETADFs) for most of the types of semantics. To provide exact characterisations, we investigated the expressiveness of SETAFs under 3-valued semantics. That is, we complemented the investigations on expressiveness of SETAFs (Dvořák et al., 2019) in terms of extension-based semantics.

- Then we determined the exact difference in expressiveness between SETADFs and SFADFs, from the perspective of realizability. In particular, we gave the exact difference in expressiveness between SETAFs and SFADFs under conflict-free, admissible, preferred, grounded, complete, stable, and two-valued model semantics.

- We showed that a SFADF can be encoded as a SETAF if and only if all acceptance conditions are satisfiable. Thus, what mainly distinguishes SETAFs and SFADFs is that in SFADFs initial arguments can be set to false.

**Chapter 10: Embedding Probabilities, Utilities and Decisions in a Generalization of ADFs** As a final contribution of this thesis in Chapter 10 we investigated combining argumentation with decision theory in the context of ADFs. Specifically we presented a generalization of ADFs, numerical ADFs (nADFs for short), that enable to encode the practical problem of computing expected utilities as an argumentation problem. We showed that when an nADF encodes such a decision problem, all semantics coincide. Thus, nADFs can be used to choose the best set of decisions (or actions) that have the maximum expected utility for a problem of interest.

The main difference between ADFs and nADFs is first of all that the language used to define acceptance conditions of nADFs is a variation of propositional logic that also allows for basic arithmetic. Furthermore, semantics of nADFs are defined based on many-valued interpretations while semantics of ADFs give three-valued interpretations. In other words, ADFs are a special case of nADFs in which formulas are limited to the standard language of propositional logic and the semantics is defined based on three-valued interpretations. Having arithmetic in the context of nADFs as well allows applying them for decision making, e.g., in the medical domain.

## 11.2 Related Work

As mentioned in the introduction to this thesis, the historical importance of argumentation as well as the amount of research in and diversity of argumentation formalisms and applications of argumentation in AI and in other scientific fields illustrate the significance of argumentation. Formal and computational argumentation methods underly a number of applications, for instance in law (Prakken and Sartor, 2015), medicine (Hunter

and Williams, 2012; Fox and Das, 2000), health promotion (Grasso et al., 2000), debating (Slonim et al., 2021), and dispute mediation (Janier et al., 2016); see (Atkinson et al., 2017) for a survey.

Dung's abstract argumentation frameworks (AFs) formalize argumentation by considering arguments as atomic entities whose acceptance status is to be determined by considering only the relation of attack among arguments. Then, semantics of AFs are criteria or methods proposed to evaluate arguments, i.e. to determine which sets of arguments can be accepted together. It has been proved in (Dung, 1995) that several important non-monotonic reasoning formalisms developed in AI, such as Pollock's defeasible reasoning (Pollock, 1987), Reiter's default logic (Reiter, 1980), and logic programming (Gabbay et al., 1998) can be reconstructed using AFs thus giving an argumentation perspective on such formalisms.

Based on Google Scholar, Dung's paper (Dung, 1995) has been cited more than 4,000 times. Dung's formalism has been widely used and studied mainly because of the simple structure of AFs, which are nothing more than directed graphs, and because of its intuitive and powerful semantics for evaluation of arguments. There are two main approaches for defining semantics of AFs, namely extension-based and labelling-based (see (Dung, 1995; Baroni et al., 2011) for an overview). In the first, semantics returns sets of arguments which can be accepted together. In the second, labellings are provided; the most popular of these being a three-valued labellings associating one of the values undecided, true, or false to arguments; extension-based semantics can be seen as providing two-valued labellings.

**Semantics** Several of the most important semantics for AFs have been introduced in (Dung, 1995), namely conflict-free, admissible, complete, preferred, and stable semantics. Further semantics for AFs have been proposed later on, for instance, stage semantics (Verheij, 1996), semi-stable semantics, first introduced in (Verheij, 1996) under a different name and then further investigated in (Caminada, 2006), strong admissibility semantics, first defined in the work of Baroni and Giacomin (2007), and later in (Caminada, 2014) without referring to the notion of strong defence, ideal semantics (Dung et al., 2007), and eager semantics (Caminada, 2007b). The relations among different semantics of AFs have been studied in (Baroni et al., 2011; Baroni and Giacomin, 2005, 2008; Caminada, 2007b; Caminada et al., 2012; Dung, 1995; Dung et al., 2007).

Much research has been devoted to AFs. For instance, AFs are used in

several diverse areas, such as multi-agent systems (McBurney et al., 2012), multi-agent negotiation (Amgoud et al., 2007) and legal reasoning (Bench-Capon and Dunne, 2005). Recently, the role of AFs in AI and Law has been discussed in (Bench-Capon, 2020).

As we have already mentioned, among all admissibility-based semantics of AFs, grounded semantics stands out for its important features. Thus, an agent may be eager to know 'why does an argument belong to the grounded extension?' There exist two approaches to answer this question.

- On the one hand, grounded discussion games have been introduced to answer the credulous decision problem of AFs (Prakken and Sartor, 1997; Caminada and Podlaszewski, 2012a,b) under grounded semantics. Discussion procedures in these games explain why an argument belongs to the grounded extension.

- On the other hand, the notion of strong admissibility semantics of AFs answers the question why an argument belongs to the grounded extension (Caminada and Dunne, 2019; Caminada, 2014). The notion of strong admissibility of AFs characterizes the unique properties of the grounded extension.

To target explaining why an argument is justified under the grounded semantics in ADFs, in the current thesis we have proposed the notion of strong admissibility semantics of ADFs in Chapter 3 and defined grounded semantics games for ADFs in Chapter 6.

The complexity of the reasoning problems that can be defined for the several semantics for AFs has been analyzed in (Dunne and Caminada, 2008; Dunne and Bench-Capon, 2002; Dvořák, 2012; Dvořák and Dunne, 2018; Dvořák and Wallner, 2020) and ranges from tractability up to the second level of the polynomial hierarchy. Computational complexity of reasoning tasks of AFs under strongly admissible semantics is studied in (Dvořák and Wallner, 2020). Moreover, Caminada and Dunne (2020) study the computational complexity of identifying strongly admissible labellings with bounded or minimal size. We analyzed the complexity of the relevant reasoning tasks for strong admissibility semantics of ADFs in Chapter 4.

Another semantics of AFs that has received increased attention and support not only in argumentation frameworks but also in logic programming (Gelfond and Lifschitz, 1988) and answer set programming (Gelfond and Lifschitz, 1991) is the stable semantics. This semantics represents the "black and white" perspective with a stable extension (or model) attacking all the arguments that are outside of the set. Thus, whereas each AF has

at least an admissible, a preferred, a complete and a unique grounded extension, it may have no stable extension. The notion of semi-stability for AFs is introduced in (Verheij, 1996; Caminada, 2006), as a way of approximating the stable semantics of AFs in situations where an AF has no stable extension.

The advantages of semi-stable semantics compared to other semantics of AFs are as follows.

1. First, in each AF, every stable extension is a semi-stable extension and every semi-stable extension is a preferred extension. Thus, semi-stable semantics are placed between stable semantics and preferred semantics.
2. Moreover, each finite AF has at least one semi-stable extension.
3. Finally, if an AF has at least one stable extension, then the semi-stable extensions are equal to the stable extensions, which is not always the case for preferred and grounded semantics.

The complexity of reasoning tasks of AFs under semi-stable semantics has been analyzed in (Dunne and Caminada, 2008; Dvořák and Woltran, 2010). Algorithms for computing the semi-stable extensions of a given AF are provided in (Caminada, 2007a; Wallner, 2014).

In Chapter 5 we have introduced the notion of semi-stable semantics for ADFs, analogously to semi-stable semantics for AFs, as a way of approximating stable semantics of ADFs.

In (Alcântara and Sá, 2018) the authors have also considered the semi-stable semantics for ADFs. To prevent confusion with the notion of semi-stable semantics presented in the current work, we call their notion semi-stable2 semantics, abbreviated SSS2. A key difference between our notion and SSS2 is that ours is compatible with the standard ADF definitions. In particular, in their discussion, the characteristic operator $\Gamma_D$, together with the semantics of ADFs, and specifically the complete semantics, are not presented in the way that was originally proposed by Brewka and Woltran (2018a; 2010). Thus, the relation among semantics of ADFs based on the definitions presented in (Alcântara and Sá, 2018) does not coincide with the relation among the semantics of ADFs presented in the main papers on ADFs, e.g., (Brewka et al., 2018a; Brewka and Woltran, 2010). Specifically, the set of complete models (interpretations) which is at the base of the definition of SSS2, presented in (Alcântara and Sá, 2018) may not coincide with the set of complete interpretations introduced in the central works on ADFs .

**Discussion Games** AF semantics can be seen declaratively as providing criteria that determine arguments that can be accepted together. A more procedural view of AF semantics, in particular to decide acceptability of arguments, underlies work on discussion games (Jakobovits and Vermeir, 1999; Prakken and Sartor, 1997; Modgil and Caminada, 2009; Caminada, 2018; Dung and Thang, 2007; van Eemeren et al., 2014). In such games acceptance of an argument according to some semantics $\sigma$ is determined by showing that there is a strategy for defending such an argument from arguments attacking it in an imagined dispute. Such a strategy can be seen as providing an explanation of why the argument of interest is to be accepted (according to $\sigma$) and thus discussion games often are intuitive means of determining acceptance of arguments accompanying the classical definitions of the semantics to the AF at hand. The dialogue perspective is also relevant for applications of argumentation.

Dialogical methods, can be traced back to the work of Hamblin (1971) and Mackenzie (1979; 1990). As an example, Socratic forms of reasoning (discussion games) are proposed in (Caminada et al., 2014) to answer the credulous decision problem of AFs under the preferred semantics. In (Cayrol et al., 2003), games are presented to answer the credulous and skeptical decision problems of AFs under the preferred semantics. A discussion game is proposed in (Dung and Thang, 2007) to answer the skeptical decision problem of AFs under the preferred semantics. It is proven that this method is sound for any AFs, while it is complete for finitary AFs. Next, Vreeswijk and Prakken (2000) present a discussion game to answer the credulous decision problem under preferred semantics and a game to answer the skeptical decision problem when an AF is coherent (i.e, when the set of preferred extensions and the set of stable extensions coincide). Furthermore, the Standard Grounded Game (Modgil and Caminada, 2009; Prakken and Sartor, 1997) and the Grounded Persuasion Game (Caminada and Podlaszewski, 2012a,b) have been presented to answer the credulous decision problem of AFs under grounded semantics. Further dialectical methods for AFs are presented in (Nofal et al., 2014; Modgil and Caminada, 2009).

In Chapters 6 and 7 we introduce, to the best of our knowledge, the first discussion games for ADFs. Particularly we do so for the grounded and preferred semantics. Thus, dialectical explanations for acceptance of arguments as for AFs can be obtained also in the context of the clearly more complex context of ADFs.

**Variations** There are various ways to determine the capabilities of knowledge representation formalisms. One way is to study the computational complexity of the reasoning tasks that are defined for such formalisms. Another approach is to study their expressivity: properties that characterise the outcomes of evaluating the formalisms via their associated semantics.

As analyzed in (Dunne and Caminada, 2008; Dunne and Bench-Capon, 2002; Dvořák, 2012; Dvořák and Dunne, 2018; Dvořák and Wallner, 2020), and illustrated in Table 2.2, the complexity of the reasoning problems that can be defined in the case of several of the most important semantics for AFs range from tractability up to the second level of the polynomial hierarchy. Thus the analysis of restricted classes of AFs is of importance, since the restrictions may make decision problems easier. The class of acyclic (also known as well-founded) AFs has been introduced by Dung in (1995). In addition, the class of symmetric AFs has been introduced in (Coste-Marquis et al., 2005). Further subclasses obtained via other constraints on the graphical structure of AFs are defined in (Dunne, 2007).

Dung (1995) shows that for acyclic AFs all the semantics considered in this work coincide. As presented in (Dvořák and Dunne, 2018) and shown in Table 2.2, the main decision problems of AFs under the grounded semantics are tractable. Thus, for the class of acyclic AFs, all the main decision problems that we present in Table 2.2 are also tractable under the preferred, complete, stable, semi-stable, and ideal semantics. On the other hand, the complexity results for acyclic ADFs have been established in (Linsbichler et al., 2018).

In (Coste-Marquis et al., 2005) it is shown that the class of symmetric AFs is coherent (i.e., preferred and stable semantics coincide), thus, each symmetric AF has at least one stable extension. Dvořák and Dunne (2018) show (Table 2.2) that preferred semantics have the highest computational complexity of the semantics we consider in this work. However, for symmetric AFs, the complexity of the decision problems under the preferred semantics are easier than in the general case. Specifically, the verification problem under the preferred semantics is coNP-complete, while this problem under the stable semantics is in P. Furthermore, it is shown in (Coste-Marquis et al., 2005) that the class of symmetric AFs is relatively-grounded (i.e., the grounded extension is given by the intersection of the preferred extensions). Thus, for the class of symmetric AFs, the skeptical decision problem under preferred semantics coincides with the credulous decision problem of AFs under grounded semantics. However, the credulous decision problem of AFs under grounded semantics is tractable, while the

skeptical decision problem under preferred semantics is on the second level of the polynomial hierarchy, namely it is $\Pi_2^P$-complete. More generally, semantic and syntactic restrictions on AFs often make decision problems easier from a complexity perspective. This fact has, for instance, been made use of in practice in the cegartix system (Dvořák et al., 2014). But studying subclasses of AFs also provides a better understanding of the semantics, by revealing under which conditions one obtains different or identical results.

Since coherence of AFs leads to a lower computational complexity of reasoning tasks under preferred semantics, Dunne and Bench-Capon (2002) study the reasons why an AF is not coherent. It is shown in (Dunne and Bench-Capon, 2002) that, if an AF is not coherent, then it contains a cycle of odd length. This means, by contraposition, that if an AF does not contain any odd-length cycle, then it is coherent.

The added modeling capabilities of ADFs representing complex links between arguments (beyond simple attacks) leads to higher computational complexity in comparison to AFs (Strass and Wallner, 2015; Dvořák and Dunne, 2018; Gaggl et al., 2021), as is shown in Table 2.3. This has been our motivation to define subclasses of ADFs analogous to those known for AFs in Chapter 8, and to investigate whether similar properties of such subclasses of AFs also hold for ADFs.

As argued in (Dunne et al., 2015), another approach to study and compare the capabilities of different semantics of AFs is achieved via the notion of expressiveness of formalisms from the perspective of realizability. Realizability is the ability of a formalism under a semantics to express specific desired sets of extensions. The so-called "signatures" capture the exact expressiveness of a formalism under a semantics by characterising the sets of extensions that can be obtained. Formally, the study of expressiveness of a formalism with respect to a semantics can be done by considering the outcomes that can be realised by the formalism under the semantics of interest. For instance, Dunne et al. (2015) show that preferred semantics of AFs are strictly more expressive than stable semantics, from the perspective of realizability. In contrast, preferred and semi-stable semantics of AFs have the same expressiveness. It is shown in (Linsbichler, 2017) that preferred and semi-stable semantics are among the most expressive semantics of AFs.

Furthermore, the notion of expressiveness can be used to compare the capability of different formalisms of argumentation. For an overview of generalizations of Dung's AFs that allow for a richer syntax, see (Brewka

et al., 2014). A generalization of AFs, namely argumentation frameworks with collective attacks (SETAFs), has been introduced in (Nielsen and Parsons, 2006) and they have received increasing attention recently (Dvořák et al., 2019; Flouris and Bikakis, 2019). Translations between SETAFs and other abstract argumentation formalisms have been studied in (Polberg, 2017). The expressiveness of SETAFs under two-valued semantics has been investigated in (Dvořák et al., 2019) in terms of signatures. It has been established in (Linsbichler et al., 2016; Linsbichler, 2017) that, for a fixed semantics, the class of bipolar ADFs is strictly more expressive than the class of SETAFs, and the class of SETAFs is more expressive than the class of AFs. In addition, (Dvořák et al., 2019) provides explicit characterisations of the two-valued signatures of SETAFs and shows that SETAFs are more expressive than AFs. Moreover, in (Linsbichler et al., 2016), an algorithm is provided to decide realizability in AFs, SETAFs, bipolar ADFs, and ADFs under the admissible, preferred, complete, model and stable semantics. In contrast, (Flouris and Bikakis, 2019) show that when allowing to add extra arguments to an AF that are not relevant for the signature, i.e. when the extensions/labellings are projected on common arguments, then SETAFs and AFs have equivalent expressivity. Furthermore, Pührer (2020b) presented explicit characterisations of the signatures of general ADFs (but not for the sub-classes discussed above).

Regarding this approach to comparing formalisms of argumentation, we complement previous work on expressiveness of AFs and ADFs (Strass, 2015; Pührer, 2015; Linsbichler et al., 2016) by comparing the expressiveness of ASSADFs, SFADFs, and SFSADFs with that of AFs, bipolar ADFs (BADFs), and ADFs, in Chapter 8. Furthermore, a sufficient condition for rewriting an ADF as a SETAF has been implicitly presented in (Wallner, 2020): that SFADFs with satisfiable acceptance conditions can be equivalently represented as SETAFs. This was a motivation for us to investigate whether this is a necessary condition for rewriting an ADF as a SETAF, as investigated in Chapter 9.

Argumentation theory can shed light on the process of decision making, from modeling to evaluating a problem. The use of ADFs as a semantic tool of modelling and evaluating of arguments in various scenarios has been presented in (Brewka et al., 2018a). For instance, it has been shown how graphical representations based on link types (supporting or attacking) can be represented by ADFs. Furthermore, ADFs have also been generalized, for example in weighted abstract dialectical frameworks (Brewka et al., 2018b) (wADFs for short), to accommodate arbitrary acceptance degrees

for the arguments.

All standard semantics of ADFs have been defined for weighted ADFs, while the semantics of weighted ADFs are defined based on many-valued interpretations, instead of three(two)-valued interpretations. A related approach in a multi-valued setting is (Dondio, 2014, 2017). Also, it is interesting to see (Alsinet et al., 2017) for an application of weighted argumentation on the Twitter social network. In addition, the notion of weights on links has been introduced by Brewka and Woltran in (2014) in order to simplify the definition of acceptance conditions in the context of GRAPPA frameworks.

Concerning a modeling example of a practical problem, we considered expected utility theory (EUT for short), which is a theory concerned with making the best decision under uncertainty (Von Neumann and Morgenstern, 1947; von Neumann and Morgenstern, 2007; Savage, 1954; Briggs, 2019; Gilboa, 2009; Mongin, 1998; Russell and Norvig, 2009). In Chapter 10, we defined ADFs that formalize situations induced by the probabilities and utilities and to evaluate EUT.

## 11.3 Future Work

**Semantics** As to future work, first of all, there are still a few noteworthy semantics for AFs which are lacking a counterpart in the ADF world. In particular, the ideal and eager semantics are unique-status semantics (i.e. returning a single interpretation) that are more credulous than the grounded semantics (which is often deemed too skeptical). Defining such semantics for ADFs would thus be of immediate relevance. Of course then establishing the relationship between these semantics for ADFs and the semantics considered in this work would also be crucial.

**Computational Complexity** There are still some open questions regarding reasoning in ADFs that follow from this thesis. Specifically, we studied the complexity of the main reasoning tasks that can be defined for ADFs with respect to the strong admissibility semantics in Chapter 4. However, the computational complexity of reasoning tasks for the semi-stable semantics of ADFs, presented in Chapter 5, is still open. Some of the reasoning tasks whose complexity require clarification are the following:

1. Whether a given interpretation is a semi-stable model or semi-two-valued model, i.e., the verification problem under the semi-stable semantics and semi-two-valued semantics;

2. Whether a given argument is credulously acceptable/deniable under the semi-stable/semi-two-valued semantics of a given ADFs, i.e., the credulous decision problem;
3. Whether a given argument is skeptically acceptable/deniable under the semi-stable/semi-two-valued semantics of a given ADF, i.e., the skeptical decision problem;
4. The problem of finding a small witness for justifying an argument, i.e, whether there exists a semi-stable model or a semi-two-valued model that satisfies a queried argument and is smaller than a given bound.

Furthermore, several subclasses of ADFs have been introduced (Brewka and Woltran, 2010; Diller et al., 2020), and it would be interesting to clarify the computational complexity of the reasoning tasks over subclasses of ADFs under semantics presented in this work, i.e., strong admissibility and semi-stable semantics of ADFs, in particular, for bipolar ADFs (Brewka and Woltran, 2010) and acyclic ADFs (Diller et al., 2020).

**Discussion Games** In this thesis, we assumed all ADFs to be finite. Thus, we presented our games, namely, preferred discussion games and grounded discussion games, for finite ADFs. Furthermore, in these games we assume that the acceptance conditions of arguments are represented by propositional formulas in ADFs. As future work, we are interested in investigating games for infinite ADFs and for ADFs for which the acceptance conditions are not restricted to propositional formulas, for example, for ADFs for which the language of the acceptance conditions are first-order logic or modal logic. Moreover, we could investigate structural discussion games for other semantics of ADFs. In addition, we could study discussion games for other decision problems of ADFs. We could also investigate whether the presented methods in the discussion games are more effective than the methods used in current ADF-solvers (Brewka et al., 2017a; Ellmauthaler and Strass, 2014). This study may lead to new ADF-solvers that work locally on an argument to answer decision problems.

**Variations** A further topic in this work was to verify whether properties known for subclasses AFs also hold for the new semantics and analogous subclasses of ADFs we considered in this work. For instance, we showed that most of the semantics of ADFs coincide for acyclic ADFs. In the same

vein, it would also be of interest to determine whether the semi-stable semantics coincides with the preferred semantics, and whether the strong admissibility semantics coincides with the admissible semantics in the class of acyclic ADFs. Furthermore, since all reasoning problems become coNP-complete for acyclic ADFs (cf. the complexity results presented formally in (Linsbichler et al., 2018)) our work offers further guidelines for designing more efficient systems for ADFs.

Also, the exact characterisation of admissible and complete semantics of SETAFs, SFADFs, and SETADFs are an interesting topic for future work. Another aspect to be investigated is to what extent our insights on labelling-based semantics for SETAFs and SFADFs can help to improve the performance of reasoning systems for such argumentation formalisms.

The last contribution of this thesis was to present a generalization of ADFs, called nADFs. The formalism of nADFs are used to model and evaluate the standard decision problem of expected utility in a single-agent system. As to future work, it can be investigated whether nADFs can be used for modeling decision problems in multi-agent systems. We leave for future work to study whether nADFs are powerful enough not only to find a set of best actions, but also to answer queries, such as for which probabilities it is reasonable to perform an action. Finally, the computational complexity of decision problems in nADFs can be studied.

# Bibliography

L. Al-Abdulkarim, K. Atkinson, and T. J. M. Bench-Capon. Abstract dialectical frameworks for legal reasoning. In *Legal Knowledge and Information Systems JURIX*, volume 271 of *Frontiers in Artificial Intelligence and Applications*, pages 61–70. IOS Press, Amsterdam, 2014.

L. Al-Abdulkarim, K. Atkinson, and T. J. M. Bench-Capon. A methodology for designing systems to reason with legal cases using abstract dialectical frameworks. *Artificial Intelligence Law*, 24(1):1–49, 2016.

J. F. L. Alcântara and S. Sá. On three-valued acceptance conditions of abstract dialectical frameworks. In *LSFA*, volume 344 of *Electronic Notes in Theoretical Computer Science*, pages 3–23. Elsevier, Amsterdam, 2018.

T. Alsinet, J. Argelich, R. Béjar, C. Fernández, C. Mateu, and J. Planes. Weighted argumentation for analysis of discussions in twitter. *International Journal of Approximate Reasoning*, 85:21–35, 2017.

L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009.

L. Amgoud, Y. Dimopoulos, and P. Moraitis. A unified and general framework for argumentation-based negotiation. In *Proceedings of the 6th international Joint Conference on Autonomous agents and Multiagent Systems*, pages 1–8. ACM Press, 2007.

S. Arora and B. Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, Cambridge, 2009.

K. J. Arrow. The use of unbounded utility functions in expected-utility maximization: Response. *The Quarterly Journal of Economics*, 88(1): 136–138, 1974.

K. Atkinson and T. J. M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

K. Atkinson and T. J. M. Bench-Capon. Taking account of the actions of others in value-based reasoning. *Artificial Intelligence*, 254:1–20, 2018.

K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. Simari, M. Thimm, and S. Villata. Towards artificial argumentation. *AI Magazine*, 38(3):25–36, 2017.

H. Ayoobi, M. Cao, R. Verbrugge, and B. Verheij. Handling unforeseen failures using argumentation-based learning. In *15th International Conference on Automation Science and Engineering (CASE 2019)*, pages 1699–1704. IEEE, 2019.

P. Baroni and M. Giacomin. Evaluating argumentation semantics with respect to skepticism adequacy. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 329–340. Springer, Berlin, 2005.

P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007.

P. Baroni and M. Giacomin. A systematic classification of argumentation frameworks where semantics agree. *Frontiers in Artificial Intelligence and Applications*, 172:37, 2008.

P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.

P. Baroni, M. Caminada, and M. Giacomin. Abstract argumentation frameworks and their semantics. *Handbook of Formal Argumentation*, 1: 157–234, 2018a.

P. Baroni, D. M. Gabbay, M. Giacomin, and L. van der Torre. *Handbook of Formal Argumentation*. College Publications, London, 2018b.

P. Baroni, F. Toni, and B. Verheij. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games: 25 years later. *Argument & Computation*, 11(1-2):1–14, 2020.

E. M. Barth and E. C. Krabbe. *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation.* Walter de Gruyter, Berlin-New York, 1982.

R. Baumann and G. Brewka. Extension removal in abstract argumentation - An axiomatic approach. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 2670–2677. AAAI Press, 2019.

T. Bench-Capon and K. Atkinson. Abstract argumentation and values. In *Argumentation in Artificial Intelligence*, pages 45–64. Springer, Berlin, 2009.

T. J. Bench-Capon and P. E. Dunne. Argumentation in ai and law: Editors' introduction. *Artificial Intelligence and Law*, 13(1):1–8, 2005.

T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3): 429–448, 2003.

T. J. M. Bench-Capon. Before and after Dung: Argumentation in AI and law. *Argument & Computation*, 11(1-2):221–238, 2020.

T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

J. Bentham. *An Introduction to the Principles of Morals and Legislation.* Doubleday, Originally published in 1789, Garden City, 1961.

P. Besnard, C. Cayrol, and M. Lagasquie-Schiex. Logical theories and abstract argumentation: A survey of existing works. *Argument & Computation*, 11(1-2):41–102, 2020.

F. Bex, J. Lawrence, M. Snaith, and C. Reed. Implementing the argument web. *Communications of the ACM*, 56(10):66–73, 2013.

S. Bistarelli and F. Santini. Abstract argumentation and (optimal) stable marriage problems. *Argument & Computation*, 11(1-2):15–40, 2020.

D. G. Bobrow. Special issue on non-monotonic logic: Preface. *Artificial Intelligence*, 13:1–4, 1980.

A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1-2):63–101, 1997.

R. Booth, M. Caminada, and B. Marshall. DISCO: A web-based implementation of discussion games for grounded and preferred semantics. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of Computational Models of Argument COMMA*, pages 453–454. IOS Press, Amsterdam, 2018.

G. Brewka and T. F. Gordon. Carneades and abstract dialectical frameworks: A reconstruction. In *Proceedings of the 2010 Conference on Computational Models of Argument: Proceedings of COMMA 2010*, pages 3–12, 2010.

G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, pages 102–111. AAAI Press, 2010.

G. Brewka and S. Woltran. Grappa: A semantical framework for graph-based argument processing. In *ECAI*, pages 153–158, 2014.

G. Brewka, P. E. Dunne, and S. Woltran. Relating the semantics of abstract dialectical frameworks and standard AFs. In *Twenty-Second International Joint Conference on Artificial Intelligence*. AAAI Press, 2011.

G. Brewka, S. Ellmauthaler, H. Strass, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks revisited. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, pages 803–809. AAAI Press, 2013.

G. Brewka, S. Polberg, and S. Woltran. Generalizations of Dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems*, 29 (1):30–38, 2014.

G. Brewka, M. Diller, G. Heissenberger, T. Linsbichler, and S. Woltran. Solving advanced argumentation problems with answer-set programming. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI 2017)*, pages 1077–1083. AAAI Press, 2017a.

G. Brewka, S. Ellmauthaler, H. Strass, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks. an overview. *IfCoLog Journal of Logics and their Applications*, 4(8):2263–2318, 2017b.

G. Brewka, S. Ellmauthaler, H. Strass, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks: An overview. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, chapter 5. College Publications, London, 2018a.

G. Brewka, H. Strass, J. P. Wallner, and S. Woltran. Weighted abstract dialectical frameworks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI, 2018)*, pages 1779–1786. AAAI Press, 2018b.

R. A. Briggs. Normative theories of rational choice: Expected utility. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2019 edition, 2019.

K. Budzynska, M. Janier, J. Kang, C. Reed, P. Saint-Dizier, M. Stede, and O. Yaskorska. Towards argument mining from dialogue. In *COMMA*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 185–196. IOS Press, Amsterdam, 2014.

E. Cabrio and S. Villata. Abstract dialectical frameworks for text exploration. In *Proceedings of International Conference on Agents and Artificial Intelligence (ICAART 2016)*, pages 85–95. SciTePress, 2016.

E. Cabrio and S. Villata. Five years of argument mining: A data-driven analysis. In *IJCAI*, pages 5427–5433. ijcai.org, 2018.

M. Caminada. Semi-stable semantics. In *COMMA*, volume 144 of *Frontiers in Artificial Intelligence and Applications*, pages 121–130. IOS Press, Amsterdam, 2006.

M. Caminada. An algorithm for computing semi-stable semantics. In *ECSQARU*, volume 4724 of *Lecture Notes in Computer Science*, pages 222–234. Springer, Berlin, 2007a.

M. Caminada. Comparing two unique extension semantics for formal argumentation: Ideal and eager. In *Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence (BNAIC 2007)*, pages 81–87. Utrecht University Press, 2007b.

M. Caminada. Strong admissibility revisited. In *Proceedings of Computational Models of Argument COMMA*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 197–208. IOS Press, Amsterdam, 2014.

M. Caminada. A discussion game for grounded semantics. In *International Workshop on Theory and Applications of Formal Argumentation*, pages 59–73. Springer, Berlin, 2015.

M. Caminada. Argumentation semantics as formal discussion. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, pages 487–518. College Publications, London, 2018.

M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

M. Caminada and P. E. Dunne. Strong admissibility revisited: Theory and applications. *Argument & Computation*, 10(3):277–300, 2019.

M. Caminada and P. E. Dunne. Minimal strong admissibility: A complexity analysis. In *Proceedings of Computational Models of Argument COMMA*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 135–146. IOS Press, Amsterdam, 2020.

M. Caminada and M. Podlaszewski. Grounded semantics as persuasion dialogue. In *Proceedings of Computational Models of Argument COMMA*, volume 245 of *Frontiers in Artificial Intelligence and Applications*, pages 478–485. IOS Press, Amsterdam, 2012a.

M. Caminada and M. Podlaszewski. User-computer persuasion dialogue for grounded semantics. In *Proceedings of BNAIC*, pages 343–344, 2012b.

M. Caminada and S. Uebis. An implementation of argument-based discussion using ASPIC-. In *COMMA*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 455–456. IOS Press, Amsterdam, 2020.

M. Caminada and B. Verheij. On the existence of semi-stable extensions. In *Proceedings of the 22nd Benelux Conference on Artificial Intelligence (BNAIC 2010)*. University of Luxemburg, 2010. URL `https://bnaic.gforge.uni.lu/proceedings.html`.

M. W. Caminada. A formal account of Socratic-style argumentation. *Journal of Applied Logic*, 6(1):109–132, 2008.

M. W. Caminada, W. A. Carnielli, and P. E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22(5):1207–1254, 2012.

M. W. Caminada, W. Dvořák, and S. Vesic. Preferred semantics as Socratic discussion. *Journal of Logic and Computation*, 26(4):1257–1292, 2014.

M. W. A. Caminada and D. M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.

C. Cayrol and M. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *ECSQARU*, volume 3571 of *Lecture Notes in Computer Science*, pages 378–389. Springer, Berlin, 2005.

C. Cayrol and M. Lagasquie-Schiex. Bipolar abstract argumentation systems. In *Argumentation in Artificial Intelligence*, pages 65–84. Springer, Berlin, 2009.

C. Cayrol, S. Doutre, and J. Mengin. On decision problems related to the preferred semantics for argumentation frameworks. *Journal of Logic and Computation*, 13(3):377–403, 2003.

F. Cerutti, S. A. Gaggl, M. Thimm, and J. Wallner. Foundations of implementations for formal argumentation. *IfCoLog Journal of Logics and their Applications*, 2017.

L. A. Chalaguine and A. Hunter. A persuasive chatbot using a crowdsourced argument graph and concerns. In *Proceedings of the 2020 conference on Computational Models of Argument: Proceedings of COMMA*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 9–20. IOS Press, Amsterdam, 2020.

G. Charwat, W. Dvořák, S. A. Gaggl, J. P. Wallner, and S. Woltran. Methods for solving reasoning problems in abstract argumentation – a survey. *Artificial intelligence*, 220:28–63, 2015.

O. Cocarascu and F. Toni. Argumentation for machine learning: A survey. In *COMMA*, volume 287 of *Frontiers in Artificial Intelligence and Applications*, pages 219–230. IOS Press, Amsterdam, 2016.

J. Collenette, K. Atkinson, and T. J. M. Bench-Capon. An explainable approach to deducing outcomes in European Court of Human Rights cases using ADFs. In *Proceedings of Computational Models of Argument COMMA 2020*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 21–32. IOS Press, Amsterdam, 2020.

S. Coste-Marquis, C. Devred, and P. Marquis. Symmetric argumentation frameworks. In *Proceedings of the 8th European Conference onSymbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 317–328. Springer, 2005.

B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 2002.

M. Davis. *Engines of Logic: Mathematicians and the Origin of the Computer*, volume 7. Norton, New York, 2000.

M. Davis, S. Maslov, G. Mints, V. Orevkov, M. Davis, A. Newell, J. Shaw, H. Simon, A. Robinson, E. Beth, et al. The prehistory and early history of automated deduction. *Automation of Reasoning 1: Classical Papers on Computational Logic 1957-1966*, 1983.

M. Denecker, V. Marek, and M. Truszczyński. Approximations, stable operators, well-founded fixpoints and applications in nonmonotonic reasoning. In *Logic-Based Artificial Intelligence*, pages 127–144. Springer, Berlin, 2000.

M. Denecker, V. W. Marek, and M. Truszczyński. Uniform semantic treatment of default and autoepistemic logics. *Artificial Intelligence*, 143 (1):79–122, 2003.

M. Denecker, V. W. Marek, and M. Truszczyński. Ultimate approximation and its application in nonmonotonic knowledge representation systems. *Information and Computation*, 192(1):84–121, 2004.

M. Diller, J. P. Wallner, and S. Woltran. Reasoning in abstract dialectical frameworks using quantified boolean formulas. In *Proceedings of Computational Models of Argument (COMMA 2014)*, pages 241–252. IOS Press, Amsterdam, 2014.

M. Diller, A. Keshavarzi Zafarghandi, T. Linsbichler, and S. Woltran. Investigating subclasses of abstract dialectical frameworks. In *Proceedings of Computational Models of Argument (COMMA 2018)*, pages 61–72. IOS Press, Amsterdam, 2018.

M. Diller, A. K. Zafarghandi, T. Linsbichler, and S. Woltran. Investigating subclasses of abstract dialectical frameworks. *Argument & Computation*, 11(1), 2020.

P. Dondio. Multi-valued and probabilistic argumentation frameworks. In *COMMA*, pages 253–260, 2014.

P. Dondio. Propagating degrees of truth on an argumentation framework: An abstract account of fuzzy argumentation. In *Proceedings of the Symposium on Applied Computing*, pages 995–1002, 2017.

P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995. ISSN 0004-3702.

P. M. Dung and P. M. Thang. A sound and complete dialectical proof procedure for sceptical preferred argumentation. In *Proceedings of the LPNMR-Workshop on Argumentation and Nonmonotonic Reasoning (ArgNMR07)*, pages 49–63, 2007.

P. M. Dung and P. M. Thang. Towards (probabilistic) argumentation for jury-based dispute resolution. In *COMMA*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 171–182. IOS Press, Amsterdam, 2010.

P. M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674, 2007.

P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer, Berlin, 2009.

P. E. Dunne. Computational properties of argument systems satisfying graph-theoretic constraints. *Artificial Intelligence*, 171(10-15):701–729, 2007.

P. E. Dunne and T. J. M. Bench-Capon. Coherence in finite argument systems. *Artificial Intelligence*, 141(1/2):187–203, 2002.

P. E. Dunne and M. Caminada. Computational complexity of semi-stable semantics in abstract argumentation frameworks. In *JELIA*, volume 5293 of *Lecture Notes in Computer Science*, pages 153–165. Springer, Berlin, 2008.

P. E. Dunne, W. Dvořák, T. Linsbichler, and S. Woltran. Characteristics of multiple viewpoints in abstract argumentation. *Artificial Intelligence*, 228:153–178, 2015.

W. Dvořák and P. E. Dunne. Computational problems in formal argumentation and their complexity. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, chapter 13, pages 631–687. College Publications, London, 2018.

W. Dvořák and J. P. Wallner. Computing strongly admissible sets. In *Proceedings of Computational Models of Argument COMMA 2020*, pages 179–190. IOS Press, Amsterdam, 2020.

W. Dvořák and S. Woltran. Complexity of semi-stable and stage semantics in argumentation frameworks. *Inf. Process. Lett.*, 110(11):425–430, 2010.

W. Dvořák, S. Ordyniak, and S. Szeider. Augmenting tractable fragments of abstract argumentation. *Artificial Intelligence*, 186:157–173, 2012.

W. Dvořák, M. Järvisalo, J. P. Wallner, and S. Woltran. Complexity-sensitive decision procedures for abstract argumentation. *Artificial Intelligence*, 206:53–78, 2014.

W. Dvořák, J. Fandinno, and S. Woltran. On the expressive power of collective attacks. *Argument & Computation*, 10(2):191–230, 2019.

W. Dvořák, A. Keshavarzi Zafarghandi, and S. Woltran. Expressiveness of SETAFs and support-free ADFs under 3-valued semantics. In *COMMA*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 191–202. IOS Press, Amsterdam, 2020.

W. Dvořák. *Computational Aspects of Abstract Argumentation*. PhD thesis, Vienna University of Technology, Institute of Information Systems, 2012.

U. Egly, S. A. Gaggl, and S. Woltran. Answer-set programming encodings for argumentation frameworks. *Argument & Computation*, 1(2):147–177, 2010.

S. Ellmauthaler. *Abstract Dialectical Frameworks; Properties, Complexity, and Implementation*. PhD thesis, Vienna University of Technology, 2012.

S. Ellmauthaler and H. Strass. The DIAMOND system for computing with abstract dialectical frameworks. In *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA 2014)*, pages 233–240. IOS Press, Amsterdam, 2014.

H. B. Enderton. *A Mathematical Introduction to Logic*. Elsevier, Amsterdam, 2001.

X. Fan and F. Toni. Assumption-based argumentation dialogues. In *Twenty-Second International Joint Conference on Artificial Intelligence*. AAAI Press, 2011.

G. Flouris and A. Bikakis. A comprehensive study of argumentation frameworks with sets of attacking arguments. *International Journal of Approximate Reasoning*, 109:55–86, 2019.

J. Fox and S. Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. The MIT Press, Cambridge, MA, 2000.

D. M. Gabbay, C. J. Hogger, and J. A. Robinson. *Handbook of Logic in Artificial Intelligence and Logic Programming: Volume 5: Logic Programming*. Clarendon Press, Oxford, 1998.

S. A. Gaggl, S. Rudolph, and H. Strass. On the computational complexity of naive-based semantics for abstract dialectical frameworks. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pages 2985–2991. AAAI Press, 2015.

S. A. Gaggl, S. Rudolph, and H. Straß. On the decomposition of abstract dialectical frameworks and the complexity of naive-based semantics. *Journal of Artificial Intelligence Research*, 70:1–64, 2021.

A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2): 95–138, 2004.

M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proceedings of the International Conference on Logic Programming, ICLP/SLP*, pages 1070–1080. The MIT Press, Cambridge, MA, 1988.

M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Gener. Comput.*, 9(3/4):365–386, 1991.

I. Gilboa. *Theory of Decision under Uncertainty*. Cambridge University Press, Cambridge and New York, NY, 2009.

F. Grasso, A. Cawsey, and R. B. Jones. Dialectical argumentation to solve conflicts in advice giving: A case study in the promotion of healthy nutrition. *Int. J. Hum. Comput. Stud.*, 53(6):1077–1115, 2000.

E. Hadoux and A. Hunter. Learning and updating user models for subpopulations in persuasive argumentation using beta distributions. In *AAMAS*, pages 1141–1149. International Foundation for Autonomous Agents and Multiagent System / ACM, 2018.

E. Hadoux and A. Hunter. Comfort or safety? Gathering and using the concerns of a participant for better persuasion. *Argument and Compututation*, 10(2):113–147, 2019.

C. L. Hamblin. Mathematical models of dialogue 1. *Theoria*, 37(2):130–155, 1971.

J. Heyninck, G. Kern-Isberner, and M. Thimm. On the correspondence between abstract dialectical frameworks and nonmonotonic conditional logics. In *FLAIRS Conference*, pages 575–580, 2020.

R. A. Howard and J. E. Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005.

A. Hunter. Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In *IJCAI*, pages 3055–3061. AAAI Press, 2015.

A. Hunter. Towards a framework for computational persuasion with applications in behaviour change. *Argument & Computation*, 9(1):15–40, 2018.

A. Hunter and M. Thimm. Probabilistic argument graphs for argumentation lotteries. In *COMMA*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 313–324. IOS Press, Amsterdam, 2014.

A. Hunter and M. Williams. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine*, 56(3): 173–190, 2012.

H. Jakobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law*, pages 53–62. ACM Press, New York, 1999.

M. Janier, M. Snaith, K. Budzynska, J. Lawrence, and C. Reed. A system for dispute mediation: The mediation dialogue game. In *COMMA*, volume 287 of *Frontiers in Artificial Intelligence and Applications*, pages 351–358. IOS Press, Amsterdam, 2016.

A. Keshavarzi Zafarghandi, R. Verbrugge, and B. Verheij. Discussion games for preferred semantics of abstract dialectical frameworks. In G. Kern-Isberner and Z. Ognjanovic, editors, *European Conference on Symbolic and Quantitative Approaches with Uncertainty*, pages 62–73. Springer, Berlin, 2019a.

A. Keshavarzi Zafarghandi, B. Verheij, and R. Verbrugge. Embedding probabilities, utilities and decisions in a generalization of abstract dialectical frameworks. In *ISIPTA*, volume 103 of *Proceedings of Machine Learning Research*, pages 246–255. PMLR, 2019b.

A. Keshavarzi Zafarghandi, R. Verbrugge, and B. Verheij. A discussion game for the grounded semantics of abstract dialectical frameworks. In *Proceedings of Computational Models of Argument COMMA 2020*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 431–442. IOS Press, Amsterdam, 2020.

A. Keshavarzi Zafarghandi, W. Dvořák, R. Verbrugge, and B. Verheij. Computational complexity of strong admissibility for abstract dialectical frameworks. In *19th International Workshop on Non-Monotonic Reasoning (NMR)*, pages 295–304, 2021a.

A. Keshavarzi Zafarghandi, R. Verbrugge, and B. Verheij. Strong admissibility for abstract dialectical frameworks. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing SAC '21*, pages 873–880, 2021b.

A. Keshavarzi Zafarghandi, R. Verbrugge, and B. Verheij. Semi-stable semantics for abstract dialectical frameworks. In *Proceeedings 18th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 422–431, 2021c.

A. Keshavarzi Zafarghandi, R. Verbrugge, and B. Verheij. Strong admissibility for abstract dialectical frameworks. *Argument & Computation*, page online first, 2021d.

E. C. Krabbe. Dialogue logic. In *Handbook of the History of Logic*, volume 7, pages 665–704. Elsevier, Amsterdam, 2006.

J. Lawrence and C. Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2020.

G. W. Leibniz. *The Art of Discovery.* Wiener 51, 1685.

T. Linsbichler. *Advances in Abstract Argumentation; Expressiveness and Dynamics.* PhD thesis, Vienna University of Technology, Institute of Information Systems, 2017.

T. Linsbichler, J. Pührer, and H. Strass. A uniform account of realizability in abstract argumentation. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, pages 252–260. IOS Press, Amsterdam, 2016.

T. Linsbichler, M. Maratea, A. Niskanen, J. P. Wallner, and S. Woltran. Novel algorithms for abstract dialectical frameworks based on complexity analysis of subclasses and SAT solving. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*, pages 1905–1911. ijcai.org, 2018.

M. Lippi and P. Torroni. Context-independent claim detection for argument mining. In *IJCAI*, pages 185–191. AAAI Press, 2015.

M. Lippi and P. Torroni. Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Techn.*, 16(2):10:1–10:25, 2016.

P. Lorenzen and K. Lorenz. *Dialogische Logik.* Wissenschaftliche Buchgesellschaft, Darmstadt, 1978.

J. Mackenzie. Four dialogue systems. *Studia logica*, 49(4):567–583, 1990.

J. D. Mackenzie. Question-begging in non-cumulative systems. *Journal of philosophical logic*, 8(1):117–133, 1979.

J. Macoubrie. Logical argument structures in decision-making. *Argumentation*, 17(3):291–313, 2003.

P. McBurney, S. Parsons, and I. Rahwan, editors. *Argumentation in Multi-Agent Systems - 8th International Workshop, ArgMAS 2011, Taipei, Taiwan, May 3, 2011, Revised Selected Papers*, volume 7543 of *Lecture Notes in Computer Science*. Springer, Berlin, 2012.

S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In G. R. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 105–129. Springer, Berlin, 2009.

S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014.

P. Mongin. Expected utility theory. In J. B. Davis, D. W. Hands, and U. Maki, editors, *The Handbook of Economic Methodology*, pages 171–178. Edward Elgar Publishing, Cheltenham, 1998.

D. Neugebauer. Generating defeasible knowledge bases from real-world argumentations using D-BAS. In *Proceedings of the 1st Workshop on Advances In Argumentation In Artificial Intelligence co-located with XVI International Conference of the Italian Association for Artificial Intelligence*, pages 105–110. CEUR-WS.org, 2017.

D. Neugebauer. *Bridging the Gap between Online Discussions and Formal Models of Argumentation*. PhD thesis, University of Düsseldorf, Germany, 2019.

S. H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Argumentation in Multi-Agent Systems (ArgMAS 2006)*, pages 54–73. Springer, Berlin, 2006.

S. Nofal, K. Atkinson, and P. E. Dunne. Algorithms for decision problems in argument systems under preferred semantics. *Artificial Intelligence*, 207:23–51, 2014.

F. Nouioua. AFs with necessities: Further semantics and labelling characterization. In *SUM*, volume 8078 of *Lecture Notes in Computer Science*, pages 120–133. Springer, Berlin, 2013.

S. M. Olmsted. *On Representing and Solving Decision Problems.* PhD thesis, Stanford University, US, 1985.

N. Oren, C. Reed, and M. Luck. Moving between argumentation frameworks. In *COMMA*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 379–390. IOS Press, Amsterdam, 2010.

C. H. Papadimitriou. *Computational Complexity*. Academic Internet Publications, 2007.

S. Polberg. Understanding the abstract dialectical framework. In *Logics in Artificial Intelligence - 15th European Conference (JELIA 2016)*, pages 430–446. Springer, Berlin, 2016.

S. Polberg. *Developing the Abstract Dialectical Framework*. PhD thesis, Vienna University of Technology, Institute of Information Systems, 2017.

S. Polberg, J. P. Wallner, and S. Woltran. Admissibility in the abstract dialectical framework. In *CLIMA*, volume 8143 of *Lecture Notes in Computer Science*, pages 102–118. Springer, Berlin, 2013.

J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.

J. L. Pollock. *Cognitive Carpentry: A Blueprint for how to Build a Person*. MIT Press, Cambridge, MA, 1995.

N. Potyka, S. Polberg, and A. Hunter. Polynomial-time updates of epistemic states in a fragment of probabilistic epistemic argumentation. In *ECSQARU*, volume 11726 of *Lecture Notes in Computer Science*, pages 74–86. Springer, Berlin, 2019.

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

H. Prakken. Historical overview of formal argumentation. *FLAP*, 4(8), 2017.

H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7(1-2): 25–75, 1997.

H. Prakken and G. Sartor. Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214–245, 2015.

G. Priest. *An introduction to non-classical logic: From if to is*. Cambridge University Press, 2008.

J. Pührer. Realizability of three-valued semantics for abstract dialectical frameworks. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, pages 3171–3177. AAAI Press, 2015.

J. Pührer. ArgueApply: A mobile app for argumentation. In *Proceedings of the 14th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR 2017)*, pages 250–262. Springer, Berlin, 2017.

J. Pührer. Realizability of three-valued semantics for abstract dialectical frameworks. *Artificial Intelligence*, 278, 2020a.

J. Pührer. Realizability of three-valued semantics for abstract dialectical frameworks. *Artificial Intelligence*, 278, 2020b.

F. P. Ramsey. Truth and probability. In H. Arlo-Costa, V. F. Hendricks, and J. van Benthem, editors, *Readings in Formal Epistemology*, pages 21–45. Springer, Berlin, 2016.

C. Reed and T. J. Norman, editors. *Argumentation Machines, New Frontiers in Argument and Computation*, volume 9 of *Argumentation Library*. Springer, Berlin, 2004.

R. Reiter. A logic for default reasoning. *Artificial intelligence*, 13(1-2): 81–132, 1980.

E. L. Rissland, K. D. Ashley, and R. P. Loui. AI and law: A fruitful synergy. *Artificial Intelligence*, 150(1-2):1–15, 2003.

S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, New Jersey, NJ, 3rd edition, 2009.

L. Savage. *The Foundations of Statistics*. Wiley, New York, NY, 1954.

A. Sen. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, pages 317–344, 1977.

M. J. Sergot, F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond, and H. T. Cory. The british nationality act as a logic program. *Communications of the ACM*, 29(5):370–386, 1986.

R. D. Shachter. Evaluating influence diagrams. *Operations research*, 34(6): 871–882, 1986.

H. Sidgwick. *The Methods of Ethics*. Hackett Publishing, London, 1981. First Edition 1874.

N. Slonim, Y. Bilu, C. Alzate, R. Bar-Haim, B. Bogin, F. Bonin, L. Choshen, E. Cohen-Karlik, L. Dankin, L. Edelstein, et al. An autonomous debating system. *Nature*, 591(7850):379–384, 2021.

L. J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.

H. Strass. Instantiating knowledge bases in abstract dialectical frameworks. In *CLIMA*, volume 8143 of *Lecture Notes in Computer Science*, pages 86–101. Springer, Berlin, 2013a.

H. Strass. Approximating operators and semantics for abstract dialectical frameworks. *Artificial Intelligence*, 205:39–70, 2013b.

H. Strass. Implementing instantiation of knowledge bases in argumentation frameworks. In *COMMA*, pages 475–476. IOS Press, Amsterdam, 2014.

H. Strass. Expressiveness of two-valued semantics for abstract dialectical frameworks. *Journal of Artificial Intelligence Research*, 54:193–231, 2015.

H. Strass. Instantiating rule-based defeasible theories in abstract dialectical frameworks and beyond. *Journal of Logic and Computation*, 28(3):605–627, 2018.

H. Strass and S. Ellmauthaler. goDIAMOND 0.6.6 – ICCMA 2017 System Description, 2017. Available at `http://argumentationcompetition.org/2017/goDIAMOND.pdf`.

H. Strass and J. P. Wallner. Analyzing the computational complexity of abstract dialectical frameworks via approximation fixpoint theory. *Artificial Intelligence*, 226:34–74, 2015.

P. M. Thang, P. M. Dung, and N. D. Hung. Towards a common framework for dialectical proof procedures in abstract argumentation. *Journal of Logic and Computation*, 19(6):1071–1109, 2009.

M. Thimm and S. Villata. The first international competition on computational models of argumentation: Results and analysis. *Artificial Intelligence*, 252:267–294, 2017.

F. Toni. Reasoning on the web with assumption-based argumentation. In *Reasoning Web*, volume 7487 of *Lecture Notes in Computer Science*, pages 370–386. Springer, Berlin, 2012.

F. Toni. A tutorial on assumption-based argumentation. *Argument & Computation*, 5(1):89–117, 2014.

S. E. Toulmin. *The Uses of Argument.* Cambridge University Press, Cambridge, 2003.

S. E. Toulmin, R. D. Rieke, and A. Janik. *An Introduction to Reasoning.* Macmillan, New York, 1984.

F. H. van Eemeren and B. Verheij. Argumentation theory in formal and computational perspective. *IFCoLog Journal of Logics and Their Applications FLAP*, 4(8), 2017.

F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. S. Henkemans, B. Verheij, and J. H. M. Wagemans, editors. *Handbook of Argumentation Theory.* Springer, New York, 2014.

B. Verheij. Two approaches to dialectical argumentation: Admissible sets and argumentation stages. *Proceedings NAIC 1996*, pages 357–368, 1996.

B. Verheij. Dialectical argumentation with argumentation schemes: An approach to legal logic. *Artificial Intelligence and Law*, 11(2-3):167–195, 2003a.

B. Verheij. DefLog: On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13(3):319–346, 2003b.

B. Verheij. A labeling approach to the computation of credulous acceptance in argumentation. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence IJCAI*, pages 623–628. AAAI Press, 2007.

B. Verheij. The Toulmin argument model in artificial intelligence. or: How semi-formal, defeasible argumentation schemes creep into logic. In G. R. Simari and I. Rahwan, editors, *Argumentation in artificial intelligence*, pages 219–238. Springer, Berlin, 2009.

B. Verheij. Arguments for ethical systems design. In *JURIX*, volume 294 of *Frontiers in Artificial Intelligence and Applications*, pages 101–110. IOS Press, Amsterdam, 2016a.

B. Verheij. Formalizing value-guided argumentation for ethical systems design. *Artificial Intelligence and Law*, 24(4):387–407, 2016b.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, 24:285–324, 2016.

J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 2nd edition, 1947.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 2007. 60th anniversary edition, with an introduction by Harold W. Kuhn and an afterword by Ariel Rubinstein.

G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceeding of the 7th European Workshop on Logic for Artificial Intelligence JELIA*, volume 1919, pages 239–253. Springer, Berlin, 2000.

J. P. Wallner. *Complexity Results and Algorithms for Argumentation-Dung's Frameworks and Beyond*. PhD thesis, Vienna University of Technology, Institute of Information Systems, 2014.

J. P. Wallner. Structural constraints for dynamic operators in abstract argumentation. *Argument & Computation*, 11(1-2):151–190, 2020.

D. Walton and E. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, 1995.

D. N. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, 2008.

B. Yun, S. Vesic, and M. Croitoru. Toward a more efficient generation of structured argumentation graphs. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018*, Frontiers in Artificial Intelligence and Applications, pages 205–212. IOS Press, Amsterdam, 2018.

# Summary

Argumentation is an essential part of our daily life both in our individual and our social activities. Arguing is so natural for all of us that we do it all the time, either with ourselves or with other people. One of the strengths of the argumentation approach is that it turns out to be powerful enough to model a wide range of formalisms for non-monotonic reasoning. Argumentation is deeply rooted in human history, and the academic study of argumentation goes back to ancient Greece in theoretical philosophy. Reasoning via argumentation has been a specific topic in philosophy at least since the time of Aristotle. The extensive trajectory of research on argumentation from Aristotle to today's computational argumentation in artificial intelligence shows how far research in argumentation has come.

We argue for different reasons. For instance, when we have an important decision to make, we may discuss it with other people in order to consider their ideas, we deliberate about it in our mind several times, and we may also use an automated system such as an argument-assistance system that simulates human reasoning. In any event, a fruitful way of progress towards a decision is through argumentation.

According to Leibniz, "the only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: Let us calculate [calculemus], without further ado, to see who is right". Put differently, developing automated methods capturing the human ability of reasoning is an old, ambitious, and ongoing research goal. Big dreams bring extraordinary results. In modern terms, one would rephrase Leibniz' dream as the aim to design a formal system and a decision procedure for making a decision without any doubt.

With the advent and development of technology, we see that different forms of argumentation can occur between a human and an autonomous system, for instance, when a person uses a chatbot, smartphone voice assistant, or automated persuasion system. Recently, even an autonomous

debating system has been developed that can perform a debate with a human expert debater. A crucial question is: How should the process of arguing occur among automated systems? To empower automated systems to argue, solid formalisms are required for modeling and evaluating argumentation.

Argumentation theory can shed light on the process of decision making, from modeling to evaluating a problem. Models of argumentation reflect how arguments relate to one another, and semantics of models of argumentation reflect how to use argumentation for making a decision under inconsistent, controversial, and incomplete information.

In this thesis we consider abstract dialectical frameworks (ADFs for short), one of the powerful formalisms of argumentation that allow arbitrary logical relationships among arguments to be expressed. An ADF can be represented by a directed graph in which nodes indicate arguments and links show the relation among arguments. Each argument in an ADF is labeled by a propositional formula, called acceptance condition. The acceptance condition of each argument expresses under which condition the argument can be accepted. The semantics of ADFs are methods proposed to evaluate the acceptance of the arguments.

We begin by focusing on the semantical evaluation of ADFs, presenting two novel semantics. Among all admissibility-based semantics of ADFs, grounded semantics stands out for its important features. To target explaining why an argument is justified under the grounded semantics in ADFs, in the current thesis we have proposed the notion of strong admissibility semantics of ADFs. Furthermore, we analyzed the complexity of the relevant reasoning tasks of ADFs under the strong admissibility semantics.

For the cases in which a given ADF does not have any two-valued model, we introduce semi-two-valued semantics and semi-stable semantics of ADFs as new points of view on the acceptance of arguments. Both are proper formal generalizations of the notion of the semi-stable semantics of AFs to ADFs. In ADFs, the user can choose whether support cycles should be accepted or rejected, by choosing semi-two-valued models or semi-stable models as semantics.

A more procedural view of ADF semantics, in particular to decide acceptability of arguments, underlies work on discussion games. Discussion games can provide an explanation of why an argument of interest is to be accepted or denied according to a given semantics. Therefore, discussion games can be regarded as intuitive means of determining acceptance of

arguments accompanying the formal definitions of the semantics to an ADF at hand. This was our motivation to introduce the first discussion games for ADFs in the next step of this thesis. Particularly we do so for the grounded and preferred semantics.

The high computational complexity of reasoning tasks for ADFs was our motivation to introduce subclasses of ADFs. We investigate how the restrictions that we consider influence the semantic evaluation of such ADFs. To determine the capabilities of knowledge representation formalisms, we study their expressivity: properties that characterize the outcomes of evaluating the formalisms via their associated semantics. Furthermore, we study the sufficient and necessary conditions for rewriting an ADF as a SETAF. We show that an ADF without any support link can be encoded as a SETAF if and only if all acceptance conditions are satisfiable.

Next, we combine argumentation with decision theory in the context of ADFs in order to model expected utility problems. We present a generalization of ADFs, called numerical ADFs. The formalism of numerical ADFs can be used to model and evaluate the standard decision problem of expected utility theory, which is a theory concerned with making the best decision under uncertainty in a single-agent system.

With this work, we hope that we have advanced the knowledge on the field of formal argumentation and in particular that we have given new insights into the semantics of abstract dialectical frameworks that reflect how to use argumentation for making a decision under inconsistent, controversial, and incomplete information.

# Samenvatting

Argumentatie maakt een essentieel deel uit van ons dagelijks leven, zowel in onze individuele als in onze sociale bezigheden. Argumenteren is zo natuurlijk voor ieder van ons dat we er continu mee bezig zijn, hetzij met onszelf, hetzij met andere mensen. Een van de sterke punten van de argumentatiebenadering is dat deze benadering krachtig genoeg blijkt te zijn voor het modelleren van een groot scala aan formalismen voor niet-monotoon redeneren. Argumentatie is diep geworteld in de menselijke geschiedenis, en de academische studie van argumentatieleer gaat in de theoretische filosofie terug tot het oude Griekenland. Redeneren via argumentatie is ten minste vanaf de tijd van Aristoteles een specifiek onderwerp in de filosofie geweest. Het uitgebreide spoor van onderzoek over argumentatie, vanaf Aristoteles tot aan de computationele argumentatie van vandaag in de kunstmatige intelligentie, laat zien hoe ver het onderzoek in de argumentatieleer is gekomen.

We argumenteren om verschillende redenen. Als we bijvoorbeeld een belangrijke beslissing moeten nemen kunnen we die bespreken met anderen om hun ideeën te overwegen, we kunnen die verschillende keren in gedachten overwegen, en we kunnen zelfs een automatisch systeem gebruiken, zoals een 'argument-assistentie'-systeem dat menselijk redeneren simuleert. In al deze gevallen is argumentatie een vruchtbare manier om tot een beslissing te komen.

Zoals Leibniz zei: 'De enige manier om onze redeneringen te rechtvaardigen is om ze net zo tastbaar te maken als die van de wiskundigen, zodat we onze fouten in een oogopslag kunnen vinden, en wanneer er onenigheden zijn tussen personen we gewoon kunnen zeggen: "Laat ons berekenen [calculemus], zonder verder gedoe, wie er gelijk heeft". Met andere woorden, het ontwikkelen van geautomatiseerde methoden om het menselijke vermogen tot redeneren te vatten is een oud, ambitieus en voortdurend onderzoeksdoel. Grote dromen leiden tot buitengewone resultaten. In moderne bewoordingen zou men Leibniz' droom omschrijven als het

streven een formeel systeem en een besluitvormingsprocedure te ontwerpen om zonder enige twijfel besluiten te nemen.

Met de komst en ontwikkeling van deze technologie zien we dat zich verschillende vormen van argumentatie kunnen voordoen tussen een mens en een autonoom systeem, bijvoorbeeld als iemand een chatbot gebruikt, of een smartphone-spraakassistent of een geautomatiseerd overtuigingssysteem. Kort geleden is er zelfs een autonoom debatsysteem ontwikkeld dat een debat kan voeren met een menselijke kampioendebatteerder. Een cruciale vraag is hier: Hoe zou het argumentatieproces verlopen tussen geautomatiseerde systemen? Om geautomatiseerde systemen in staat te stellen te argumenteren zijn er degelijke formalismen nodig om argumentatie te modelleren en te evalueren.

Argumentatietheorie kan licht werpen op het besluitvormingsproces, vanaf het modelleren tot het evalueren van een probleem. Argumentatiemodellen weerspiegelen hoe argumenten zich tot elkaar verhouden, en semantiek van argumentatiemodellen laat zien hoe men argumentatie zou moeten gebruiken om in het geval van inconsistente, controversiële en onvolledige informatie een beslissing te nemen.

In dit proefschrift kijken we naar abstracte dialectische raamwerken ('abstract dialectical frameworks', of ADF's), een krachtig formalisme van de argumentatieleer dat het uitdrukken van willekeurige logische verbanden tussen argumenten mogelijk maakt. Een ADF kan weergegeven worden als een gerichte graaf waarin knopen argumenten voorstellen en pijlen het verband tussen argumenten laten zien. Elk argument in een ADF is gelabeld met een propositionele formule, die de acceptatieconditie wordt genoemd. De acceptatieconditie van elk argument drukt uit onder welke voorwaarde het argument geaccepteerd kan worden. Een semantiek voor ADF's is een methode die gebruikt wordt om de acceptatie van de argumenten te evalueren.

Om te beginnen focussen we op de semantische evaluatie van ADF's, waarbij we twee nieuwe semantieken presenteren. Onder de op toelaatbaarheid gebaseerde semantieken van ADF's vallen gegronde semantieken op vanwege hun belangrijke eigenschappen. Om uit te kunnen leggen waarom een argument gerechtvaardigd is onder de gegronde semantiek in ADF's hebben we in dit proefschrift de notie van sterke toelaatbaarheidssemantiek van ADF's voorgesteld. Verder hebben we de complexiteit van de relevante redeneertaken van ADF's onder de sterke toelaatbaarheidssemantiek geanalyseerd.

Voor de gevallen waarin een gegeven ADF geen enkel tweewaardig model

heeft introduceren we semi-tweewaardige semantiek en semi-stabiele semantiek van ADF's als nieuwe zienswijzen op de acceptatie van argumenten. Beide zijn correcte formele generalisaties van het idee van semi-stabiele uitbreidingen voor AF's naar ADF's. In ADF's kan de gebruiker door semi-tweewaardige modellen of semi-stabiele modellen als semantiek te kiezen, besluiten of ondersteuningscycli geaccepteerd of afgewezen moeten worden.

Een meer procedurele visie op ADF-semantiek, in het bijzonder om de aanvaardbaarheid van argumenten te bepalen, ligt ten grondslag aan onderzoek over discussiespellen. Discussiespellen kunnen uitleggen waarom een belangrijk argument geaccepteerd of geweigerd moet worden volgens een gegeven semantiek. Daarom kunnen discussiespellen beschouwd worden als een intuïtieve manier om de toelaatbaarheid van argumenten in een ADF ten opzichte van een gegeven semantiek ADF te bepalen. Dit was onze motivatie om in de volgende stap van dit proefschrift de eerste discussiespellen voor ADF's te introduceren. Dit doen we in het bijzonder voor de gegronde en geprefereerde semantiek.

De hoge computationele complexiteit van de redeneertaken voor ADF's was onze motivatie om deelklassen van ADF's te introduceren. We onderzoeken hoe de restricties die we beschouwen de semantische evaluatie van zulke ADF's beïnvloeden. Om de mogelijkheden van kennisrepresentatieformalismen te bepalen bestuderen we hun expressiviteit. Verder bestuderen we de voldoende en noodzakelijke voorwaarden om een ADF als een SETAF te herschrijven. We laten zien dat een ADF zonder enige ondersteuningslink als een SETAF gecodeerd kan worden dan en slechts dan als alle acceptatievoorwaarden vervulbaar zijn.

Vervolgens combineren we argumentatieleer met beslistheorie in de context van ADF's om verwachtenutsproblemen te modelleren. We stellen numerieke ADF's voor als een generalisatie van ADF's. Het formalisme van numerieke ADF's kan gebruikt worden voor het modelleren en evalueren van het standaard probleem uit de beslistheorie, een theorie die zich bezig houdt met het maken van de beste beslissing in onzekere omstandigheden in een systeem met één agent.

We hopen dat we met dit werk de kennis op het gebied van formele argumentatie bevorderd hebben, en in het bijzonder dat we nieuwe inzichten hebben gegeven in de semantiek van abstracte dialectische raamwerken (ADF's) die weerspiegelen hoe argumentatie gebruikt kan worden om beslissingen te nemen onder inconsistente, controversiële en onvolledige informatie.