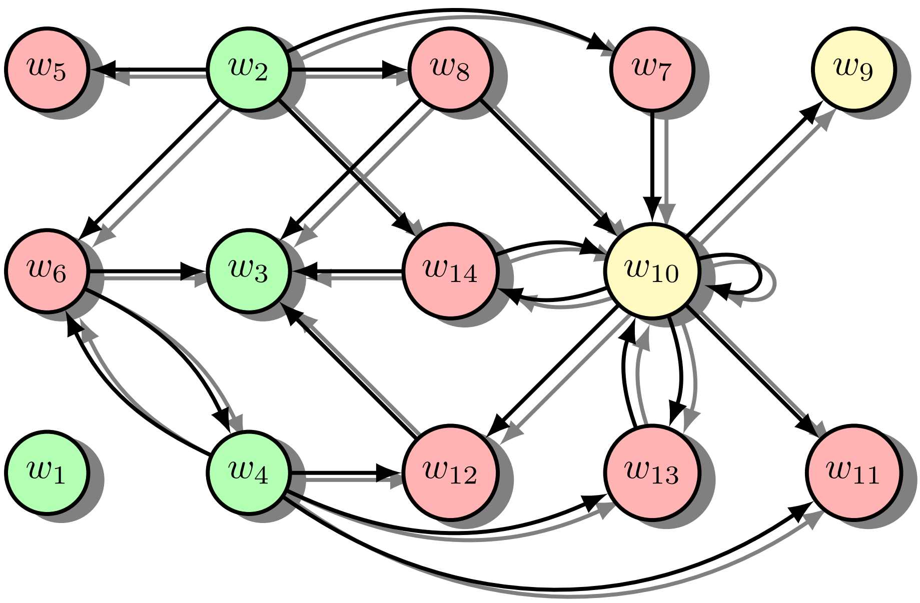
# Extension-based Semantics of Argumentation Frameworks for Agent Interactions




**Cristian Gratie**

# Extension-based Semantics of Argumentation Frameworks for Agent Interactions

# Extensiegebaseerde semantiek van argumentatiesystemen voor agent-interacties

(met een samenvatting in het Nederlands)

# Semantici bazate pe extensii ale sistemelor de argumentare pentru interacţiunile dintre agenţi

(cu un rezumat în limba română)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties in het openbaar
te verdedigen op woensdag 5 december 2012 des middags te 4.15 uur

door

**Cristian Gratie**

geboren op 26 november 1986 te Iaşi, Roemenië

| | |
|---|---|
| Promotoren: | Prof. dr. J.-J. Ch. Meyer |
| | Prof. dr. A. M. Florea |
| Co-promotor: | Dr. G. A. W. Vreeswijk |

*to my grandmother*

# Contents

# Acknowledgements

After five years of bachelor courses with very diverse subjects, the prospect of spending three years on a single topic is a different kind of challenge. A little more than three years ago, and only a few months after my graduation, I already knew that I would take this challenge, yet I was still somewhat undecided about the actual direction to follow. Then I attended the European Agent Systems Summer School in Turin, Italy. Among many other interesting courses that have broadened my understanding of agent systems, the one dedicated to argumentation, given by Sanjay Modgil, stood out for me and decided the topic I would pursue. Thank you, Sanjay.

The journey has been long and has not always taken me where I thought it would, but the new "places" I have seen proved to be more interesting then the ones I had initially set out for. During this journey, a lot of people have helped me, directly or indirectly, sometimes even unknowingly, and I thank them all for this. First of all, I would like to thank my advisers: Adina Magda Florea, from University "Politehnica" of Bucharest, John-Jules Meyer and Gerard Vreeswijk, from Utrecht University. This work would not have been possible without their constant support and guidance.

I thank Adina especially for her guidance on how to do research and for the valuable knowledge that she has passed on to me, also before the start of my PhD, as my diploma project adviser. She has managed to find a good balance between, on one hand, giving me enough freedom to test my own ideas and, on the other hand, filtering out my bad ideas and not letting me follow ten directions at once. I thank John-Jules especially for the useful and prompt feedback that he gave me whenever I needed it, also for following my proofs and challenging the way I presented them or the storyline of the thesis. Furthermore, I thank him for helping me put everything together and for guiding me towards a more realistic schedule for writing the thesis. Also, I am grateful to Gerard for having guided my way towards a better understanding of modal logic and for helping me make the transition from my initial, somewhat naive and literature-independent ideas for describing semantics to the modal approach.

Next, I would like to thank the members of the reading committee: Wiebe van der Hoek from Liverpool University, United Kingdom, Henry Prakken from Utrecht University, Netherlands, Olivier Boissier from Ecole Nationale Superieure des Mines de Saint-Etienne, France, and Costin Bădică from University of Craiova, Romania. Their useful comments and suggestions have helped improve the overall quality of this thesis.

I have not been in Utrecht for long, but the time I have spent there was most productive. I would like to thank all the members of the Intelligent Systems group for making me feel welcome there. Special thanks go to Paolo Turrini for helping me bypass some misconceptions I used to have about the normal forms for global modalities. Furthermore, special thanks go to Max Knobbout and Eric Kok, for sharing their offices with me and also for the discussions we had about our theses. Also, I thank Edith Stap for managing the logistics of my stay in Utrecht. Last, but not least, I would like to thank my flatmates from Janshovenstraat 12 bis, Iulian, Patryk and Angie, for the encouragements that have helped me through the actual writing of the thesis and also for the fun times that we have spent together.

This work, and also the two research mobilities at Utrecht University, would have been impossible without the financial support that I have received these three years. First of all, I have been supported by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/88/1.5/S/61178. Also, I would mention the ERRIC project (Empowering Romanian Research on Intelligent Information Technologies), number 264207/FP7-REGPOT-2010-1, for funding my research mobility to Ecole Nationale Superieure des Mines de Saint-Etienne, France, where most ideas presented in the agent-related chapter of the thesis were consolidated. I would like to thank Mihaela Puică and Iulia Măriuţa for helping me with logistical aspects related to these two projects while I was abroad.

I am also grateful to COST Action IC0801 on Agreement Technologies for facilitating the dissemination of results and for promoting networking among researchers interested in topics related to agreement technologies. It has been very fortunate for me that this Action included a special working group dedicated to argumentation.

I would like to thank all the members of the Artificial Intelligence and Multi-Agent Systems group from my university for their support. Particularly, I thank Andrei Mogoş for helping me improve my paper writing and oral presentation skills.

I would also like to thank all my friends, for understanding the "I'm busy working for my PhD" line, but also for not accepting this excuse all the time, for asking me about the status of my research, for worrying about my sleep pattern and for many other little things that have helped me through the more difficult times. It is impossible to give an exhaustive list, but I should mention, at least, Mihaela Puică, Bogdan Iancu and Alina Sbîrlea.

Last, but certainly not least, I would like to mention the persons closest and dearest to me. I thank my father Constantin, my mother Zoiţa, my sister Diana-Elena and my grandmother Nina, for their constant support on all levels, throughout my doctoral studies, as well as throughout my whole life. I also thank Mădălina, for being by my side these years.

Bucharest, 7 September 2012

*Chapter 1*

# Introduction

Argumentation plays an important part in human interaction, whether it is used for persuasion, negotiation or simply for sharing one's point of view on a certain topic. Even at the abstract level, choosing the acceptable arguments from a given set of conflicting arguments is a challenging problem, one that was given multiple solutions in the argumentation literature, in the form of argumentation semantics.

While some of the argumentation semantics have been extensively studied and compared to one another, others have not been included in many surveys. Our aim is to consider an extended array of semantics and deal with the following issues:

- the comparison of argumentation semantics with respect to their properties
- the modal expressive power that is needed in order to describe argumentation semantics
- argumentation semantics that can cope with dynamic environments

We rely on the abstract approach to argumentation proposed in (Dung, 1995), where the central concept is that of argumentation frameworks, in essence directed graphs in which nodes are arguments and the arrows encode the attack relation between arguments.

> The main goal of abstract argumentation is "to give an analysis of the nature of human argumentation in its full generality. This is done in two steps. In the first step, a formal, abstract but simple theory of argumentation is developed to capture the notion of acceptability of arguments. In the next step, we demonstrate the 'correctness' (or 'appropriateness') of our theory. It is clear that the 'correctness' of our theory cannot be 'proved' formally. The only way to accomplish this task is to provide relevant and convincing examples." (Dung, 1995, p. 324)

Dung's work and subsequent research has shown that this abstract approach can model logic programming, nonmonotonic and defeasible reasoning approaches from Artificial Intelligence, as well as social and economic problems. The most important question in abstract argumentation is whether an argument is acceptable or not.

> "Arguments distinguish themselves from proofs by the fact that they are defeasible, that is, the validity of their conclusions can be disputed by other arguments. Whether a claim can be accepted therefore depends not only on the existence of an argument that supports this claim, but also on the existence of possible counter arguments, that can then themselves be attacked by counter arguments, etc." (Baroni et al., 2011a).

Thus, the acceptability of arguments cannot be decided in isolation, so the question becomes: which set(s) of arguments can be accepted? In the argumentation literature an answer to this question, which may be an algorithm for computing accepted sets or a list of properties that accepted sets must satisfy, is referred to as an argumentation semantics. The sets of accepted arguments are called extensions of the argumentation framework.

## 1.1 Research Questions

### 1.1.1 Properties of argumentation semantics

Many argumentation semantics have been proposed in the literature, either as a solution to problematic behavior identified for existing semantics on particular examples, or for satisfying certain desirable properties not satisfied by previous proposals. In our survey from Chapter 2 we discuss a total of 43 semantics, including also some auxiliary concepts that we have promoted as semantics.

The comparison of argumentation semantics is a challenging task, as the methods used for defining them are rather diverse. A possible solution consists in comparing them with respect to a set of desirable properties (Baroni and Giacomin, 2007). However, there are argumentation semantics that were not analyzed even with respect to some of the most elementary properties. On the other hand, we feel that there is also a need for a terminology-independent, example-based method for comparing or distinguishing semantics from one another.

Desirable properties can also be used for defining novel argumentation semantics, as is the case of resolution-based semantics (Baroni and Giacomin, 2008). While this approach has been used for several of the more intricate properties, we find it surprising that some of the basic properties were not yet used for defining semantics.

All the considerations above lead to our first set of research questions:

**Research Questions 1**: Given a fixed set of evaluation principles, what are the novel semantics that can be defined based on their satisfaction? Which of the existing semantics satisfy those principles? What are the relations between semantics and what makes an argumentation semantics unique?

We answer these questions in Chapter 3, where we also provide a graphical representation for the properties of argumentation semantics and for the relations between semantics.

### 1.1.2 Modal logic and argumentation

The use of modal logic for abstract argumentation was discussed in (Grossi, 2010), using the fact that both argumentation frameworks and Kripke models rely on assigning some meaning to directed graphs. The link between the two domains allows the transfer of techniques and theoretical results from one to the other. Furthermore, the use of modal logic as a meta-language for describing semantics can help software agents reason about and compare semantics.

In his paper, Grossi used the global modal language and managed to describe several of the classical argumentation semantics, but also identified semantics that cannot be captured by the chosen language. A stronger logic was proposed in (Grossi, 2011) in order to capture more semantics. While the use of logics (modal or not) for argumentation has also been discussed in other works, the issue of finding the right expressive power needed for describing a certain argumentation semantics has not yet received attention in the literature.

Furthermore, there is no extensive work aiming to decide the modal definability of at least the mainstream argumentation semantics with respect to very simple modal logics. Also the relation between modal formulas and the properties of argumentation semantics is insufficiently covered in the literature.

**Research Questions 2**: Given a modal language, what are the argumentation semantics that can be described using it? What is the relation between the modal formulas and the properties of the semantics they describe?

We provide an answer to these questions in Chapter 5. Our research on this topic has also generated valuable results about the normal forms and satisfiability of modal formulas containing the converse and the global modalities. These results are presented in Chapter 4.

### 1.1.3 Argumentation semantics for MAS

There is extensive research on the use of argumentation for various specific tasks within a multi-agent system, such as persuasion dialogues, decision making, belief revision and more. On the other hand, there is little research on using abstract argumentation for modeling the whole agent system. Having such a model would allow various knowledge representation languages to be plugged in without any change to the formal model or to its implementation.

Such an approach was recently proposed in (Kakas et al., 2011). In the proposed model, an agent consists of several modules and each task of a module is handled by an argumentation theory. However, the argumentation theory contains a lot more information than the abstract arguments: possible outcomes that can be supported by abstract arguments, parameters that characterize abstract arguments, a partial ordering of the parameters, a partial ordering of the abstract arguments. The approach relies in fact on value-based argumentation frameworks (Bench-Capon, 2003), an extension of the original model proposed by Dung. While the modularity of the approach avoids computational issues related to argumentation semantics, the architecture is rather intricate and the model is rather far from Dung's abstract frameworks.

**Research Questions 3**: Are there argumentation semantics that are suitable for use within MAS and that can cope with the dynamic and uncertain nature of arguments from such a system? What kind of information that pertains to the MAS can be encoded with abstract arguments? How do the changes in the system translate into changes in the underlying argumentation-based representation?

We answer these questions in Chapter 6, where we provide several new labeling-based semantics that allow an agent to identify the arguments that are to be attacked in order for its goals to be satisfied. We also discuss an argumentation-based multi-agent systems that takes advantage of the properties of these semantics.

## 1.2 Outline of the Thesis

In Chapter 2 we provide an introduction to abstract argumentation, focused on an extensive survey of argumentation semantics, covering a total of 43 semantics (including auxiliary concepts that we have promoted as semantics).

In Chapter 3 we discuss the properties of argumentation semantics. We focus on a fixed set of principles and test their satisfaction for all the 43 semantics from Chapter 2. We also introduce 5 novel principle-based semantics and discuss their properties as well. Furthermore, we provide an intuitive graphical representation of both the relations between semantics and the properties satisfied by each semantics.

In Chapter 4 we discuss the normal forms of modal languages and also their satisfaction. The important results are related to the satisfaction of formulas that contain the global and the converse modalities.

In Chapter 5 we discuss the use of modal logic for argumentation. More precisely, we discuss the normal form of formulas that describe semantics. Then we analyze the global modal language and the global converse modal language with respect to definability of argumentation semantics.

In Chapter 6 we propose several labeling-based semantics that can cope with the challenges of using abstract argumentation for multi-agent systems. The proposal is based on temporarily ignoring arguments so that the labels of the remaining arguments satisfy a certain constraint. We provide a negotiation scenario that uses the proposed labelings. We also refine the scenario into a formal model of an argumentation-based negotiation game that can serve as a starting point for the implementation of a multi-agent system using only abstract argumentation frameworks and propositional constraints.

The thesis ends with a discussion of the main contributions, grouped by research questions, in Chapter 7.

*Chapter 2*

# Argumentation Semantics

Argumentation frameworks, as proposed in (Dung, 1995), abstract away from the content of arguments (premises, conclusion, inference rules) and only rely on the attack relation between arguments in order to provide general results about the acceptability of arguments. Given an argumentation framework, in essence a directed graph, one can choose sets of arguments that can be accepted together (extensions) and can do so using various approaches, called semantics in the argumentation literature.

This chapter aims to provide an argumentation background, focused especially on argumentation semantics. In Section 2.1 we introduce argumentation frameworks (Dung, 1995), the general concept of argumentation semantics and the basic notations. The rest of the chapter provides an extensive survey of argumentation semantics. Section 2.2 describes the classical argumentation semantics, introduced together with argumentation frameworks in (Dung, 1995). We continue in Section 2.3 with the naive, stage (Verheij, 1996) and semi-stable (Caminada, 2006b) semantics. Section 2.4 is dedicated to the resolution-based semantics (Baroni and Giacomin, 2008; Baroni et al., 2011b), while in Section 2.5 we discuss justification states and the parameterized ideal semantics (Dung et al., 2007; Dvorak et al., 2011). SCC-recursiveness and the corresponding semantics (Baroni et al., 2005) are presented in Section 2.6. We cover prudent semantics (Coste-Marquis et al., 2005) in Section 2.7. Section 2.8 is dedicated to the enhanced preferred semantics. We conclude the chapter with a round-up of all the presented semantics in Section 2.9

## 2.1 Preliminaries

Argumentation frameworks, as proposed in (Dung, 1995), rely on the idea that at the most abstract level we can work just with the attack relation between arguments and forget the structured representation of the arguments. This approach can provide generic results that are meaningful when translated back to the logical representation. In his seminal paper, Dung applies abstract argumentation to nonmonotonic reasoning, logic programming and $n$-person games.

**Definition 1.** *An* ***argumentation framework*** *is a pair* $F = (\mathcal{A}, \mathcal{R})$*, where* $\mathcal{A}$ *is a finite set of arguments and* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary attack relation on* $\mathcal{A}$*. We say that an argument* $a$ ***attacks*** *an argument* $b$ *and we write this as* $a \rightarrow b$ *iff* $(a, b) \in \mathcal{R}$*. If* $(a, b) \notin \mathcal{R}$*, we say that* $a$ *does not attack* $b$ *and write this as* $a \nrightarrow b$*. We will use* $\mathcal{AF}$ *to refer to the set of all argumentation frameworks and* $\mathcal{A}rg$ *to refer to the set of argument symbols.*

In other words, an argumentation framework is simply a directed graph where the nodes are arguments and the (directed) edges are attacks between arguments. We will use lowercase letters as argument names. For an example argumentation framework see Figure 2.1. The framework is given by $F = (\mathcal{A}, \mathcal{R})$, with $\mathcal{A} = \{a, b, c\}$ and $\mathcal{R} = \{(a, b), (b, c), (c, b)\}$.

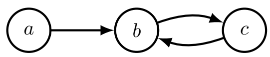



**Figure 2.1:** *A simple argumentation framework.*

It is often useful to extend the notion of attack to cover sets of arguments and be able to say that a set attacks or is attacked by an argument.

**Definition 2.** *The notion of attack and the notation can be extended to cover sets of arguments as follows:*

- *a set* $S$ *attacks an argument* $a$ *iff* $S$ *contains an argument* $b$ *that attacks* $a$
- *an argument* $a$ *attacks a set* $S$ *iff* $a$ *attacks an argument* $b$ *that is in* $S$
- *a set of arguments* $S$ *attacks a set* $T$ *iff there is an argument* $a$ *in* $S$ *and an argument* $b$ *in* $T$ *such that* $a$ *attacks* $b$*.*

$$\begin{aligned} a \rightarrow S &\Leftrightarrow \exists b(b \in S \wedge a \rightarrow b) \\ S \rightarrow a &\Leftrightarrow \exists b(b \in S \wedge b \rightarrow a) \\ S \rightarrow T &\Leftrightarrow \exists ab(a \in S \wedge b \in T \wedge a \rightarrow b) \end{aligned} \tag{2.1}$$

Note that we shall use $\Rightarrow$ for material implication, so that there is no confusion with the attacks between arguments. We will use this convention throughout the thesis.

Several extensions of the original approach have been proposed, enriching argumentation frameworks with additional information deemed useful. Bipolar argumentation frameworks (Cayrol and Lagasquie-Schiex, 2005) introduce a support relation between arguments in addition to the attack relation. In value-based argumentation frameworks (Bench-Capon, 2003) arguments are said to promote certain values and a partial ordering of values is used. Hierarchical argumentation frameworks (Modgil, 2006) allow attacks that go from an argument to the attack between two other arguments for encoding meta-arguments about preferences. Another approach using attacks between arguments and attacks is proposed in (Baroni et al., 2009) as argumentation frameworks with recursive attacks. Fuzzy argumentation frameworks (Janssen et al., 2008) use fuzzy attacks, while weighted argumentation frameworks assign a weight to each attack (Dunne et al., 2009). While such extensions are not the subject of this thesis, they were presented here so that the interested reader may find some quick references.

In order to decide whether an argument is accepted or not, one must account for the arguments that attack it and their status, which in turn depends on their attackers, and so on. This means that, in fact, one does not generally accept arguments in isolation, but decides on sets of arguments that can be accepted simultaneously. Such sets are called extensions in the argumentation literature.

So, given an argumentation framework, what are its extensions? This question is answered by several argumentation semantics, which provide the constraints that extensions must satisfy, or sometimes algorithms for computing them.

**Definition 3.** *An* ***argumentation semantics*** $\mathcal{S}em$ *can be seen as a function* $\mathcal{E}_{\mathcal{S}em} : \mathcal{AF} \to \mathcal{P}(\mathcal{P}(\mathcal{A}rg))$ *that gives, for any argumentation framework* $F = (\mathcal{A}, \mathcal{R}) \in \mathcal{AF}$*, its extensions with respect to semantics* $\mathcal{S}em$*:* $\mathcal{E}_{\mathcal{S}em}(F) \subseteq \mathcal{P}(\mathcal{P}(\mathcal{A}))$.

Note that $\mathcal{P}(X)$ denotes the subsets of $X$.

In this chapter we aim to provide an up-to-date survey of existing argumentation semantics. We have grouped semantics according to their similarities, into sections. For each section, a special argumentation framework was chosen as an example, so that most of the presented semantics provide distinct extensions.

## 2.2 Classical Semantics

The notion of classical argumentation semantics generally refers to argumentation semantics that were introduced in (Dung, 1995). We will discuss each of them and use the argumentation framework from Figure 2.2 as an example.

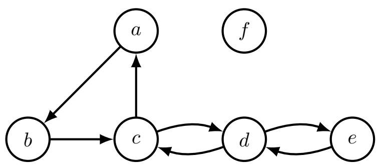



**Figure 2.2:** *Example argumentation framework for the classical semantics.*

**Definition 4.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework. A set of arguments* $S \subseteq \mathcal{A}$ *is* ***conflict-free*** *($\mathcal{CF}$) iff there are no arguments* $a$ *and* $b$ *in* $S$ *such that* $a$ *attacks* $b$*. Alternatively, using the extended notion of attack,* $S$ *is conflict-free iff* $S$ *does not attack any of its arguments.*

$$\mathcal{E}_{\mathcal{CF}}(F) = \{S \subseteq \mathcal{A} \mid \forall a (a \in S \Rightarrow S \not\to a)\} \tag{2.2}$$

For the framework from Figure 2.2 we have that $\mathcal{E}_{\mathcal{CF}}(F) = \{\emptyset, \{a\}, \{a,d\}, \{a,d,f\}, \{a,e\}, \{a,e,f\}, \{a,f\}, \{b\}, \{b,d\}, \{b,d,f\}, \{b,e\}, \{b,e,f\}, \{b,f\}, \{c\}, \{c,e\}, \{c,e,f\}, \{c,f\}, \{d\}, \{d,f\}, \{e\}, \{e,f\}, \{f\}\}$. Note that the empty set is always conflict-free. Furthermore, sets containing a single argument are also conflict-free, unless the argument attacks itself.

**Definition 5.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework. A set of arguments* $S \subseteq \mathcal{A}$ ***defends*** *an argument* $a$ *iff it attacks all the attackers of* $a$*. The set* $S$ *is*

***admissible*** *($\mathcal{AS}$) iff $S$ is conflict-free and $S$ defends all the arguments it contains.*

$$\mathcal{E}_{\mathcal{AS}}(F) = \{S \subseteq \mathcal{E}_{\mathcal{CF}}(F) \mid \forall a(a \in S \Rightarrow \forall b(b \to a \Rightarrow S \to b))\} \tag{2.3}$$

In order to determine the admissible sets for the example framework, we can examine each of the conflict-free sets and see which of them defend all their elements. It is easy to see, for example, that the conflict free set $\{b, d\}$ is not admissible, because $b$ is attacked by $a$ but neither $b$ nor $d$ attacks $a$. The set $\{a, d\}$, on the other hand, is admissible because $d$ defends itself against all its attackers and also defends $a$ against $c$. In the end we get $\mathcal{E}_{\mathcal{AS}}(F) = \{\emptyset, \{a, d\}, \{a, d, f\}, \{d\}, \{d, f\}, \{e\}, \{e, f\}, \{f\}\}$. Note that the empty set is always admissible, as it cannot be attacked.

**Definition 6.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S \subseteq \mathcal{A}$ is a* ***complete extension*** *($\mathcal{CO}$) of $F$ iff $S$ is admissible and all arguments that are defended by $S$ are in $S$.*

$$\mathcal{E}_{\mathcal{CO}}(F) = \{S \in \mathcal{E}_{\mathcal{AS}}(F) \mid \forall a(\forall b(b \to a \Rightarrow S \to b) \Rightarrow a \in S)\} \tag{2.4}$$

Alternatively, we can define complete extensions as conflict-free sets $S$ such that for any argument $a$, $a$ is in $S$ iff $S$ defends $a$.

In order to determine the complete extensions of our example framework, we look at the admissible sets and check whether they are also complete extensions. Note that the empty set is not complete in this case, because $f$ has no attackers and, thus, is defended by $\emptyset$. In fact, $f$ is defended by any set of arguments, so none of the admissible sets that do not contain $f$ is complete. Furthermore, the admissible set $\{d, f\}$ is not complete either, because it defends $a$. The other admissible sets are complete, so we have $\mathcal{E}_{\mathcal{CO}} = \{\{a, d, f\}, \{e, f\}, \{f\}\}$.

**Definition 7.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S$ is a* ***preferred extension*** *($\mathcal{PR}$) of $F$ iff $S$ is a maximal (with respect to set inclusion) admissible set.*

$$\mathcal{E}_{\mathcal{PR}}(F) = \{S \in \mathcal{E}_{\mathcal{AS}}(F) \mid \forall S'(S' \in \mathcal{E}_{\mathcal{AS}}(F) \wedge S \subseteq S' \Rightarrow S = S')\} \tag{2.5}$$

Note that in (2.5) we have formally written the maximality condition. Due to space constraints and also for improving the readability of the formulas, we will prefer the following formulation:

$$\mathcal{E}_{\mathcal{PR}}(F) = \{S \in \mathcal{E}_{\mathcal{AS}}(F) \mid S \text{ is maximal w.r.t. } \subseteq\} \tag{2.6}$$

Looking at the admissible sets of our example framework, it is not difficult to see that $\mathcal{E}_{\mathcal{PR}}(F) = \{\{a, d, f\}, \{e, f\}\}$. Also note that, since the empty set is always admissible, there will always exist at least one preferred extension.

**Definition 8.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S \subseteq \mathcal{A}$ is a* ***stable extension*** *($\mathcal{ST}$) of $F$ iff $S$ is conflict-free and $S$ attacks all the arguments it does not contain.*

$$\mathcal{E}_{\mathcal{ST}}(F) = \{S \in \mathcal{E}_{\mathcal{CF}}(F) \mid \forall a(a \in \mathcal{A} \setminus S \Rightarrow S \to a)\} \tag{2.7}$$

Instead of testing all conflict-free sets, we can use the fact that all stable extensions are also preferred (Dung, 1995). Thus, for computing the stable extensions of the example framework it is enough to test the two preferred extensions. We see that $\{e, f\}$ is not stable, as it does not attack any of the arguments $a$, $b$ and $c$. For $\{a, d, f\}$, on the other hand, we have that $a$ attacks $b$, while $d$ attacks both $c$ and $e$, so the set is stable. We get that $\mathcal{E}_{\mathcal{ST}}(F) = \{\{a, d, f\}\}$.

The main downside of the stable semantics is that it does not always provide extensions. Indeed, let us consider a simple argumentation framework consisting of a cycle of length 3: $F' = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$. The conflict-free sets of this framework are $\mathcal{E}_{\mathcal{CF}}(F') = \{\varnothing, \{a\}, \{b\}, \{c\}\}$. The empty set cannot be stable, as it does not attack any argument. The singleton sets only attack one of the remaining two arguments, so they are not stable either. Thus, there is no stable extension for $F'$.

**Definition 9.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. The* ***characteristic function*** *of $F$, $\mathcal{F}_F : \mathcal{P}(\mathcal{A}) \to \mathcal{P}(\mathcal{A})$, is the function that gives, for a set of arguments $S \subseteq A$, the set of arguments defended by $S$. The* ***grounded extension*** *($\mathcal{GR}$) of $F$ is the least fixed point of the characteristic function.*

$$\begin{aligned} \mathcal{F}_F(S) &= \{a \in \mathcal{A} \mid \forall b (b \in A \wedge b \to a \Rightarrow S \to b)\} \\ \mathcal{E}_{\mathcal{GR}}(F) &= \{S \subseteq A \mid \mathcal{F}_F(S) = S, S \textit{ is maximal w.r.t. } \subseteq\} \end{aligned} \tag{2.8}$$

It is shown in (Dung, 1995) that the grounded extension exists and is unique. Moreover, the grounded extension is also the minimal complete extension. Since we already have the complete extensions for the example framework, we use this property for computing the grounded extension and we get $\mathcal{E}_{\mathcal{GR}}(F) = \{\{f\}\}$. Note that, although we know that the grounded semantics always gives exactly one extension, we prefer to enclose this extension in a set so that our notation is uniform across all semantics.

We can also determine the grounded extension using the characteristic function and this is in fact the most efficient way to do it if we do not need the other complete extensions as well. As shown in (Dung, 1995), the grounded extension is the stationary point of the sequence $\mathcal{F}_F(\varnothing)$, $\mathcal{F}_F(\mathcal{F}_F(\varnothing))$, ..., $\mathcal{F}_F^{(k)}(\varnothing)$. In our case, we have $\mathcal{F}_F(\varnothing) = \{f\}$, since $f$ is unattacked, then $\mathcal{F}_F(\{f\}) = \{f\}$, as no additional arguments are defended by $f$. As expected, we obtained the same grounded extension.

## 2.3 The Naive, Stage and Semi-stable Semantics

The ideas underlining the stage semantics were first proposed in (Verheij, 1996), but with a rather different terminology from what is currently used in the argumentation literature. Thus, we prefer the formalization from (Caminada, 2010) for the stage semantics. Similar ideas, but applied to admissible sets, are used for defining the semi-stable semantics in (Caminada, 2006b). Naive extensions – or maximal conflict-free sets – although not discussed in Dung's original work, have gained popularity with their use within the SCC-recursive schema (Baroni et al., 2005). We will use the argumentation framework in Figure 2.3 as an example for these three semantics.

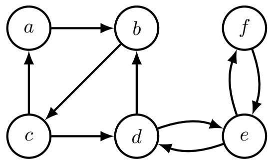



**Figure 2.3:** *Example argumentation framework for the naive, stage and semi-stable semantics.*

**Definition 10.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S \subseteq \mathcal{A}$ is a* ***naive extension*** *or, alternatively, a* ***maximal conflict-free set*** *($\mathcal{MCF}$) of $F$ iff $S$ is a maximal (with respect to set inclusion) conflict-free set.*

$$\mathcal{E}_{\mathcal{MCF}}(F) = \{S \in \mathcal{E}_{\mathcal{CF}}(F) \mid S \text{ is maximal w.r.t. } \subseteq\} \tag{2.9}$$

In some papers the naive extensions of an argumentation framework are denoted with $naive(F)$. We prefer the $\mathcal{E}_{\mathcal{MCF}}(F)$ notation for uniformity with the other semantics. For our example argumentation framework we have $\mathcal{E}_{\mathcal{MCF}}(F) = \{\{a,d,f\}, \{a,e\}, \{b,e\}, \{b,f\}, \{c,e\}, \{c,f\}\}$.

**Definition 11.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $S \subseteq \mathcal{A}$ be a set of arguments. We use $S^+$ to denote the set of arguments attacked by $S$. The* ***range*** *of $S$ is defined as $S \cup S^+$. The set $S$ is a* ***stage extension*** *($\mathcal{STA}$) of $F$ iff $S$ is conflict-free and has a maximal range (with respect to set inclusion) among all the conflict-free sets of $F$. The set $S$ is a* ***semi-stable extension*** *($\mathcal{SST}$) of $F$ iff $S$ is admissible and has a maximal range (with respect to set inclusion) among all the admissible sets of $F$.*

$$\begin{aligned}
&S^+ = \{a \in \mathcal{A} \mid S \to a\}\\
&range(S) = S \cup S^+\\
&\mathcal{E}_{\mathcal{STA}}(F) = \{S \in \mathcal{E}_{\mathcal{CF}}(F) \mid range(S) \text{ maximal w.r.t. } \subseteq\}\\
&\mathcal{E}_{\mathcal{SST}}(F) = \{S \in \mathcal{E}_{\mathcal{AS}}(F) \mid range(S) \text{ maximal w.r.t. } \subseteq\}
\end{aligned} \tag{2.10}$$

It is not difficult to see that stage extensions must also be naive. Indeed, given two conflict-free sets $S_1$ and $S_2$, we have that $S_1 \subsetneq S_2 \Rightarrow range(S_1) \subsetneq range(S_2)$. So we can determine the stage extensions by testing range maximality for naive extensions only. We have:

$$\begin{aligned}
&range(\{a,d,f\}) = \{a,b,d,e,f\}\\
&range(\{a,e\}) = \{a,b,d,e,f\}\\
&range(\{b,e\}) = \{b,c,d,e,f\}\\
&range(\{b,f\}) = \{b,c,e,f\} \subseteq range(\{b,e\})\\
&range(\{c,e\}) = \{a,c,d,e,f\}\\
&range(\{c,f\}) = \{a,c,d,e,f\}
\end{aligned} \tag{2.11}$$

Thus, we get $\mathcal{E}_{\mathcal{STA}}(F) = \{\{a,d,f\}, \{a,e\}, \{b,e\}, \{c,e\}, \{c,f\}\}$.

For computing the semi-stable extensions of the example framework, we will use the fact that every semi-stable extension is also a preferred extension (Caminada, 2006b). Thus, we first determine the admissible sets of $F$ as $\mathcal{E}_{\mathcal{AS}}(F) = \{\varnothing, \{e\}, \{f\}\}$, then the preferred extensions $\mathcal{E}_{\mathcal{PR}}(F) = \{\{e\}, \{f\}\}$. Now, we have $range(\{e\}) = \{d, e, f\}$ and $range(\{f\}) = \{e, f\} \subseteq range(\{e\})$. Thus, there is a single semi-stable extension for our example framework: $\mathcal{E}_{\mathcal{SST}}(F) = \{\{e\}\}$.

An important property of $\mathcal{SST}$ is the fact that for any argumentation framework $F$ that has at least one stable extension, the semi-stable and the stable semantics coincide (Caminada, 2006b). The same holds for the stage semantics, as shown in (Caminada, 2010).

$$\mathcal{E}_{\mathcal{ST}}(F) \neq \varnothing \Rightarrow \mathcal{E}_{\mathcal{SST}}(F) = \mathcal{E}_{\mathcal{STA}}(F) = \mathcal{E}_{\mathcal{ST}}(F) \tag{2.12}$$

Since every argumentation framework has at least one admissible set, it follows that it also has at least one semi-stable extension. Coupled with the above property, this means that the semi-stable semantics may be a suitable alternative to the traditional stable semantics.

## 2.4 Resolution-based Semantics

Resolution-based [1] semantics were introduced in (Baroni and Giacomin, 2008) using a principle-based approach: the authors were looking for semantics satisfying certain desirable properties. We only provide a minimalist definition here; for a more detailed account and the rationale behind these semantics, the reader may see (Baroni and Giacomin, 2008) or the more recent (Baroni et al., 2011b).

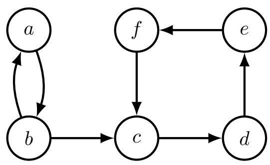



**Figure 2.4:** *Example argumentation framework for the resolution-based semantics.*

**Definition 12.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. Two arguments $a$ and $b$ are **conflicting** iff $a \to b$ or $b \to a$. For an argumentation framework $F$, the set of all conflicting pairs of arguments is denoted by $\mathcal{CONF}(F)$. The **full resolutions** of $F$, denoted by $\mathcal{RES}(F)$, are all argumentation frameworks that are based on the same set of arguments $\mathcal{A}$ and have an attack relation which preserves the conflicting pairs of $F$ and is minimal (with respect to set inclusion) among all such relations. The **resolution-based version** ($\mathcal{S}em^*$) of a given argumentation semantics $\mathcal{S}em$ takes all the extensions prescribed by $\mathcal{S}em$ for the frameworks in*

[1] Note that the use of the term "resolution" here is not related to the principle of resolution used in automated theorem proving.

*$\mathcal{RES}(F)$ and selects the minimal sets (with respect to set inclusion).*

$$\begin{aligned}
&\mathcal{CONF}(F) = \{(a,b) \in \mathcal{A} \times \mathcal{A} \mid a \neq b, (a,b) \in \mathcal{R} \text{ or } (b,a) \in \mathcal{R}\}\\
&\mathcal{RES}(F) = \{F' = (\mathcal{A}, \mathcal{R}') \mid \mathcal{CONF}(F') = \mathcal{CONF}(F), \mathcal{R}' \text{ is minimal w.r.t. } \subseteq\}\\
&\mathcal{E}_{\mathcal{S}em^*}(F) = \{E \in \bigcup_{F' \in \mathcal{RES}(F)} \mathcal{E}_{\mathcal{S}em}(F') \mid E \text{ is minimal w.r.t. } \subseteq\}
\end{aligned} \tag{2.13}$$

The main idea behind the approach is the fact that, given two arguments $a$ and $b$ that attack each other, one may choose to resolve the mutual conflict using preferences. For example, if $a$ is preferred to $b$, then the attack $b \to a$ can be ignored. Note that we still have $a \to b$, so the arguments are still in conflict. The full resolution of an argumentation framework actually consists in replacing all mutual attacks with single attacks. An important property, used as a lemma in (Baroni and Giacomin, 2008), is the fact that the complete extensions of a full resolution of $F$ are also complete extensions in $F$.

The instances discussed in (Baroni and Giacomin, 2008) are the resolution-based versions of the grounded ($\mathcal{GR}^*$) and preferred ($\mathcal{PR}^*$) semantics. The more recent paper (Baroni et al., 2011b) also talks about the resolution-based semi-stable ($\mathcal{SST}^*$) and ideal ($\mathcal{ID}^*$) semantics. The authors show that, of these four, only $\mathcal{GR}^*$ satisfies all the proposed properties.

Note that the only full resolution of a framework containing no mutual attacks will be the framework itself and the resolution-based semantics will coincide in this case with the corresponding base semantics. We will use the framework $F$ from Figure 2.4 as an example. The two full resolutions of $F$, $F_1$ and $F_2$, are depicted in Figure 2.5 (a) and (b), respectively.

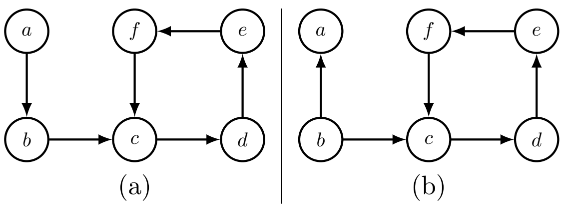



**Figure 2.5:** *The full resolutions of the framework $F$ from Figure 2.4.*

We start with the resolution-based grounded semantics, which is also the most important of them. We have $\mathcal{E}_{\mathcal{GR}}(F_1) = \{\{a\}\}$ and $\mathcal{E}_{\mathcal{GR}}(F_2) = \{\{b,d,f\}\}$, which leads to $\mathcal{E}_{\mathcal{GR}^*}(F) = \{\{a\}, \{b,d,f\}\}$.

For the resolution-based preferred semantics we have $\mathcal{E}_{\mathcal{PR}}(F_1) = \{\{a,c,e\}, \{a,d,f\}\}$ and $\mathcal{E}_{\mathcal{PR}}(F_2) = \{\{b,d,f\}\}$, so $\mathcal{E}_{\mathcal{PR}^*}(F) = \{\{a,c,e\}, \{a,d,f\}, \{b,d,f\}\}$.

Using the already determined preferred extensions we get the semi-stable extensions as $\mathcal{E}_{\mathcal{SST}}(F_1) = \{\{a,c,e\},\{b,d,f\}\}$ and $\mathcal{E}_{\mathcal{SST}}(F_2) = \{\{b,d,f\}\}$ and, thus, $\mathcal{E}_{\mathcal{SST}^*} = \{\{a,c,e\}, \{a,d,f\}, \{b,d,f\}\}$.

Note that we have not introduced the ideal semantics yet, but we need to use it. This is in fact due to a cyclic reference: the resolution-based approach is used on the ideal semantics, but the construction of the ideal semantics can be applied to the resolution-based grounded semantics to yield its ideal version. We have chosen to break the cycle like this because $\mathcal{ID}$ is easier to anticipate than $\mathcal{GR}^*$. The ideal extension of an argumentation framework, to be formally introduced in the following section, is the maximal (with respect to set inclusion) admissible set that is included in all preferred extensions of the framework.

For the first resolution, $F_1$, we have that the intersection of the preferred extensions is $\{a\}$, which is also admissible, so $\mathcal{E}_{\mathcal{ID}}(F_1) = \{\{a\}\}$. For the second resolution, $F_2$, we have a single preferred extension, so the ideal extension will coincide with it: $\mathcal{E}_{\mathcal{ID}}(F_2) = \{\{b, d, f\}\}$. Thus, for the resolution-based ideal semantics we get $\mathcal{E}_{\mathcal{ID}^*}(F) = \{\{a\}, \{b, d, f\}\}$.

We can see that for our example we have $\mathcal{E}_{\mathcal{PR}^*}(F) = \mathcal{E}_{\mathcal{SST}^*}(F)$ and also $\mathcal{E}_{\mathcal{ID}^*}(F) = \mathcal{E}_{\mathcal{GR}^*}(F)$. This is generally not the case. Indeed, we have seen already that $\mathcal{PR}$ is generally distinct from $\mathcal{SST}$ and we shall see that $\mathcal{ID}$ is distinct from $\mathcal{GR}$. Using the previous observation about frameworks containing no mutual attacks, we can easily find examples where the resolution-based semantics no longer coincide.

## 2.5 Parameterized Ideal Semantics

In this section we discuss the justification state of arguments with respect to semantics, then introduce the parameterized ideal semantics (Dvorak et al., 2011).

**Definition 13.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{S}em$ be an argumentation semantics. We say that an argument $a \in \mathcal{A}$ is **skeptically justified** (or **skeptically accepted**) with respect to $\mathcal{S}em$ if $a$ is a member of all the extensions prescribed by $\mathcal{S}em$. The argument $a$ is **credulously justified** (or **credulously accepted**) with respect to $\mathcal{S}em$ iff there is at least one extension that contains it. The argument is **defeated** if no extension of $F$ using $\mathcal{S}em$ contains it.*

A more detailed account of justification states is provided in (Baroni et al., 2004), where finer grained justification states are introduced. For the purpose of the current discussion, Definition 13 will suffice.

Given a unique status semantics (one that gives a single extension for any argumentation framework, such as the grounded semantics), the meanings of skeptical and credulous justification overlap. For multiple-status semantics, on the other hand, they are distinct.

Given some argumentation semantics $\mathcal{S}em$ (multiple status), one may attempt to use the set of all credulously (or skeptically) accepted arguments as an extension. However, the set of credulously accepted arguments is often not even conflict-free. The set of skeptically accepted arguments, on the other hand, is conflict-free, provided that the extensions of $\mathcal{S}em$ are conflict-free as well. We introduce the skeptical version of an argumentation semantics in 14, together with the notation that we will use.

**Definition 14.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{S}em$ be an argumentation semantics. The **skeptical version of** $\mathcal{S}em$ (notation $\mathcal{S}em^S$) returns a single extension consisting of the intersection of all the $\mathcal{S}em$ extensions* *of F.*

$$\mathcal{E}_{\mathcal{S}em^S}(F) = \{ \bigcap_{S \in \mathcal{E}_{\mathcal{S}em}(F)} S \} \tag{2.14}$$

We will only discuss the skeptical versions of the following semantics: complete, preferred, semi-stable, stage, naive and resolution-based grounded. While some of these skeptical versions are not particularly interesting in themselves, they will help complete the picture of relations between argumentation semantics, especially with respect to the ideal versions of the same semantics, presented in (Dvorak et al., 2011). We will use the framework in Figure 2.6 as an example.

For the complete semantics, note that the grounded extension is included in all complete extensions and is also a complete extension itself (Dung, 1995), so we have $\mathcal{CO}^S = \mathcal{GR}$. Thus, the skeptical complete extension of the example framework is given by $\mathcal{E}_{\mathcal{CO}^S}(F) = \mathcal{E}_{\mathcal{GR}}(F) = \{\{m, n\}\}$. Since it does not actually denote a new semantics, we will not use the $\mathcal{CO}^S$ notation further on.

Before moving on, a few words about computing the extensions for large frameworks. Considering all sets of arguments and testing their properties is generally not feasible, especially for manual computation. As an alternative one can compute the extensions for each connected component of the framework, then aggregate them to form extensions of the large framework. This relies on the additivity property, which states that the $\mathcal{S}em$ extensions of an argumentation framework $F$ can be computed as the union of $\mathcal{S}em$ extensions computed for the connected components of $F$. We will formally introduce additivity in Chapter 3, where we will also show that most argumentation semantics satisfy this property. For all the examples provided in this thesis, the extensions were computed (or double checked) using a Haskell implementation of the actual definitions (no optimization).

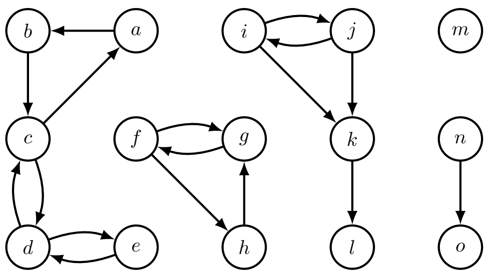



**Figure 2.6:** *Example argumentation framework for the skeptical and ideal versions of argumentation semantics*

The example framework has four preferred extensions $\mathcal{E}_{\mathcal{PR}}(F) = \{\{a, d, f, i, l, m, n\}, \{a, d, f, j, l, m, n\}, \{e, f, i, l, m, n\}, \{e, f, j, l, m, n\}\}$, so the skeptical preferred extension is $\mathcal{E}_{\mathcal{PR}^S}(F) = \{\{f, l, m, n\}\}$. Only two of the preferred extensions are also semi-stable: $\mathcal{E}_{\mathcal{SST}}(F) = \{\{a, d, f, i, l, m, n\}, \{a, d, f, j, l, m, n\}\}$, which leads to $\mathcal{E}_{\mathcal{SST}^S}(F) = \{\{a, d, f, l, m, n\}\}$. Note that, for this particular example, the skeptical versions of both semantics give extensions that are not admissible. Indeed, argument $l$, present in both, is not defended against $k$.

For the example framework we have that the stage and semi-stable semantics coincide. We have already seen that they are generally distinct, but here we preferred not to extend the size of the framework even more in order to exhibit this. So we have $\mathcal{E}_{\mathcal{STA}^S}(F) = \{\{a, d, f, l, m, n\}\}$.

We move on to the naive semantics. We will not list all the naive extensions of the example framework due to their large number (90). Instead, let us first see that $\{a, d, f, i, l, m, n\}$ and $\{c, e, g, k, m, o\}$ are naive extensions, as they are conflict-free sets and are also in conflict with all the arguments they do not contain. Thus, the only argument that can be in the skeptical version of $\mathcal{MCF}$ is $m$. Since $m$ is not in conflict with any other argument, it can be added to conflict-free sets that do not contain it already. Thus, $m$ must be in all naive extensions of $F$ and we get $\mathcal{E}_{\mathcal{MCF}^S}(F) = \{\{m\}\}$.

A general characterization of $\mathcal{MCF}^S$ can be deduced from (Dvorak et al., 2011): the elements of the skeptical naive extension are the arguments that are not self-attacking and are only in conflict with self-attacking arguments. Indeed, all arguments $b$ that are in conflict with an argument $a \in \mathcal{E}_{\mathcal{MCF}^S}(F)$ should not be in any naive extension of $F$ and, thus, they should not be part of any conflict-free set. But, for the singleton set $\{b\}$ not to be conflict-free, the argument $b$ must be self-attacking.

$$\mathcal{E}_{\mathcal{MCF}^S}(F) = \{a \in \mathcal{A} \mid (a, a) \notin \mathcal{R} \wedge \forall b (b \to a \texttt{ or } a \to b \Rightarrow (b, b) \in \mathcal{R})\} \qquad (2.15)$$

For the resolution-based grounded semantics, note that our example framework has 4 mutual attacks, so $F$ has 16 full resolutions. The corresponding grounded extensions are: $\{a, d, f, i, l, m, n\}$, $\{a, d, f, j, l, m, n\}$, $\{a, d, i, l, m, n\}$, $\{a, d, j, l, m, n\}$, $\{e, f, i, l, m, n\}$, $\{e, f, j, l, m, n\}$, $\{e, i, l, m, n\}$, $\{e, j, l, m, n\}$, $\{f, i, l, m, n\}$, $\{f, j, l, m, n\}$, $\{i, l, m, n\}$ and $\{j, l, m, n\}$. Note that we got only 12 distinct sets. Of these, we pick the minimal with respect to set inclusion and we get $\mathcal{E}_{\mathcal{GR}^*}(F) = \{\{i, l, m, n\}, \{j, l, m, n\}\}$ and, thus, $\mathcal{E}_{\mathcal{GR}^{*S}}(F) = \{\{l, m, n\}\}$.

As we have seen already, the skeptical version of an argumentation semantics $\mathcal{S}em$ does not necessarily provide an admissible extensions, even when $\mathcal{S}em$ gives complete extensions. The ideal version of $\mathcal{S}em$ deals with this in a straightforward manner: simply look for maximal admissible sets that are included in all the extensions yielded by $\mathcal{S}em$. Definition 15 follows the parametric approach from (Dvorak et al., 2011).

**Definition 15.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{S}em$ be an argumentation semantics that promises at least one extension. A set of arguments $S \subseteq \mathcal{A}$ is an **ideal set** with respect to $\mathcal{S}em$ ($\mathcal{S}em^{ids}$) iff it is admissible and is included in all $\mathcal{S}em$ extensions of $F$. The set $S$ is the **ideal extension** with respect to $\mathcal{S}em$ ($\mathcal{S}em^{id}$) of $F$ iff it is the maximal (with respect to set inclusion) ideal set (with respect to the same base semantics $\mathcal{S}em$).*

$$\begin{aligned} &\mathcal{E}_{\mathcal{S}em^{ids}}(F) = \{S \in \mathcal{E}_{\mathcal{AS}}(F) \mid \forall S' (S' \in \mathcal{E}_{\mathcal{S}em}(F) \Rightarrow S \subseteq S')\} \\ &\mathcal{E}_{\mathcal{S}em^{id}}(F) = \{S \in \mathcal{E}_{\mathcal{S}em^{ids}}(F) \mid S \textit{ is maximal w.r.t. } \subseteq\} \end{aligned} \qquad (2.16)$$

Note that for some argumentation semantics, the corresponding ideal version had already been defined prior to the parameterized approach from (Dvorak et al., 2011). The ideal version of the preferred semantics was introduced in (Dung et al., 2007) as "the" ideal semantics. We shall use $\mathcal{ID}$ instead of $\mathcal{PR}^{id}$ to refer to it and also $\mathcal{IDS}$ instead of $\mathcal{PR}^{ids}$ for the corresponding ideal sets. Also, the ideal version of the semi-stable semantics was introduced in (Caminada, 2007) as the eager semantics. We will use $\mathcal{EAG}$ instead of $\mathcal{SST}^{id}$ to refer to it and $\mathcal{EAGS}$ (eager sets) instead of $\mathcal{SST}^{ids}$ to refer to the corresponding ideal sets. Furthermore, the approach for defining the ideal versions is also discussed in (Caminada and Pigozzi, 2011) in the context of argument labelings, where it is regarded as a form of judgement aggregation.

Whenever the base semantics $\mathcal{S}em$ gives conflict-free extensions, the ideal version $\mathcal{S}em^{id}$ gives a unique extension (Dvorak et al., 2011). This justifies the formulation we have used in Definition 15.

Since the skeptical complete extension, which is the same as the grounded extension, is already admissible, the ideal version of the complete semantics is also the same as the grounded semantics $\mathcal{CO}^{id} = \mathcal{GR}$, so there is no need for using $\mathcal{CO}^{id}$ further on. The corresponding ideal sets are simply admissible sets included in the grounded extension. Since the grounded semantics is better described either as the least fixed point of the characteristic function or as the minimal complete extension, these sets are of little significance so we will not use $\mathcal{CO}^{ids}$ from now on either. On the other hand we will keep $\mathcal{IDS}$, $\mathcal{EAGS}$, $\mathcal{MCF}^{ids}$ and $\mathcal{GR}^{*ids}$, as they play an important role in defining the corresponding ideal semantics.

For the argumentation semantics that give complete extensions, the corresponding ideal extension is also complete (Dvorak et al., 2011). This holds for $\mathcal{ID}$, $\mathcal{EAG}$ and $\mathcal{GR}^{*id}$, but not for $\mathcal{STA}^{id}$ and $\mathcal{MCF}^{id}$. For the first three semantics, this also means that we can choose the ideal extension as the maximal complete extension included in all extensions of the corresponding base semantics. We introduce complete ideal sets in Definition 16.

**Definition 16.** *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{S}em \in \{\mathcal{PR}, \mathcal{SST}, \mathcal{GR}^*\}$. A set of arguments $S$ is a **complete ideal set** ($\mathcal{S}em^{cids}$) with respect to $\mathcal{S}em$ iff $S$ is a complete extension of $F$ and $S$ is included in all the extensions of $F$ using $\mathcal{S}em$.*

$$\mathcal{E}_{\mathcal{S}em^{cids}}(F) = \{S \in \mathcal{E}_{\mathcal{CO}}(F) \mid \forall S'(S' \in \mathcal{E}_{\mathcal{S}em}(F) \Rightarrow S \subseteq S')\} \tag{2.17}$$

We have already discussed the complete semantics. We now take the remaining base semantics one by one and determine the ideal sets, complete ideal sets (where appropriate) and ideal extension, for the example framework from Figure 2.6.

For the preferred, semantics, the ideal sets should be included in the skeptical preferred extension $\{f, l, m, n\}$. We have $\mathcal{E}_{\mathcal{IDS}}(F) = \{\varnothing, \{f\}, \{f, m\}, \{f, m, n\}, \{f, n\}, \{m\}, \{m, n\}, \{n\}\}$. Of these, only the sets containing both $m$ and $n$ are complete: $\mathcal{E}_{\mathcal{PR}^{cids}}(F) = \{\{f, m, n\}, \{m, n\}\}$. Last, but not least, we have $\mathcal{E}_{\mathcal{ID}}(F) = \{\{f, m, n\}\}$. This example also shows that $\mathcal{ID} \neq \mathcal{PR}^S$.

For the semi-stable semantics, we know that eager sets should be included in the skeptical semi-stable extension $\{a, d, f, l, m, n\}$. We get $\mathcal{E}_{\mathcal{EAGS}}(F) = \{\varnothing, \{a, d\}, \{a, d, f\}, \{a, d, f, m\}, \{a, d, f, m, n\}, \{a, d, f, n\}, \{a, d, m\}, \{a, d, m, n\}, \{a, d, n\}, \{d\}, \{d, f\}, \{d, f, m\}, \{d, f, m, n\}, \{d, f, n\}, \{d, m\}, \{d, m, n\}, \{d, n\}, \{f\}, \{f, m\}, \{f, m, n\}, \{f, n\}, \{m\}, \{m, n\}, \{n\}\}$.

The corresponding complete ideal sets must contain both $m$ and $n$, as they are unattacked arguments. Additionally, sets that contain $d$ must also contain $a$, since $a$ is defended by $d$. We get $\mathcal{E}_{\mathcal{SST}^{cids}}(F) = \{\{a, d, f, m, n\}, \{a, d, m, n\}, \{f, m, n\},$

$\{m, n\}\}$. Thus, the eager extension is $\mathcal{E}_{\mathcal{EAG}}(F) = \{\{a, d, f, m, n\}\}$ and we also see that $\mathcal{EAG} \neq \mathcal{SST}^S$.

Since the stage and semi-stable semantics coincide for the example framework, we can directly write $\mathcal{E}_{\mathcal{STA}^{ids}}(F) = \mathcal{E}_{\mathcal{EAGS}}(F)$ and $\mathcal{E}_{\mathcal{STA}^{id}}(F) = \mathcal{E}_{\mathcal{EAG}}(F) = \{\{a, d, f, m, n\}\}$.

For the naive semantics we have a rather easy job, since the skeptical naive extension is just $\{m\}$. We have $\mathcal{E}_{\mathcal{MCF}^{ids}}(F) = \{\emptyset, \{m\}\}$ and $\mathcal{E}_{\mathcal{MCF}^{id}} = \{\{m\}\}$. Note that, in general, $\mathcal{MCF}^{id} \neq \mathcal{MCF}^S$, although for this example they provide the same extension.

The last base semantics we need to cover is the resolution-based grounded semantics. The ideal sets must be included in the corresponding skeptical extension $\{l, m, n\}$ so we have $\mathcal{E}_{\mathcal{GR}^{*ids}}(F) = \{\emptyset, \{m\}, \{m, n\}, \{n\}\}$. In this case there is only one complete ideal set $\mathcal{E}_{\mathcal{GR}^{*cids}}(F) = \{\{m, n\}\}$, which is also the ideal extension $\mathcal{E}_{\mathcal{GR}^{*id}}(F) = \{\{m, n\}\}$. We have $\mathcal{GR}^{*id} \neq \mathcal{GR}^{*S}$. Also note that $\mathcal{GR}^{*id}$ is in general distinct from $\mathcal{GR}$.

## 2.6 SCC-recursiveness

In this section we introduce SCC-recursiveness (Baroni et al., 2005) and the SCC-recursive semantics defined in the same paper. We also discuss the newer *stage*2 semantics (Dvorak and Gaggl, 2012).

The general idea behind SCC-recursiveness is to compute semantics taking advantage of the decomposition of the argumentation framework along its strongly connected components (SCC's). We will only provide here the minimal definitions required for introducing the concept. For more details and the rationale behind the idea, the reader may consult (Baroni et al., 2005).

We start with a formal definition for the strongly connected components.

**Definition 17.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. We define the **path equivalence relation** $PE_F \subseteq \mathcal{A} \times \mathcal{A}$ as follows:*

*(i) $(a, a) \in PE$ for all arguments $a \in \mathcal{A}$*

*(ii) for any two arguments $a$ and $b$, $(a, b) \in PE$ iff there is a path in $\mathcal{R}$ from $a$ to $b$ and a path from $b$ to $a$*

*$PE$ so defined is an equivalence relation and its equivalence classes are called **strongly connected components** (SCC's). We will use $SCCS_F$ to refer to the set of strongly connected components of $F$.*

We have talked about justification states in the previous section and have seen how to distinguish types of arguments with respect to an argumentation semantics. Similarly, given an argumentation framework and an extension, we can distinguish three types of arguments with respect to how the extension interacts with them considering also the SCC decomposition of the framework. These classes are introduced in Definition 18.

**Definition 18.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework, $E \subseteq \mathcal{A}$ a set of arguments and $S \in SCCS_F$ a strongly connected component of $F$. The elements of $S$ can be partitioned into three sets with respect to $E$:*

*(i) $D_F(S, E) = \{a \in S \mid (E \setminus S) \to a\}$ – the set of arguments attacked by $E$ from outside $S$*

*(ii)* $U_F(S,E) = \{a \in S \mid (E \setminus S) \not\to a \wedge \forall b(b \in \mathcal{A} \setminus S \wedge b \to a \Rightarrow (E \setminus S) \to b)\}$ – *the set of arguments not attacked by E from outside S and defended by E against all attackers that are not in S.*

*(iii)* $P_F(S,E) = S \setminus (D_F(S,E) \cup U_F(S,E))$ – *arguments that are neither attacked by E from outside of S, nor defended by E against attacks coming from outside S.*

We will use $UP_F(S,E)$ as a notation for $U_F(S,E) \cup P_F(S,E)$, which is the same as $S \setminus D_F(S,E)$. We are now ready to define SCC-recursive semantics.

**Definition 19.** *An argumentation semantics $\mathcal{S}em$ is said to be* ***SCC-recursive*** *iff, for any argumentation framework $F = (\mathcal{A}, \mathcal{R})$ we have $\mathcal{E}_{\mathcal{S}em}(F) = \mathcal{GF}(F, \mathcal{A})$, where the generic recursive function $\mathcal{GF}$ is defined as follows: for any argumentation framework $F = (\mathcal{A}, \mathcal{R})$ and any two sets of arguments $E, C \subseteq \mathcal{A}$ it holds that $E \in \mathcal{GF}(F, C)$ iff*

- *if $|SCCS_F| = 1$ then*

$$E \in \mathcal{BF}_{\mathcal{S}em}(F, C) \tag{2.18}$$

- *otherwise, for all strongly connected components S, we have*

$$(E \cap S) \in \mathcal{GF}(F{\downarrow}_{UP_F(S,E)}, U_F(S,E) \cap C) \tag{2.19}$$

*where $\mathcal{BF}_{\mathcal{S}em}$ stands for a* ***base function*** *that, given an argumentation framework $F = (\mathcal{A}, \mathcal{R})$ such that $|SCCS_F| = 1$ and a set of arguments $C \subseteq \mathcal{A}$, returns a set of arguments. The base function depends on the argumentation semantics $\mathcal{S}em$.*

The intuition behind Definition 19 is that the extensions of SCC-recursive semantics can be computed based on the SCC-decomposition of the argumentation framework by choosing the elements of the extension from each strongly connected component $S$ as follows:

- if the $S$ is unattacked, the base function can be used for selecting arguments from $S$ for the extension
- if $S$ is attacked by other SCC's, then we first choose the extension's arguments from the parent SCC's, then we can select the elements from $S$ accounting for the following:
  - the arguments that are attacked by arguments that we have selected for the extension from the parent SCC's of $S$ – these arguments will be ignored
  - the arguments that are defended by elements of the extension chosen from parent SCC's against attacks that come from outside of $S$ – this set is given as input to the generic function

This approach progressively computes the extensions of the SCC-recursive semantics. Alternatively, one can consider all possible sets of arguments $E \subseteq \mathcal{A}$ and test the property specified in Definition 19.

It is shown in (Baroni et al., 2005) that all the classical semantics are SCC-recursive so this approach proves to be an alternative to the traditional definitions of these semantics. On the other hand, we have showed in (Gratie and Florea, 2012c) that $\mathcal{ID}$, $\mathcal{IDS}$, $\mathcal{EAG}$ and $\mathcal{EAGS}$ are not SCC-recursive. We also provided a generalization of SCC-recursiveness that is able to capture aspects of the local behavior of argumentation semantics (Gratie et al., 2012b).

We will use the argumentation framework in Figure 2.7, which has 4 strongly connected components: $S_1 = \{a,b,c,d\}$, $S_2 = \{e\}$, $S_3 = \{f,g\}$ and $S_4 = \{h\}$. We have the following attacks between these components: $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$.

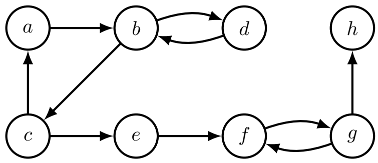



**Figure 2.7:** *Example argumentation framework for the SCC-recursive semantics.*

In order to define new SCC-recursive semantics $\mathcal{S}em$, one takes a base function $\mathcal{BF}_{\mathcal{S}em}(F,C)$ and uses it to compute the generic function $\mathcal{GF}(F,C)$ according to the rules specified in Definition 19. Then the extensions of $\mathcal{S}em$ are given by $\mathcal{E}_{\mathcal{S}em}(F) = \mathcal{GF}(F,\mathcal{A})$. Four new semantics using this approach were proposed in (Baroni et al., 2005), in the search for a uniform treatment of odd and even length cycles.

The $\mathcal{AD}1$ and $\mathcal{AD}2$ semantics use base functions that give admissible sets. The extensions of these semantics are complete (Baroni et al., 2005).

**Definition 20.** *The $\mathcal{AD}1$ semantics is defined as the SCC-recursive semantics given by the following base function:*

$$\mathcal{BF}_{AD1}(F,C) = \begin{cases} \mathcal{E}_{\mathcal{PR}}(F), & \text{if } C = \mathcal{A} \\ \{\varnothing\}, & \text{otherwise} \end{cases} \tag{2.20}$$

**Definition 21.** *The $\mathcal{AD}2$ semantics is defined as the SCC-recursive semantics given by the following base function:*

$$\mathcal{BF}_{AD2}(F,C) = \begin{cases} \{E \in \mathcal{AS}^*(F) \mid E \text{ is maximal w.r.t. } \subseteq\}, & \text{if } C = \mathcal{A} \\ \{\varnothing\}, & \text{otherwise} \end{cases} \tag{2.21}$$

*where $\mathcal{AS}^*(F) = \{E \in \mathcal{E}_{\mathcal{AS}}(F) \mid \forall a(a \rightarrow E \Rightarrow \forall b(b \rightarrow a \Rightarrow b \in E))\}$ (admissible sets that contain all the attackers of arguments that attack them).*

For the example framework we have $\mathcal{E}_{\mathcal{AD}1}(F) = \{\{c,d,f,h\}, \{c,d,g\}\}$. Let us see that $E = \{c,d,f,h\}$ satisfies the required property. We have $UP_F(S_1,E) = S_1$ and $U_F(S_1,E) = S_1$, as $S_1$ is unattacked. Thus, $\mathcal{GF}(F{\downarrow}_{UP_F(S_1,E)}, \mathcal{A} \cap U_F(S_1,E)) = \mathcal{GF}(F{\downarrow}_{S_1}, S_1) = \mathcal{BF}_{AD1}(F{\downarrow}_{S_1}, S_1) = \mathcal{E}_{\mathcal{PR}}(F{\downarrow}_{S_1}) = \{\{c,d\}\}$. Since $E \cap S_1 = \{c,d\}$, the condition is satisfied for $S_1$.

We move on to $S_2$. We have $UP_F(S_2,E) = \varnothing$ and $U_F(S,E) = \varnothing$, because $e$ is attacked by $c$. Thus, $\mathcal{GF}(F{\downarrow}_{UP_F(S_2,E)}, \mathcal{A} \cap U_F(S_2,E)) = \mathcal{GF}(F{\downarrow}_{\varnothing}, \varnothing) = \mathcal{E}_{\mathcal{PR}}(F{\downarrow}_{\varnothing}) = \{\varnothing\}$. Since $E \cap S_2 = \varnothing$ the condition is again satisfied.

For $S_3$ we have $UP_F(S_3,E) = S_3$ and $U_F(S_3,E) = S_3$, because the extension $E$ defends $S_3$ against the outer attacker $e$. Furthermore, $\mathcal{GF}(F{\downarrow}_{UP_F(S_3,E)}, \mathcal{A} \cap U_F(S_3,E)) = \mathcal{GF}(F{\downarrow}_{S_3}, S_3) = \mathcal{E}_{\mathcal{PR}}(F{\downarrow}_{S_3}) = \{\{f\},\{g\}\}$ and the condition is again satisfied, as $E \cap S_3 = \{f\}$.

For the last SCC, we have $UP_F(S_4, E) = U_F(S_4, E) = S_4$, because the only outer attacker of $S_4$, argument $g$, is attacked by $E$. Thus, $\mathcal{GF}(F{\downarrow}_{UP_F(S_4,E)}, \mathcal{A} \cap U_F(S_4, E)) = \mathcal{GF}(F{\downarrow}_{S_4}, S_4) = \mathcal{E}_{\mathcal{PR}}(F{\downarrow}_{S_4}) = \{\{h\}\}$ and we also have $E \cap S_4 = \{h\}$, which concludes our proof.

For the $\mathcal{AD}2$ semantics we have $\mathcal{E}_{\mathcal{AD}2}(F) = \{\varnothing\}$. Note that the main reason for getting only the empty set as an extension is the fact that the non-empty admissible sets of $S_1$, namely $\{d\}$ and $\{c, d\}$ don't satisfy the property required by the base function since they don't contain $a$, an attacker of $b$, which in turn attacks $d$.

In contrast with the $\mathcal{AD}1$ and $\mathcal{AD}2$ semantics, which were based on admissible base functions, the $\mathcal{CF}1$ and $\mathcal{CF}2$ semantics are both based on naive extensions.

**Definition 22.** *The $\mathcal{CF}1$ semantics is defined as the SCC-recursive semantics given by the following base function:*

$$\mathcal{BF}_{CF1}(F, C) = \mathcal{E}_{\mathcal{MCF}}(F{\downarrow}_C) \tag{2.22}$$

**Definition 23.** *The $\mathcal{CF}2$ semantics is defined as the SCC-recursive semantics given by the following base function:*

$$\mathcal{BF}_{CF2}(F, C) = \mathcal{E}_{\mathcal{MCF}}(F) \tag{2.23}$$

Although the $\mathcal{CF}1$ and $\mathcal{CF}2$ semantics do not give admissible extensions, they do satisfy several other desirable properties (Baroni et al., 2005). We shall discuss some of them in Chapter 3.

For the example framework we have $\mathcal{E}_{\mathcal{CF}1}(F) = \{\{a, d, g\}, \{b, e, g\}, \{c, d, f, h\}, \{c, d, g\}\}$ and $\mathcal{E}_{\mathcal{CF}2}(F) = \{\{a, d, e, g\}, \{b, e, g\}, \{c, d, f, h\}, \{c, d, g\}\}$. Note that $\mathcal{CF}2$ does not use the second argument of the generic function (the set of defended arguments, $C$).

**Definition 24.** *The stage2 ($\mathcal{STA}2$) semantics is given by the following base function:*

$$\mathcal{BF}_{STA2}(F, C) = \mathcal{E}_{\mathcal{STA}}(F) \tag{2.24}$$

The *stage*2 semantics, introduced in (Dvorak and Gaggl, 2012), aims to deal with several issues identified in the behavior of $\mathcal{CF}1$ and $\mathcal{STA}$, while retaining the desirable properties of both semantics. The authors manage to do that by replacing $\mathcal{MCF}$ with $\mathcal{STA}$ in the definition of $\mathcal{CF}2$. Thus, the *stage*2 semantics can be seen as a refinement of $\mathcal{CF}1$. In addition, for any argumentation framework $F$ where the preferred and stable semantics coincide it holds that $\mathcal{E}_{\mathcal{ST}}(F) = \mathcal{E}_{\mathcal{STA}}(F) = \mathcal{E}_{\mathcal{STA}}(F)$.

The *stage*2 extensions for the example framework are $\mathcal{E}_{\mathcal{STA}2} = \{\{c, d, f, h\}, \{c, d, g\}\}$. It is shown in (Dvorak and Gaggl, 2012) that all *stage*2 extensions are also $\mathcal{CF}2$ extensions, but from this example we can see that the two semantics are distinct. A more extensive comparison with the other semantics will be provided in Chapter 3.

## 2.7 Prudent Semantics

This section is dedicated to prudent semantics (Coste-Marquis et al., 2005). The approach relies on the intuition that an odd-length path of attacks can be seen as an indirect attack, while an even-length path of attacks can be interpreted as an indirect defense. Two arguments are said to be controversial if one of them both indirectly attacks and indirectly defends the other. The authors argue that allowing such pairs of arguments in the extensions is not prudent. They follow an approach similar to that of (Dung, 1995) for defining prudent versions of the classical semantics As we shall see, however, some of the prudent semantics are rather distinct from their classical counterparts. We will use the framework from Figure 2.8 as an example.

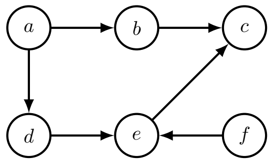



**Figure 2.8:** *Argumentation framework for the prudent semantics.*

We will change the definitions of prudent semantics by introducing a definition for $p$-conflict-free sets. This is only implicitly present in (Coste-Marquis et al., 2005), in the definition of $p$-admissible sets.

**Definition 25.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. We say that an argument $a$ **indirectly attacks** (notation $a \rightarrow^i b$) an argument $b$ iff there is an odd-length path from $a$ to $b$ in $F$. A set of arguments $S$ is $p$-**conflict-free** ($\mathcal{CF}^P$) iff there is no indirect conflict between its elements.*

$$\begin{aligned} &a \rightarrow^i b \Leftrightarrow \exists (a_0 = a) \rightarrow a_1 \rightarrow \ldots \rightarrow (a_{2n+1} = b) \\ &\mathcal{E}_{\mathcal{CF}^P}(F) = \{S \subseteq \mathcal{A} \mid \forall ab(a \in S \textbf{ and } b \in S \Rightarrow a \not\rightarrow^i b)\} \end{aligned} \tag{2.25}$$

It is easy to notice that every $p$-conflict-free set must also be conflict-free, but not vice versa. For the example framework we have $\mathcal{E}_{\mathcal{CF}}(F) = \{\varnothing, \{a\}, \{a,c\}, \{a,c,f\}, \{a,e\}, \{a,f\}, \{b\}, \{b,d\}, \{b,d,f\}, \{b,e\}, \{b,f\}, \{c\}, \{c,d\}, \{c,d,f\}, \{c,f\}, \{d\}, \{d,f\}, \{e\}, \{f\}\}$ and the only indirect attack (that is not also a direct attack) is between $a$ and $c$. Thus we should avoid all conflict-free sets that contain both $a$ and $c$. We get $\mathcal{E}_{\mathcal{CF}^P}(F) = \{\varnothing, \{a\}, \{a,e\}, \{a,f\}, \{b\}, \{b,d\}, \{b,d,f\}, \{b,e\}, \{b,f\}, \{c\}, \{c,d\}, \{c,d,f\}, \{c,f\}, \{d\}, \{d,f\}, \{e\}, \{f\}\}$.

**Definition 26.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S$ is $p$-**admissible** ($\mathcal{AS}^P$) iff it is $p$-conflict-free and defends all its elements.*

$$\mathcal{E}_{\mathcal{AS}^P}(F) = \{S \in \mathcal{E}_{\mathcal{CF}^P}(F) \mid \forall a(a \in S \Rightarrow \forall b(b \rightarrow a \Rightarrow S \rightarrow b))\} \tag{2.26}$$

Note that, in our example, arguments $b$, $d$ and $e$ cannot be defended (as they are attacked by unattacked arguments). Furthermore, the only defender of $c$ is $a$, but $a$ indirectly attacks $c$. With these observations, we get $\mathcal{E}_{\mathcal{AS}^P}(F) = \{\varnothing, \{a\},$

$\{a,f\}$, $\{f\}\}$. It's easy to see that all $p$-admissible sets are also admissible. The converse generally does not hold. Indeed, we have $\mathcal{E}_{\mathcal{AS}}(F) = \{\varnothing,\ \{a\},\ \{a,c,f\},\ \{a,f\},\ \{f\}\}$.

**Definition 27.** *Let $F = (\mathcal{A},\mathcal{R})$ be an argumentation framework. A set of arguments $S$ is a **preferred $p$-extension** ($\mathcal{PR}^P$) of $F$ iff it is a maximal (with respect to set inclusion) $p$-admissible set.*

$$\mathcal{E}_{\mathcal{PR}^P}(F) = \{S \in \mathcal{E}_{\mathcal{AS}^P}(F) \mid S \text{ is maximal w.r.t. } \subseteq\} \tag{2.27}$$

For our framework we have $\mathcal{E}_{\mathcal{PR}^P}(F) = \{\{a,f\}\}$. Note that, since preferred extensions may contain arguments $a$ and $b$ such that $a$ indirectly attacks $b$, the prudent preferred extensions are generally distinct from the preferred ones. Indeed, for the example framework we have that $\mathcal{E}_{\mathcal{PR}}(F) = \{\{a,c,f\}\}$ and $a$ indirectly attacks $c$.

**Definition 28.** *Let $F = (\mathcal{A},\mathcal{R})$ be an argumentation framework. A set of arguments $S$ is a **stable $p$-extension** iff it is $p$-conflict-free and attacks all arguments it does not contain.*

$$\mathcal{E}_{\mathcal{ST}^P}(F) = \{S \in \mathcal{E}_{\mathcal{CF}^P}(F) \mid \forall a(a \notin S \Rightarrow S \to a)\} \tag{2.28}$$

Note that all stable $p$-extensions are also stable, but not vice versa. In fact, the prudent stable semantics can be defined as

$$\mathcal{E}_{\mathcal{ST}^P}(F) = \mathcal{E}_{\mathcal{CF}^P}(F) \cap \mathcal{E}_{\mathcal{ST}}(F) \tag{2.29}$$

In our case, the only stable extension of $F$, namely $\{a,c,f\}$, is not $p$-conflict-free, which means that $\mathcal{E}_{\mathcal{ST}^P}(F) = \varnothing$. Alternatively, we can rely on the fact that every stable $p$-extension is also a preferred $p$-extensions (Coste-Marquis et al., 2005). This means that we need to examine the set $\{a,f\}$, which turns out not to be a stable $p$-extension because it does not attack $c$. Thus, we have reached the same result, that there is no stable $p$-extension for the example framework.

We will next discuss the prudent version of the grounded semantics. Note that the repeated application of the characteristic function used in Definition 9 ensures that only admissible sets are generated, but cannot enforce $p$-admissibility as well. This observation leads to the use of a different function for defining the grounded $p$-extension.

**Definition 29.** *Let $F = (\mathcal{A},\mathcal{R})$ be an argumentation framework. For a set of arguments $S$, the **prudent characteristic function** $\mathcal{F}_F^P(S)$ gives the set of all arguments $a$ such that $S$ defends $a$ and $S \cup \{a\}$ is without indirect conflicts. The **grounded $p$-extension** of $F$ ($\mathcal{GR}^P$) is defined as the stationary value of the sequence $(\mathcal{F}_F^{P,i}(\varnothing))_{i\in\mathbb{N}}$.*

$$\begin{aligned}
&\mathcal{F}_F^P(S) = \{a \in \mathcal{A} \mid \forall b(b \to a \Rightarrow S \to b) \textbf{ and } S \cup \{a\} \in \mathcal{E}_{\mathcal{CF}^P}(F)\}\\
&\mathcal{F}_F^{P,0}(\varnothing) = \varnothing\\
&\mathcal{F}_F^{P,i+1}(\varnothing) = \mathcal{F}_F^P(\mathcal{F}_F^{P,i}(\varnothing)) \text{ for all } i \in \mathbb{N}\\
&\mathcal{E}_{\mathcal{GR}^P}(F) = \{\mathcal{F}_F^{P,j}(\varnothing)\} \text{ with } j \text{ such that } \mathcal{F}_F^{P,i+1}(\varnothing) = \mathcal{F}_F^{P,i}(\varnothing) \text{ for all } i \geq j
\end{aligned} \tag{2.30}$$

Note that the prudent-characteristic function is not monotonic (with respect to set inclusion), but its successive application for the empty set gives only $p$-admissible sets and eventually leads to a fixed point (Coste-Marquis et al., 2005). For our framework we have $\mathcal{F}_F^P(\varnothing) = \{a, f\}$ and $\mathcal{F}_F^P(\{a, f\}) = \{a, f\}$, so $\mathcal{E}_{\mathcal{GR}^P}(F) = \{\{a, f\}\}$. Note that the (non-prudent) grounded semantics gives a different extension $\mathcal{E}_{\mathcal{GR}}(F) = \{\{a, c, f\}\}$.

**Definition 30.** *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S$ is a **complete $p$-extension** of $F$ ($\mathcal{CO}^P$) iff $S$ is $p$-conflict-free and it is a fixed point of the prudent characteristic function.*

$$\mathcal{E}_{\mathcal{CO}^P}(F) = \{S \in \mathcal{E}_{\mathcal{CO}^P}(F) \mid \mathcal{F}_F^P(S) = S\} \tag{2.31}$$

Alternatively, we can define complete $p$-extensions as $p$-admissible sets $S$ that contain all arguments $a$ such that $S$ defends $a$ and $S \cup \{a\}$ is $p$-conflict-free. Thus, in order to determine the complete $p$-extensions for the example framework we only need to consider the $p$-admissible sets. But since we have seen already that the empty set defends both $a$ and $f$, our complete extensions should contain them, so we end up with $\mathcal{E}_{\mathcal{CO}^P}(F) = \{\{a, f\}\}$. In general it is not the case that the grounded, preferred and complete prudent semantics give the same result.

While the grounded extension of an argumentation framework $F$ is included in all complete extensions of $F$, it is in general not the case that the grounded $p$-extension is included in all preferred $p$-extensions (Coste-Marquis et al., 2005). In fact, as shown in the paper, whenever $F$ has at least one stable $p$-extension, the intersection of all preferred $p$-extensions of $F$ is included in the grounded $p$-extension of $F$.

As a final observation, note that computing the prudent semantics is not the same as adding the indirect attacks to the framework and computing the classical semantics in the new framework. The distinction is given by the fact that, for prudent semantics, indirect attacks do not count as defense.

## 2.8 Enhanced Preferred Semantics

We now turn to enhanced preferred extensions (Zhang and Lin, 2010). The authors aim to provide an argumentation semantics that can deal with frameworks that only have the empty set as an admissible extension. They use an approach similar to that from (Dung, 1995), but working with pairs of sets of arguments $(S, H)$, where $H$ contains arguments that are ignored. We start with the definition of defense.

**Definition 31.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A pair $(S, H) \subseteq \mathcal{A} \times \mathcal{A}$ **defends** an argument $a \in \mathcal{A}$ iff the following conditions hold:*

$$\begin{aligned} & a \notin H \\ & H \cap S = \varnothing \\ & \forall b (b \notin H \textbf{ and } b \to a \Rightarrow S \to b) \end{aligned} \tag{2.32}$$

In other words, the pair $(S, H)$ defends $a$ iff $S$ and $H$ are disjoint, $a$ is not ignored and $S$ defends $a$ against all attackers that are not in $H$. Based on this notion of defense, we can introduce admissible pairs.

**Definition 32.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A pair $(S, H) \subseteq \mathcal{A} \times \mathcal{A}$ is an **admissible pair** iff the following conditions hold:*

*(a) $S \neq \varnothing$ or $H = \mathcal{A}$*

*(b) $(S, H)$ defends all the arguments from $S$*

It is not difficult to see that, in fact, $(S, H)$ is an admissible pair iff $S$ is a non-empty admissible set of the restricted framework $F\downarrow_{\mathcal{A} \setminus H}$ or $S = \varnothing$ and $H = \mathcal{A}$.

**Definition 33.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A pair $(S, H)$ is a **minimal admissible pair** iff it is a minimal (with respect to the cardinality of $H$) admissible pair. $(S, H)$ is an **enhanced preferred extension** of $F$ iff it is a maximal pair (with respect to set inclusion applied to $S$) among the minimal admissible sets of $F$. Given an enhanced preferred extension $(S, H)$, the set $S$ is a **proper enhanced preferred extension** of $F$ ($\mathcal{EPS}$).*

To sum up, in order to determine the (proper) enhanced preferred extensions of an argumentation framework $F$, one needs to find the non-empty admissible sets of all subframeworks of $F$, then choose the maximal sets among those that were computed by ignoring a minimal number of arguments. Thus, if $F$ has at least one non-empty admissible set, all minimal admissible pairs $(S, H)$ will have $H = \varnothing$, which means that $S \in \mathcal{E}_{\mathcal{AS}}(F) \setminus \{\varnothing\}$. In this case we get that the enhanced preferred and the preferred semantics coincide: $\mathcal{E}_{\mathcal{EPS}}(F) = \mathcal{E}_{\mathcal{PR}}(F)$.

If $F$ has only $\varnothing$ as an admissible set, we check whether we can obtain non-empty admissible sets by ignoring one argument. If we still end up only with the empty set, we try ignoring two arguments, and so on. Once we have obtained non-empty admissible sets, we choose the maximal ones among them as the enhanced preferred extensions of $F$.

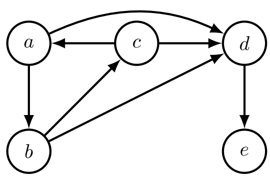



**Figure 2.9:** *Example argumentation framework for the enhanced preferred semantics.*

We will use the framework from Figure 2.9 as an example. Let us first see that the preferred semantics only gives us the empty set. This means that we should look for admissible pairs $(S, H)$ with $|H| = 1$. We get the following minimal admissible pairs: $(\{b\}, \{a\})$, $(\{b, e\}, \{a\})$, $(\{c\}, \{b\})$, $(\{c, e\}, \{b\})$, $(\{a\}, \{c\})$, $(\{a, e\}, \{c\})$, $(\{e\}, \{d\})$. From these, we take the ones having the first set maximal with respect to set inclusion and get $\mathcal{E}_{\mathcal{EPS}}(F) = \{\{a, e\}, \{b, e\}, \{c, e\}\}$. Note that these sets are not admissible in $F$, but only in the subframework defined by the corresponding ignored set (the second element of the admissible pair). They are, however, conflict-free in $F$. In Chapter 3 we shall discuss other properties of this semantics.

As a final remark, we feel that set inclusion rather than cardinality should be used for defining minimal admissible pairs. We provide a labeling-based approach using this idea in Chapter 6.

## 2.9 Round-up

We conclude this section with an overview of all the argumentation semantics that we have discussed. There are, in total, 43 distinct semantics, as follows:

- (6) the classical semantics: conflict-free – $\mathcal{CF}$, admissible – $\mathcal{AS}$, complete – $\mathcal{CO}$, stable – $\mathcal{ST}$, preferred – $\mathcal{PR}$ and grounded – $\mathcal{GR}$
- (3) the naive ($\mathcal{MCF}$), stage ($\mathcal{STA}$) and semi-stable ($\mathcal{SST}$) semantics
- (4) the resolution-based versions for the grounded ($\mathcal{GR}^*$), preferred ($\mathcal{PR}^*$), ideal ($\mathcal{ID}^*$) and semi-stable ($\mathcal{SST}^*$) semantics
- (5) skeptical versions of the preferred ($\mathcal{PR}^S$), semi-stable ($\mathcal{SST}^S$), stage ($\mathcal{STA}^S$), naive ($\mathcal{MCF}^S$) and resolution-based grounded ($\mathcal{GR}^{*S}$) semantics
- (5) ideal sets based on the preferred ($\mathcal{IDS}$ – "the" ideal sets), semi-stable ($\mathcal{EAGS}$ – eager sets), stage ($\mathcal{STA}^{ids}$), naive ($\mathcal{MCF}^{ids}$) and resolution-based grounded ($\mathcal{GR}^{*ids}$) semantics
- (3) complete ideal sets for the preferred ($\mathcal{PR}^{cids}$), semi-stable ($\mathcal{SST}^{cids}$) and resolution-based grounded ($\mathcal{GR}^{*cids}$) semantics
- (5) ideal semantics based on the preferred ($\mathcal{ID}$ – "the" ideal semantics), semi-stable ($\mathcal{EAG}$ – eager), stage ($\mathcal{STA}^{id}$), naive ($\mathcal{MCF}^{id}$) and resolution-based grounded ($\mathcal{GR}^{*id}$) semantics
- (5) SCC-recursive semantics: $\mathcal{CF}1$, $\mathcal{CF}2$, $\mathcal{AD}1$, $\mathcal{AD}2$ and the *stage*2 ($\mathcal{STA}2$) semantics
- (6) prudent versions for the conflict-free ($\mathcal{CF}^P$), admissible ($\mathcal{AS}^P$), complete ($\mathcal{CO}^P$), stable ($\mathcal{ST}^P$), preferred ($\mathcal{PR}^P$) and grounded ($\mathcal{GR}^P$) semantics
- (1) the enhanced preferred semantics – $\mathcal{EPS}$

To the best of our knowledge, this is the most comprehensive survey of argumentation semantics, at least with respect to the number of argumentation semantics that are considered. We will add five more semantics of our own to this list and provide an overview of the relations between them in Chapter 3.

*Chapter 3*

# Properties of Argumentation Semantics

In this chapter we focus on the properties of argumentation semantics. We discuss several principles proposed for evaluating semantics in Section 3.1. Most principles come from (Baroni and Giacomin, 2007) and (Baroni et al., 2011a), but we also introduce a new property that will be very useful for the results presented in Chapter 5. In Section 3.2 we discuss the relations between argumentation semantics and provide a graphical representation for these relations, backed up by argumentation frameworks that exhibit the unique traits of each semantics. We conclude the chapter with a summary in Section 3.3.

## 3.1 Evaluation Principles

In this section we explore several principles for evaluating argumentation semantics, as presented in (Baroni and Giacomin, 2007) and (Baroni et al., 2011a). We focus on those principles that are relevant for the results presented in the thesis and we provide an overview of their satisfaction, covering also semantics that have not been analyzed in this context in the argumentation literature.

### 3.1.1 Cardinality

First, we explore the characterization of argumentation semantics with respect to the cardinality of the set of extensions they provide. (Baroni et al., 2011a).

**Definition 34.** *Let $\mathcal{S}em$ be an argumentation semantics.*

*(i) $\mathcal{S}em$ is* ***universally defined*** *iff $\mathcal{S}em$ always provides at least one extension.*
*(ii) $\mathcal{S}em$ is a* ***unique-status*** *semantics iff $\mathcal{S}em$ always provides exactly one extension.*
*(iii) $\mathcal{S}em$ is a* ***multiple-status*** *semantics iff $\mathcal{S}em$ can provide multiple extensions.*

The three possible cases distinguished in (Baroni et al., 2011a) with respect to cardinality are: $\geq 0$ – multiple-status, not universally defined; $\geq 1$ – multiple-status, universally defined; $= 1$ – unique-status, universally defined.

The only argumentation semantics that are not universally defined are $\mathcal{ST}$ (stable) (Dung, 1995) and $\mathcal{ST}^P$ (prudent stable) (Coste-Marquis et al., 2005).

The grounded semantics ($\mathcal{GR}$) is unique-status (Dung, 1995). The ideal ($\mathcal{ID}$) (Dung et al., 2007) and the eager ($\mathcal{EAG}$) semantics (Caminada, 2006b) are also unique-status. A general proof for this is also provided in (Dvorak et al., 2011), in the context of parameterized ideal semantics, showing that the ideal semantics based on stage ($\mathcal{STA}^{id}$), naive ($\mathcal{MCF}^{id}$) and resolution-based grounded ($\mathcal{GR}^{*id}$) semantics are also unique-status. The proof only relies on the fact that the base semantics provide conflict-free extensions.

The skeptical versions of the preferred ($\mathcal{PR}^S$), semi-stable ($\mathcal{SST}^S$), stage ($\mathcal{STA}^S$), naive ($\mathcal{MCF}^S$) and resolution-based grounded ($\mathcal{GR}^{*S}$) semantics are unique-status by construction (the intersection of the extensions provided by the base semantics is the unique extension of the corresponding skeptical semantics). All the other argumentation semantics are multiple-status.

### 3.1.2 Conflict-freeness and admissibility

In this subsection we discuss possibly the most important principles related to the actual content of the extensions prescribed by an argumentation semantics, namely conflict-freeness and admissibility. While these ideas are also present in (Dung, 1995), they were formulated as general principles for argumentation semantics in (Baroni and Giacomin, 2007).

**Definition 35.** *Let $\mathcal{S}em$ be an argumentation semantics.*

*(i) $\mathcal{S}em$ satisfies the **conflict-free principle** iff, for any argumentation framework $F$, the $\mathcal{S}em$ extensions of $F$ are conflict-free sets.*

*(ii) $\mathcal{S}em$ satisfies the **admissibility principle** iff, for any argumentation framework $F$, the $\mathcal{S}em$ extensions of $F$ are admissible sets.*

All argumentation semantics satisfy the conflict-free principle, this being the very minimal constraint that all reasonable sets of arguments should satisfy in order to be accepted together (there is no rational reason for accepting conflicting arguments).

Admissibility on the other hand is not always satisfied, as in some approaches other properties are considered more useful. Conflict-free sets ($\mathcal{CF}$) are not admissible, but also they are not really used as a "real" argumentation semantics. All the other classical semantics ($\mathcal{CO}$, $\mathcal{ST}$, $\mathcal{PR}$, $\mathcal{GR}$) are admissible by construction (Dung, 1995).

The naive semantics ($\mathcal{MCF}$) does not satisfy admissibility, as the maximality constraint is not enough to ensure that contained arguments are defended. The stage semantics ($\mathcal{STA}$) does not satisfy admissibility either (Verheij, 1996). The semi-stable semantics, on the other hand, does satisfy it, by construction (maximal range is enforced for admissible sets) (Caminada, 2006b).

The resolution-based semantics ($\mathcal{GR}^*$, $\mathcal{PR}^*$, $\mathcal{SST}^*$ and $\mathcal{ID}^*$) satisfy admissibility because they provide complete extensions whenever the base semantics provides complete extensions (Baroni et al., 2011b).

The parameterized ideal sets ($\mathcal{IDS}$, $\mathcal{EAGS}$, $\mathcal{STA}^{ids}$, $\mathcal{MCF}^{ids}$ and $\mathcal{GR}^{*ids}$) and the parameterized ideal semantics ($\mathcal{ID}$, $\mathcal{EAG}$, $\mathcal{STA}^{id}$, $\mathcal{MCF}^{id}$ and $\mathcal{GR}^{*id}$) satisfy admissibility by construction (Dung et al., 2007; Caminada, 2007; Dvorak et al., 2011). Complete ideal sets ($\mathcal{PR}^{cids}$, $\mathcal{SST}^{cids}$ and $\mathcal{GR}^{*cids}$) are also admissible, as they provide complete extensions.

The skeptical semantics do not generally satisfy admissibility: we have already seen this for $\mathcal{PR}^S$, $\mathcal{SST}^S$, $\mathcal{STA}^S$ and $\mathcal{GR}^{*S}$, on the example framework from Figure 2.6. For that example, the skeptical naive extension happened to be admissible. On the other hand, let us see the very simple framework from Figure 3.1. We have $\mathcal{E}_{\mathcal{MCF}}(F) = \{\{b\}\}$, so $\mathcal{E}_{\mathcal{MCF}^S} = \{\{b\}\}$, but the set $\{b\}$ is not admissible.

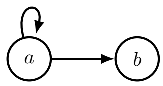



**Figure 3.1:** *The skeptical naive extension is not always admissible.*

The SCC-recursive semantics $\mathcal{AD}1$ and $\mathcal{AD}2$ satisfy admissibility (as they give complete extensions), while $\mathcal{CF}1$ and $\mathcal{CF}2$ are not admissible (Baroni et al., 2005). The *stage*2 semantics ($\mathcal{STA}2$) also fails to satisfy admissibility (Dvorak and Gaggl, 2012).

Aside from $\mathcal{CF}^P$, all prudent semantics ($\mathcal{AS}^P$, $\mathcal{CO}^P$, $\mathcal{ST}^P$, $\mathcal{PR}^P$ and $\mathcal{GR}^P$) satisfy admissibility, as they rely on $p$-admissible sets, which must also be admissible (Coste-Marquis et al., 2005).

The enhanced preferred semantics ($\mathcal{EPS}$) coincides with the preferred semantics whenever the latter provides at least one non-empty extension, but provides extensions that are not admissible otherwise (Zhang and Lin, 2010).

### 3.1.3 Reinstatement

The reinstatement principles (Baroni and Giacomin, 2007) require that, under certain constraints, arguments that are defended by an extension are included in the extension.

**Definition 36.** *Let $\mathcal{S}em$ be an argumentation semantics.*

*(i) $\mathcal{S}em$ satisfies the **strong reinstatement principle** iff, for all argumentation frameworks $F$, the $\mathcal{S}em$ extensions of $F$ contain all the arguments that they defend.*

*(ii) $\mathcal{S}em$ satisfies the **weak reinstatement principle** iff, for all argumentation frameworks $F$, the grounded extension of $F$ is included in all $\mathcal{S}em$ extensions of $F$.*

*(iii) $\mathcal{S}em$ satisfies the **$\mathcal{CF}$-reinstatement principle** iff, for all argumentation frameworks $F$, all $\mathcal{S}em$ extensions $E$ contain all arguments $a$ such that $E$ defends $a$ and $E \cup \{a\}$ is conflict-free.*

Note that we have departed slightly from the original terminology from (Baroni and Giacomin, 2007) by using "strong reinstatement" rather than just "reinstatement" in order to better distinguish between the three kinds of reinstatement. In their paper, Baroni and Giacomin show that reinstatement (strong reinstatement, in our case) implies both weak-reinstatement and $\mathcal{CF}$-reinstatement. In this context, we will often refer to weak-reinstatement and $\mathcal{CF}$-reinstatement as the weaker forms of reinstatement.

Given some argumentation semantics $\mathcal{S}em$, the following cases are possible:

(i) $\mathcal{S}em$ satisfies strong reinstatement and, hence, the weaker forms of reinstatement as well.
(ii) $\mathcal{S}em$ satisfies both weak-reinstatement and $\mathcal{CF}$-reinstatement, but not strong reinstatement
(iii) $\mathcal{S}em$ satisfies only weak reinstatement
(iv) $\mathcal{S}em$ satisfies only $\mathcal{CF}$-reinstatement
(v) $\mathcal{S}em$ does not satisfy any form of reinstatement.

From the very definition, the complete semantics $\mathcal{CO}$ satisfies strong reinstatement. Thus, all argumentation semantics that give complete extensions satisfy all forms of reinstatement. This observation covers the following:

- the stable ($\mathcal{ST}$), grounded ($\mathcal{GR}$) and preferred ($\mathcal{PR}$) semantics (Dung, 1995)
- the semi-stable semantics ($\mathcal{SST}$) (Caminada, 2006b)
- the resolution-based semantics: $\mathcal{GR}^*$, $\mathcal{PR}^*$, $\mathcal{ID}^*$ and $\mathcal{SST}^*$ (they provide complete extensions because the corresponding base semantics provide complete extensions) (Baroni et al., 2011b)
- the ideal semantics ($\mathcal{ID}$) (Dung et al., 2007), the eager semantics ($\mathcal{EAG}$) (Caminada, 2007) and the ideal semantics based on the resolution-based grounded semantics ($\mathcal{GR}^{*id}$) – the ideal version provides complete extensions because the base semantics satisfies reinstatement (Dvorak et al., 2011)
- the SCC-recursive semantics $\mathcal{AD}1$ and $\mathcal{AD}2$, shown to be complete in (Baroni et al., 2005)
- the complete ideal sets ($\mathcal{PR}^{cids}$, $\mathcal{SST}^{cids}$ and $\mathcal{GR}^{*cids}$) are complete by construction.

The survey in (Baroni and Giacomin, 2007) also covers the prudent semantics and shows that the prudent stable semantics $\mathcal{ST}^P$ satisfies strong reinstatement, while the prudent versions of grounded ($\mathcal{GR}^P$), complete ($\mathcal{CO}^P$) and preferred ($\mathcal{PR}^P$) semantics do not satisfy any form of reinstatement. Since $\mathcal{CF}^P$ and $\mathcal{AS}^P$ are even more general than these, we can conclude that they don't satisfy any kind of reinstatement either.

The $\mathcal{CF}2$ is shown to satisfy both weak and $\mathcal{CF}$-reinstatement, but not strong reinstatement (Baroni and Giacomin, 2007). The example used in the paper for showing that $\mathcal{CF}2$ does not satisfy strong reinstatement (an odd-length cycle) works for $\mathcal{CF}1$ as well. Also, it is shown in (Baroni et al., 2005) that $\mathcal{CF}1$ satisfies weak-reinstatement. We show that $\mathcal{CF}$-reinstatement is satisfied as well.

**Proposition 1.** *The $\mathcal{CF}1$ semantics satisfies $\mathcal{CF}$-reinstatement.*

*Proof.* We will prove the following: for any argumentation framework $F$, any set of arguments $C$ and any extension $E \in \mathcal{CF}1(F, C)$, all arguments $a \in C$ such that $E \cup \{a\}$ is conflict-free and $E$ defends $a$ are included in $E$. For clarity, we use $\mathcal{CF}1$ instead of $\mathcal{GF}$ for the generic recursive function. We proceed by induction on the size of the framework (the number of arguments). For the base case, the empty framework, the claim holds trivially, as the only extension is the empty set and there are no arguments.

For the induction step, we assume that the claim holds for all smaller frameworks and prove it for the current framework. We first consider the case when $|SCCS_F| > 1$. Let $S$ denote the strongly connected component that contains $a$. Then we must have $(E \cap S) \in \mathcal{CF}1(F\downarrow_{UP_F(S,E)}, C \cap U_F(S,E))$. Since $E$ defends $a$, we have that $a \in U_F(S,E)$. Since $a \in C$, we get that $a \in C \cap U_F(S,E)$. Furthermore, let us see that $E \cap S$ defends $a$ in $F\downarrow_{UP_F(S,E)}$. Indeed, let $b \in UP_F(S,E)$ be an attacker of $a$. Since $E$ defends $a$, there is an argument $c \in E$ such that $c \to b$. Now, if $c \notin S$, it would mean that $b \in D_F(S,E)$, which contradicts the fact that $b \in UP_F(S,E)$. So $c \in S \Rightarrow c \in E \cap S$ and $E \cap S$ defends $a$. Now we can apply the induction hypothesis and conclude that $a \in E \cap S$, so also $a \in E$.

Now let us consider the case $|SCCS_F| = 1$. Then $E \in \mathcal{MCF}(F\downarrow_C)$. Suppose that the argument $a$ is not in $E$. Then $E \cup \{a\}$ would be a larger conflict-free set in $F\downarrow_C$, which is a contradiction. Thus, we must have $a \in E$. This concludes our proof. □

Another principle-based evaluation is presented in (Dvorak and Gaggl, 2012), where the authors show that the *stage*2 semantics ($\mathcal{STA}2$) only satisfies the weak and $\mathcal{CF}$-reinstatement principles, while the naive ($\mathcal{MCF}$) and stage ($\mathcal{STA}$) only satisfy $\mathcal{CF}$-reinstatement.

For the skeptical semantics, let us first see that whenever the base semantics satisfies strong reinstatement, so does its skeptical version. Indeed, if the intersection of all extensions defends an argument, then each extension will defend it as well and, using the strong reinstatement property for the base semantics, must contain that argument. But then the intersection will contain it as well. So we conclude that the skeptical versions of the preferred ($\mathcal{PR}^S$), semi-stable ($\mathcal{SST}^S$) and resolution-based grounded ($\mathcal{GR}^{*S}$) semantics satisfy strong reinstatement.

Now let us see the framework in Figure 3.2. We have that $\mathcal{E}_{\mathcal{MCF}}(F) = \mathcal{E}_{\mathcal{STA}}(F) = \{\{a\},\{b\}\}$, so $\mathcal{E}_{\mathcal{MCF}^S}(F) = \mathcal{E}_{\mathcal{MCF}^{id}}(F) = \mathcal{E}_{\mathcal{STA}^S}(F) = \mathcal{E}_{\mathcal{STA}^{id}}(F) = \{\varnothing\}$. But $a$ is defended by the empty set (because it is unattacked) and also $\varnothing \cup \{a\}$ is conflict-free. Furthermore, the grounded extension of $F$ is $\{a\}$. These show that both weak and $\mathcal{CF}$-reinstatement (and hence strong reinstatement as well) are violated, so all four semantics fail to satisfy any form of reinstatement.

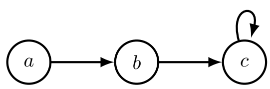



**Figure 3.2:** *Argumentation framework showing that the skeptical and ideal versions of $\mathcal{MCF}$ and $\mathcal{STA}$ do not satisfy any form of reinstatement.*

All argumentation semantics that have the empty set as an extension even for frameworks containing unattacked arguments do not satisfy any of the reinstatement principles. This covers the conflict-free ($\mathcal{CF}$) and admissible ($\mathcal{AS}$) sets, but also all ideal sets ($\mathcal{IDS}$, $\mathcal{EAGS}$, $\mathcal{STA}^{ids}$, $\mathcal{MCF}^{ids}$ and $\mathcal{GR}^{*ids}$).

Since the satisfaction of the reinstatement principles is not covered in the literature for the enhanced preferred semantics, we provide the relevant results in Proposition 2.

**Proposition 2.** *The enhanced preferred semantics ($\mathcal{EPS}$) satisfies weak and $\mathcal{CF}$-reinstatement but does not satisfy strong reinstatement.*

*Proof.* As shown in (Dvorak and Gaggl, 2012), semantics that can select non-empty conflict-free sets out of odd-length cycles do not satisfy strong reinstatement. Since $\mathcal{EPS}$ falls into this category, we get that it does not satisfy strong reinstatement.

For the weak reinstatement, we have two cases. If the framework has at least one non-empty admissible set, then $\mathcal{EPS}$ coincides with $\mathcal{PR}$ so its extensions will satisfy weak reinstatement. If the empty set is the only admissible set, then weak reinstatement is trivially satisfied, as the grounded extension is the empty set in this case.

For the $\mathcal{CF}$-reinstatement, let $F$ be an argumentation framework and let $S$ be an enhanced preferred extension that comes from the minimal admissible pair $(S, H)$. Let $a$ be an argument such that $S$ defends $a$ and $S \cup \{a\}$ is conflict-free. Suppose that $a \in H$. Then the pair $(S \cup \{a\}, H \setminus \{a\})$ is also admissible, which violates the minimality of $H$ with respect to cardinality. So we have that $a \notin H$. But then $(S \cup \{a\}, H)$ is also an admissible pair. Since $S$ is maximal among the minimal admissible pairs, it must be that $S \cup \{a\} = S \Rightarrow a \in S$, which proves $\mathcal{CF}$-reinstatement. □

### 3.1.4 Non-interference, additivity and directionality

The intuition behind non-interference (Baroni et al., 2011a) is that, given an isolated set of arguments (no conflict with outside arguments) one should be able to select the elements of the extension independently from the rest of the framework. We formalize this in Definition 37.

**Definition 37.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S$ is said to be* ***isolated*** *in $F$ iff $\mathcal{R} \cap ((\mathcal{A} \times (\mathcal{A} \setminus S)) \cup ((\mathcal{A} \setminus S) \times \mathcal{A})) = \varnothing$ (there is no conflict between the set and the rest of the framework). We will use $\mathcal{IS}(F)$ to refer the isolated sets of $F$.*

*An argumentation semantics $\mathcal{S}em$ is said to satisfy the* ***non-interference principle*** *iff, for any argumentation framework $F = (\mathcal{A}, \mathcal{R})$ and any isolated set of arguments $S \in \mathcal{IS}(F)$ we have that*

$$\mathcal{AE}_{\mathcal{S}em}(F, S) = \mathcal{E}_{\mathcal{S}em}(F{\downarrow_S}) \tag{3.1}$$

*where $\mathcal{AE}_{\mathcal{S}em}(F, S) \triangleq \{E \cap S \mid E \in \mathcal{E}_{\mathcal{S}em}(F)\}$.*

In words, the intersection of an extension with an isolated set $S$ should be equal to one of the extensions prescribed by $\mathcal{S}em$ for the restricted framework $F{\downarrow_S}$, and vice versa: given an extension of the restricted framework, one should be able to complete it with arguments from the rest of the framework so as to get an extension of $F$.

**Definition 38.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. A set of arguments $S$ is said to be* ***unattacked*** *in $F$ iff no argument from $S$ is attacked by an argument that is not in $S$, i.e. $(\mathcal{A} \setminus S) \nrightarrow S$. We will use $\mathcal{US}(F)$ to refer to all unattacked sets of $F$.*

*An argumentation semantics $\mathcal{S}em$ is said to satisfy the* ***directionality principle*** *iff, for any argumentation framework $F = (\mathcal{A}, \mathcal{R})$ and any unattacked set $S \in \mathcal{US}(F)$, the following holds:*

$$\mathcal{AE}_{\mathcal{S}em}(F, S) = \mathcal{E}_{\mathcal{S}em}(F{\downarrow}_S) \tag{3.2}$$

The directionality principle (Baroni and Giacomin, 2007) requires the same property as non-interference, but for unattacked sets. Since isolated sets are also unattacked, directionality is a stronger property.

We also introduce a new property, closely related to non-interference, which we call additivity. The relevant concepts are introduced in Definition 39.

**Definition 39.** *An argumentation semantics $\mathcal{S}em$ is* ***additive*** *(or, alternatively, $\mathcal{S}em$ satisfies the* ***additivity principle****) iff, for any two disjoint ($\mathcal{A}_1 \cap \mathcal{A}_2 = \varnothing$) argumentation frameworks $F_1 = (\mathcal{A}_1, \mathcal{R}_1)$ and $F_2 = (\mathcal{A}_2, \mathcal{R}_2)$, the following relation holds:*

$$\mathcal{E}_{\mathcal{S}em}(F_1 \uplus F_2) = \mathcal{E}_{\mathcal{S}em}(F_1) \uplus \mathcal{E}_{\mathcal{S}em}(F_2) \tag{3.3}$$

*where $F_1 \uplus F_2 \triangleq (\mathcal{A}_1 \cup \mathcal{A}_2, \mathcal{R}_1 \cup \mathcal{R}_2)$ denotes the* ***disjoint union*** *of $F_1$ and $F_2$ and, for any two sets of extensions $\mathcal{E}_1$ and $\mathcal{E}_2$, $\mathcal{E}_1 \uplus \mathcal{E}_2 \triangleq \{S_1 \cup S_2 \mid S_1 \in \mathcal{E}_1, S_2 \in \mathcal{E}_2\}$.*

In words, the additivity property says that the extensions of the disjoint union of two argumentation frameworks can be computed by considering the extensions of each framework and taking all the unions between such pairs. We provide an equivalent formulation of this principle, one that is more similar to SCC-recursiveness, but using weakly connected components instead of SCC's.

**Definition 40.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. Two arguments are* ***connected*** *in $F$ iff there is a path from $a$ to $b$ ignoring the direction of the attacks (the path may also be of length 0 – every argument is connected to itself). Connectedness is an equivalence relation and its equivalence classes are called* ***connected components****. We will use $CC_F$ to refer to the set of connected components of $F$.*

**Proposition 3.** *An argumentation semantics $\mathcal{S}em$ satisfies the additivity principle iff, for any argumentation framework $F$ and any set of arguments $E$, the following holds:*

$$E \in \mathcal{E}_{\mathcal{S}em}(F) \Leftrightarrow \forall S (S \in CC_F \Rightarrow (E \cap S) \in \mathcal{E}_{\mathcal{S}em}(F{\downarrow}_S)) \tag{3.4}$$

The additivity property, coupled with the alternative formulation in terms of connected components, is quite useful for compactly writing the extensions of large argumentation frameworks for semantics that generally provide many extensions, provided that these semantics do satisfy the property. We will make use of this observation in the next section.

While additivity and non-interference are very similar, they are still distinct properties. The exact connection between the two is given in Proposition 4.

**Proposition 4.** *Let $\mathcal{S}em$ be an argumentation semantics. The following properties hold:*

*(i) If $\mathcal{S}em$ is universally defined and additive, then $\mathcal{S}em$ satisfies the non-interference principle.*

*(ii) If $\mathcal{S}em$ satisfies the non-interference principle then, for any two disjoint argumentation frameowrks $F_1$ and $F_2$, $\mathcal{E}_{\mathcal{S}em}(F_1 \uplus F_2) \subseteq \mathcal{E}_{\mathcal{S}em}(F_1) \uplus \mathcal{E}_{\mathcal{S}em}(F_2)$*

*Proof.* For (i), let us see that, given an argumentation framework $F = (\mathcal{A}, \mathcal{R})$ and an isolated set $S$, the restricted frameworks $F{\downarrow_S}$ and $F{\downarrow_{\mathcal{A}\setminus S}}$ are disjoint and, even more, $F = F{\downarrow_S} \uplus F{\downarrow_{\mathcal{A}\setminus S}}$. Since $\mathcal{S}em$ is assumed additive, we have $\mathcal{E}_{\mathcal{S}em}(F) = \mathcal{E}_{\mathcal{S}em}(F{\downarrow_S}) \uplus \mathcal{E}_{\mathcal{S}em}(F{\downarrow_{\mathcal{A}\setminus S}})$. But then we have $\mathcal{A}\mathcal{E}_{\mathcal{S}em}(F, S) = \{(E_1 \cup E_2) \cap S \mid E_1 \in \mathcal{E}_{\mathcal{S}em}(F{\downarrow_S}), E_2 \in \mathcal{E}_{\mathcal{S}em}(F{\downarrow_{\mathcal{A}\setminus S}})\} = \{E_1 \mid E_1 \in \mathcal{E}_{\mathcal{S}em}(F{\downarrow_S}) \text{ and } \mathcal{E}_{\mathcal{S}em}(F{\downarrow_{\mathcal{A}\setminus S}}) \neq \emptyset\} = \mathcal{E}_{\mathcal{S}em}(F{\downarrow_S})$, which means that $\mathcal{S}em$ satisfies non-interference. Note that in the last equality we have also used the fact that $\mathcal{S}em$ is universally defined.

For (ii) let us see that, given two disjoint frameworks $F_1 = (\mathcal{A}_1, \mathcal{R}_1)$ and $F_2 = (\mathcal{A}_2, \mathcal{R}_2)$, we have the following $\mathcal{A}\mathcal{E}_{\mathcal{S}em}(F_1 \uplus F_2, \mathcal{A}_1) = \mathcal{E}_{\mathcal{S}em}(F_1) \Rightarrow \forall E(E \in \mathcal{E}_{\mathcal{S}em}(F_1 \uplus F_2) \Rightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em}(F_1))$. Similarly, $\forall E(E \in \mathcal{E}_{\mathcal{S}em}(F_1 \uplus F_2) \Rightarrow E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em}(F_2))$. But for extensions $E$ of the joint framework we have $E = (E \cap \mathcal{A}_1) \cup (E \cap \mathcal{A}_2)$, so we can deduce the desired $\mathcal{E}_{\mathcal{S}em}(F_1 \uplus F_2) \subseteq \mathcal{E}_{\mathcal{S}em}(F_1) \uplus \mathcal{E}_{\mathcal{S}em}(F_2)$.

To see that the inclusion may be strict, consider the argumentation frameworks

$$\begin{aligned} \mathcal{F}_1 &= (\{a, b\}, \{(a, b), (b, a)\}) \\ \mathcal{F}_2 &= (\{c, d\}, \{(c, d), (d, c)\}) \end{aligned} \tag{3.5}$$

and $F = F_1 \uplus F_2$. As we shall see further on, the only argumentation semantics that does not satisfy additivity does not satisfy non-interference either, so we are going to use a fictitious semantics $\mathcal{S}em$ here. Suppose that the argumentation semantics $\mathcal{S}em$ gives the following extensions:

$$\begin{aligned} \mathcal{E}_{\mathcal{S}em}(F_1) &= \{\{a\}, \{b\}\} \\ \mathcal{E}_{\mathcal{S}em}(F_2) &= \{\{c\}, \{d\}\} \\ \mathcal{E}_{\mathcal{S}em}(F) &= \{\{a, c\}, \{b, d\}\} \end{aligned} \tag{3.6}$$

It is not difficult to see that we have $\mathcal{A}\mathcal{E}_{\mathcal{S}em}(F, \{a, b\}) = \mathcal{E}_{\mathcal{S}em}(F{\downarrow_{\{a,b\}}})$ and $\mathcal{A}\mathcal{E}_{\mathcal{S}em}(F, \{c, d\}) = \mathcal{E}_{\mathcal{S}em}(F{\downarrow_{\{c,d\}}})$, so the non-interference property is satisfied. However, note that $\mathcal{E}_{\mathcal{S}em}(F_1) \uplus \mathcal{E}_{\mathcal{S}em}(F_2) = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\} \neq \mathcal{E}_{\mathcal{S}em}(F)$. □

In words, non-interference ensures only a part of additivity, while additivity implies non-interference, but only for universally defined semantics. Since we have seen already that most argumentation semantics are universally defined, we will prove additivity and get non-interference for free.

### Additivity

In what follows, we will show that almost all argumentation semantics satisfy the additivity principle. We start with SCC-recursive semantics.

**Proposition 5.** *All SCC-recursive semantics are additive.*

*Proof.* Let $F_1 = (\mathcal{A}_1, \mathcal{R}_1)$ and $F_2 = (\mathcal{A}_2, \mathcal{R}_2)$ be two disjoint argumentation frameworks and let $F = F_1 \uplus F_2$ be their disjoint union. We have $E \in \mathcal{E}_{\mathcal{S}em}(F) \Leftrightarrow E \in \mathcal{GF}(F, \mathcal{A}_1 \cup \mathcal{A}_2) \Leftrightarrow \forall S(S \in SCCS_F \Rightarrow (E \cap S) \in \mathcal{GF}(F{\downarrow_{UP_F(S,E)}}, (\mathcal{A}_1 \cup$

$\mathcal{A}_2) \cap U_F(S,E)))$. But since $SCCS_F = SCCS_{F_1} \cup SCCS_{F_2}$ (strongly connected components cannot span across both frameworks), we can write $UP_F(S,E) = UP_{F_1}(S, E \cap \mathcal{A}_1)$ and $U_F(S,E) = U_{F_1}(S, E \cap \mathcal{A}_1)$ for $S \in SCCS_{F_1}$ and similar relations for $S \in SCCS_{F_2}$. From these, using the SCC-recursiveness relation for $F_1$ and $F_2$ we get the desired result $E \in \mathcal{E}_{\mathcal{S}em}(F) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em}(F_2)$. □

So, we have managed to show that the SCC-recursive semantics $\mathcal{AD}1$, $\mathcal{AD}2$, $\mathcal{CF}1$, $\mathcal{CF}2$ and $\mathcal{STA}2$ are additive. Additionally, using the fact that the classical semantics are SCC-recursive as well (Baroni et al., 2005), we get that $\mathcal{CF}$, $\mathcal{AS}$, $\mathcal{CO}$, $\mathcal{ST}$, $\mathcal{GR}$ and $\mathcal{PR}$ are also additive. Note that we have also included conflict-free sets, as it is rather easy to see that taking $\mathcal{BF}_{CF}(F,C) = \mathcal{E}_{\mathcal{CF}}(F)$ we can show that $\mathcal{CF}$ is SCC-recursive as well. Next, we cover the stage and semi-stable semantics.

**Proposition 6.** *The stage and the semi-stable semantics are additive.*

*Proof.* We only provide the proof for the stage semantics, as the one for the semi-stable semantics is very similar. Let $\mathcal{F}_1 = (\mathcal{A}_1, \mathcal{R}_1)$ and $\mathcal{F}_2 = (\mathcal{A}_2, \mathcal{R}_2)$ be two disjoint argumentation frameworks and let $F = F_1 \uplus F_2$ be their disjoint union. We will show that $E \in \mathcal{E}_{\mathcal{STA}}(F_1 \uplus F_2) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{STA}}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{STA}}(F_2)$. For the direct implication, consider $E \in \mathcal{E}_{\mathcal{STA}}(F)$. Then $E$ is conflict-free and, since $\mathcal{CF}$ is additive, we have that $E \cap \mathcal{A}_1$ is conflict-free as well. All that is left to prove is that $range_F(E \cap \mathcal{A}_1)$ is maximal among the conflict-free sets of $F_1$. Suppose that this is not the case. Then there exists $E' \in \mathcal{E}_{\mathcal{CF}}(F_1)$ such that $range_{F_1}(E \cap \mathcal{A}_1) \subsetneq range_{F_1}(E')$. But then $range_F(E' \cup (E \cap \mathcal{A}_2)) = range_{F_1}(E') \cup range_{F_2}(E \cap \mathcal{A}_2) \supsetneq range_{F_1}(E \cap \mathcal{A}_1) \cup range_{F_2}(E \cap \mathcal{A}_2) = range_F(E)$, which contradicts the maximality of $range_F(E)$. The proof for the converse is similar. The key observation is that the range of a set $S \subseteq \mathcal{A}_1$ is included in $\mathcal{A}_1$, even when computed with respect to $F_1 \uplus F_2$. □

For the rest of the semantics, we employ the following strategy: we first prove some general results that allow us to infer the additivity of an argumentation semantics using the additivity of other semantics (Theorem 1), then we use these results and cover the large majority of argumentation semantics, and finally we deal with the few remaining semantics.

**Theorem 1.** *For any argumentation semantics $\mathcal{S}em$, $\mathcal{S}em_1$ and $\mathcal{S}em_2$, the following results hold:*

*(a) if $\mathcal{S}em$ is additive, then the maximal semantics based on $\mathcal{S}em$ – $\mathcal{S}em^M$ – given by $\mathcal{E}_{\mathcal{S}em}(F) = \{S \in \mathcal{E}_{\mathcal{S}em}(F) \mid S$ is maximal w.r.t. $\subseteq\}$ is also additive*

*(b) if $\mathcal{S}em$ is additive, then the minimal semantics based on $\mathcal{S}em$ – $\mathcal{S}em^m$ – given by $\mathcal{E}_{\mathcal{S}em}(F) = \{S \in \mathcal{E}_{\mathcal{S}em}(F) \mid S$ is minimal w.r.t. $\subseteq\}$ is also additive*

*(c) if $\mathcal{S}em$ is additive, then the skeptical semantics based on $\mathcal{S}em$ – $\mathcal{S}em^S$ – is also additive*

*(d) if $\mathcal{S}em$ is additive, then the ideal sets based on $\mathcal{S}em$ – $\mathcal{S}em^{ids}$ – are also additive*

*(e) if $\mathcal{S}em$ is additive, then the ideal semantics based on $\mathcal{S}em$ – $\mathcal{S}em^{id}$ – is also additive*

*(f) if $\mathcal{S}em$ is additive, then the resolution-based version of $\mathcal{S}em$ – $\mathcal{S}em^*$ – is also additive*

*(g) if $\mathcal{S}em_1$ and $\mathcal{S}em_2$ are additive, then their intersection – $\mathcal{S}em_1 \cap \mathcal{S}em_2$ – given by $\mathcal{E}_{\mathcal{S}em_1 \cap \mathcal{S}em_2}(F) = \mathcal{E}_{\mathcal{S}em_1}(F) \cap \mathcal{E}_{\mathcal{S}em_2}(F)$ is also additive*

*Proof.* Let $F_1 = (\mathcal{A}_1, \mathcal{R}_1)$ and $F_2 = (\mathcal{A}_2, \mathcal{R}_2)$ be two disjoint argumentation frameworks and let $F = F_1 \uplus F_2$ be their disjoint union. Also consider a set of arguments $E \subseteq \mathcal{A}_1 \cup \mathcal{A}_2$. We will use them for all the proofs.

(a) We will prove that $E \in \mathcal{E}_{\mathcal{S}em^M}(F) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em^M}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em^M}(F_2)$. For the direct implication, we have $E \in \mathcal{E}_{\mathcal{S}em^M}(F)$. Then we also have $E \in \mathcal{E}_{\mathcal{S}em}(F)$ which, based on the additivity of $\mathcal{S}em$, leads to $E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em}(F_1)$. Let us see that $E \cap \mathcal{A}_1$ is maximal among the $\mathcal{S}em$ extensions of $F_1$. Suppose that this is not the case. Then there is an extension $E' \in \mathcal{E}_{\mathcal{S}em}(F_1)$ such that $E \cap \mathcal{A}_1 \subsetneq E'$. But then we also have $E \subsetneq E' \cup (E \cap \mathcal{A}_2)$ and, since $\mathcal{S}em$ is additive, $E' \cup (E \cap \mathcal{A}_2) \in \mathcal{E}_{\mathcal{S}em}(F)$, which contradicts the maximality of $E$. Similarly we get $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em}(F_2)$. The proof of the converse is based on the same approach.

(b) The proof is very similar to that of (a)

(c) We have:

$$\begin{aligned}\mathcal{E}_{\mathcal{S}em^S}(F) &= \{ \bigcap_{E \in \mathcal{E}_{\mathcal{S}em}(F)} E \} = \{ \bigcap_{E_1 \in \mathcal{E}_{\mathcal{S}em}(F_1), E_2 \in \mathcal{E}_{\mathcal{S}em}(F_2)} (E_1 \cup E_2) \} \\ &=^* \{ \bigcap_{E_1 \in \mathcal{E}_{\mathcal{S}em}(F_1)} E_1 \cup \bigcap_{E_2 \in \mathcal{E}_{\mathcal{S}em}(F_2)} E_2 \} = \mathcal{E}_{\mathcal{S}em^S}(F_1) \uplus \mathcal{E}_{\mathcal{S}em^S}(F_2)\end{aligned} \tag{3.7}$$

which is the desired result. Note that the starred equality holds because the two frameworks are disjoint, while the last one is a consequence of the fact that the skeptical semantics provide a single extension. We have also used the additivity of $\mathcal{S}em$.

(d) We show that $E \in \mathcal{E}_{\mathcal{S}em^{ids}}(F) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em^{ids}}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em^{ids}}(F_2)$. For the direct implication, consider an ideal set $E \in \mathcal{E}_{\mathcal{S}em^{ids}}(F)$. Then $E$ is admissible and, since $\mathcal{AS}$ is additive, $E \cap \mathcal{A}_1$ is admissible as well. Furthermore, $E$ is included in all $\mathcal{S}em$ extensions of $F$, so it is also included in their intersection, the skeptical $\mathcal{S}em$ extension of $F$. Using (c) we get that $E \cap \mathcal{A}_1 \subseteq \mathcal{E}_{\mathcal{S}em^S}(F_1)$, so $E \cap \mathcal{A}_1$ is included in all $\mathcal{S}em$ extensions of $F_1$, which means that $E \cap \mathcal{A}_1$ is a $\mathcal{S}em$ ideal set of $F_1$. Similarly we get $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em^{ids}}(F_2)$. The proof for the converse uses the same approach.

(e) The result follows from (d) and (a).

(f) The resolution-based version of $\mathcal{S}em$ is in fact the minimal version of the semantics $\mathcal{S}em^r$, given by $\mathcal{E}_{\mathcal{S}em^r}(F) = \bigcup_{F' \in \mathcal{RES}(F)} \mathcal{E}_{\mathcal{S}em}(F')$. In virtue of (b), in order to show that $\mathcal{S}em^*$ is additive it is enough to prove that $\mathcal{S}em^r$ is additive. We will prove $E \in \mathcal{E}_{\mathcal{S}em^r}(F) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em^r}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em^r}(F_2)$. For the direct implication, consider an extension $E \in \mathcal{E}_{\mathcal{S}em^r}(F)$. Then there is a full resolution $F' \in \mathcal{RES}(F)$ such that $E \in \mathcal{E}_{\mathcal{S}em}(F')$. But then, since $\mathcal{S}em$ is additive, $E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em}(F' \downarrow_{\mathcal{A}_1})$. But $F' \downarrow_{\mathcal{A}_1} \in \mathcal{RES}(F_1)$ and, thus $E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em^r}(F_1)$. Similarly we show that $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em^r}(F_2)$. The proof for the converse relies on the same intuition, namely that the restriction of a full resolution to an isolated set is a full resolution of the restriction of the original framework to the same set.

(g) We have $E \in \mathcal{E}_{\mathcal{S}em_1 \cap \mathcal{S}em_2}(F) \Leftrightarrow E \in \mathcal{E}_{\mathcal{S}em_1}(F)$ and $E \in \mathcal{E}_{\mathcal{S}em_2}(F) \Leftrightarrow$ $(E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em_1}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em_1}(F_2))$ and $(E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em_2}(F_1)$ and

$E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em_2}(F_2)) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{S}em_1 \cap \mathcal{S}em_2}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{S}em_1 \cap \mathcal{S}em_2}(F_2)$, which leads to the desired result. □

We now use the results from Theorem 1. The naive semantics ($\mathcal{MCF}$) is additive because it is the maximal version of $\mathcal{CF}$, so we can use (a). The resolution-based versions ($\mathcal{GR}^*$, $\mathcal{PR}^*$, $\mathcal{SST}^*$ and $\mathcal{ID}^*$), the skeptical versions ($\mathcal{PR}^S$, $\mathcal{SST}^S$, $\mathcal{STA}^S$, $\mathcal{MCF}^S$ and $\mathcal{GR}^{*S}$), the ideal sets ($\mathcal{IDS}$, $\mathcal{EAGS}$, $\mathcal{STA}^{ids}$, $\mathcal{MCF}^{ids}$ and $\mathcal{GR}^{*ids}$) and the ideal semantics ($\mathcal{ID}$, $\mathcal{EAG}$, $\mathcal{STA}^{id}$, $\mathcal{MCF}^{id}$ and $\mathcal{GR}^{*id}$) are covered by items (f), (c), (d) and (e), respectively.

Furthermore, we can use the fact that complete ideal sets $\mathcal{S}em^{cids}$ are ideal sets $\mathcal{S}em^{ids}$ that are also complete, i.e. $\mathcal{S}em^{cids} = \mathcal{S}em^{ids} \cap \mathcal{CO}$, and apply (g) to get that $\mathcal{PR}^{cids}$, $\mathcal{SST}^{cids}$ and $\mathcal{GR}^{*cids}$ are additive.

Next, we deal with the prudent semantics.

**Proposition 7.** *All prudent semantics ($\mathcal{CF}^P$, $\mathcal{AS}^P$, $\mathcal{CO}^P$, $\mathcal{ST}^P$, $\mathcal{GR}^P$ and $\mathcal{PR}^P$) are additive.*

*Proof.* The proofs are very similar to those that we have provided already, so we will just give the general idea for each case. We consider, as usual, two disjoint argumentation frameworks $F_1$ and $F_2$.

For $\mathcal{CF}^P$, the proof relies on the simple observation that indirect attacks can only occur between arguments from the same framework ($F_1$ or $F_2$). For $\mathcal{AS}^P$ we use the fact that $\mathcal{AS}^P = \mathcal{CF}^P \cap \mathcal{AS}$ and Theorem 1 (g). We use the same argument for $\mathcal{ST}^P$, since $\mathcal{ST}^P = \mathcal{CF}^P \cap \mathcal{ST}$. For $\mathcal{PR}^P$ we use the fact that it is the maximal version of $\mathcal{AS}^P$ and Theorem 1 (a). For $\mathcal{GR}^P$ the proof relies on $\mathcal{F}^P_{F_1 \uplus F_2}(S) = \mathcal{F}^P_{F_1}(S \cap \mathcal{A}_1) \cup \mathcal{F}^P_{F_2}(S \cap \mathcal{A}_2)$. The proof for $\mathcal{CO}^P$ is based on the same intuition □

The only argumentation semantics we have not discussed yet is $\mathcal{EPS}$. The argumentation framework in Figure 3.3 shows that $\mathcal{EPS}$ is not additive. Indeed, we have $\mathcal{E}_{\mathcal{EPS}}(F) = \{\{a\},\{b\},\{c\},\{d\},\{e\},\{f\}\}$, but $\mathcal{E}_{\mathcal{EPS}}(F_1) \uplus \mathcal{E}_{\mathcal{EPS}}(F_2) = \{\{a\},\{b\},\{c\}\} \uplus \{\{d\},\{e\},\{f\}\} = \{\{a,d\},\{a,e\},\{a,f\},\{b,d\},\{b,e\},\{b,f\},\{c,d\}, \{c,e\},\{c,f\}\}$, so additivity is violatd.

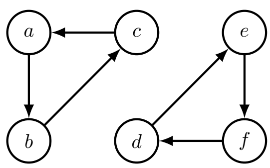



**Figure 3.3:** *The enhanced preferred semantics is not additive:* $F_1 = (\{a,b,c\}, \{(a,b),(b,c),(c,a)\})$, $F_2 = (\{d,e,f\},\{(d,e),(e,f),(f,d)\})$

To sum up, we have showed that all semantics except $\mathcal{EPS}$ are additive. So, although additivity has not yet been considered in the argumentation literature as a property of extension-based semantics, it turns out that it is nevertheless satisfied by the majority of semantics. It is also obvious that this property is desirable, as it can be regarded as a different form of non-interference.

### Non-interference

From Proposition 4 we get that all argumentation semantics that are universally defined and additive also satisfy the non-interference principle. This covers all the semantics we have considered in Chapter 2, except $\mathcal{ST}$ and $\mathcal{ST}^P$ (not universally defined) and $\mathcal{EPS}$ (not additive). So we just need to deal with these three semantics.

The example we provided in Figure 3.3 for showing that $\mathcal{EPS}$ is not additive also shows that $\mathcal{EPS}$ does not satisfy non-interference. Indeed, for the isolated set $S = \{a, b, c\}$, we have $\mathcal{AE}_{\mathcal{EPS}}(F, S) = \{\varnothing, \{a\}, \{b\}, \{c\}\}$, while $\mathcal{E}_{\mathcal{EPS}}(F{\downarrow_S}) = \{\{a\}, \{b\}, \{c\}\}$.

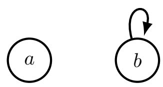



**Figure 3.4:** *$\mathcal{ST}$ and $\mathcal{ST}^P$ do not satisfy the non-interference principle.*

Furthermore, the two argumentation semantics that are not universally defined, $\mathcal{ST}$ and $\mathcal{ST}^P$, do not satisfy non-interference either, as can be seen for the very simple example from figure 3.4. Indeed, let $F$ be the framework in the figure and let $S = \{a\}$ be the isolated set. We have $\mathcal{E}_{\mathcal{ST}}(F) = \mathcal{E}_{\mathcal{ST}^P}(F) = \varnothing$, so $\mathcal{AE}_{\mathcal{ST}}(F, S) = \mathcal{AE}_{\mathcal{ST}^P}(F, S) = \varnothing$. But $\mathcal{E}_{\mathcal{ST}}(F{\downarrow_S}) = \mathcal{E}_{\mathcal{ST}^P}(F{\downarrow_S}) = \{\{a\}\}$.

Note that the satisfaction of non-interference is also discussed in (Baroni et al., 2011a) for $\mathcal{CO}$, $\mathcal{GR}$, $\mathcal{PR}$, $\mathcal{ST}$, $\mathcal{SST}$, $\mathcal{ID}$, $\mathcal{STA}$ and $\mathcal{CF}2$, but a justification is only provided for the stable semantics.

### Directionality

As far as directionality is concerned, we will start by mentioning the results from (Baroni and Giacomin, 2007) and (Baroni et al., 2011a). First, we have that directionality implies non-interference, so the stable and the prudent stable semantics do not satisfy directionality. The other classical semantics ($\mathcal{CF}$, $\mathcal{AS}$, $\mathcal{CO}$, $\mathcal{GR}$, $\mathcal{PR}$), the ideal semantics ($\mathcal{ID}$) and the universally defined SCC-recursive semantics ($\mathcal{AD}1$, $\mathcal{AD}2$, $\mathcal{CF}1$, $\mathcal{CF}2$ and $\mathcal{STA}2$) do satisfy it. Of the resolution-based semantics, it is shown in (Baroni et al., 2011b) that only $\mathcal{GR}^*$ satisfies directionality, while $\mathcal{PR}^*$, $\mathcal{SST}^*$ and $\mathcal{ID}^*$ fail to do so.

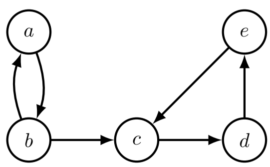



**Figure 3.5:** *$\mathcal{SST}$, $\mathcal{STA}$ and the argumentation semantics based on them do not satisfy the directionality principle.*

The semi-stable ($\mathcal{SST}$) and the stage ($\mathcal{STA}$) semantics do not satisfy directionality (Baroni et al., 2011a), based on the example argumentation framework

| $\mathcal{S}em$ | $\mathcal{E}_{\mathcal{S}em}(F)$ | $\mathcal{AE}_{\mathcal{S}em}(F, S)$ | $\mathcal{E}_{\mathcal{S}em}(F{\downarrow_S})$ |
|---|---|---|---|
| $\mathcal{SST}$, $\mathcal{STA}$ | $\{\{b, d\}\}$ | $\{\{b\}\}$ | $\{\{a\}, \{b\}\}$ |
| $\mathcal{SST}^S$, $\mathcal{STA}^S$ | $\{\{b, d\}\}$ | $\{\{b\}\}$ | $\{\varnothing\}$ |
| $\mathcal{EAGS}$, $\mathcal{STA}^{ids}$ | $\{\varnothing, \{b\}, \{b, d\}\}$ | $\{\varnothing, \{b\}\}$ | $\{\varnothing\}$ |
| $\mathcal{SST}^{cids}$ | $\{\varnothing, \{b, d\}\}$ | $\{\varnothing, \{b\}\}$ | $\{\varnothing\}$ |
| $\mathcal{EAG}$, $\mathcal{STA}^{id}$ | $\{\{b, d\}\}$ | $\{\{b\}\}$ | $\{\varnothing\}$ |

**Table 3.1:** *Extensions for various semantics based on $\mathcal{SST}$ and $\mathcal{STA}$, for the argumentation framework from Figure 3.5. The unattacked set $S$ is $\{a, b\}$.*

that we reproduce in Figure 3.5. We use the same framework for extending the result to argumentation semantics that are based on $\mathcal{SST}$ and $\mathcal{STA}$. The relevant results are presented in Table 3.1. What we get is that $\mathcal{SST}^S$, $\mathcal{EAGS}$, $\mathcal{SST}^{cids}$, $\mathcal{EAG}$, $\mathcal{STA}^S$, $\mathcal{STA}^{ids}$ and $\mathcal{STA}^{id}$ do not satisfy directionality.

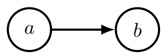



**Figure 3.6:** *$\mathcal{MCF}$ and semantics based on it do not satisfy directionality.*

| $\mathcal{S}em$ | $\mathcal{E}_{\mathcal{S}em}(F)$ | $\mathcal{AE}_{\mathcal{S}em}(F, S)$ | $\mathcal{E}_{\mathcal{S}em}(F{\downarrow_S})$ |
|---|---|---|---|
| $\mathcal{MCF}$ | $\{\{a\}, \{b\}\}$ | $\{\varnothing, \{a\}\}$ | $\{\{a\}\}$ |
| $\mathcal{MCF}^S$ | $\{\varnothing\}$ | $\{\varnothing\}$ | $\{\{a\}\}$ |
| $\mathcal{MCF}^{ids}$ | $\{\varnothing\}$ | $\{\varnothing\}$ | $\{\varnothing, \{a\}\}$ |
| $\mathcal{MCF}^{id}$ | $\{\varnothing\}$ | $\{\varnothing\}$ | $\{\{a\}\}$ |

**Table 3.2:** *Extensions for semantics based on $\mathcal{MCF}$, for the argumentation framework from Figure 3.6. The unattacked set $S$ is $\{a\}$.*

Furthermore, a very simple argumentation framework, such as that from Figure 3.6, shows that $\mathcal{MCF}$ and semantics based on it ($\mathcal{MCF}^S$, $\mathcal{MCF}^{ids}$ and $\mathcal{MCF}^{id}$) also fail to satisfy directionality. The relevant results are presented in Table 3.2.

Next, we provide a general positive result showing that whenever an argumentation semantics $\mathcal{S}em$ does satisfy the directionality principle, the argumentation semantics based on $\mathcal{S}em$ satisfy it as well.

**Theorem 2.** *Let $\mathcal{S}em$ be an argumentation semantics that satisfies directionality. Then the following results hold:*

*(a) the skeptical version of $\mathcal{S}em$ ($\mathcal{S}em^S$) satisfies directionality*

*(b) the ideal sets based on $\mathcal{S}em$ ($\mathcal{S}em^{ids}$) satisfy directionality*

*(c) if $\mathcal{S}em$ gives complete extensions, then the complete ideal sets based on $\mathcal{S}em$ ($\mathcal{S}em^{cids}$) satisfy directionality*

*(d) if $\mathcal{S}em$ gives conflict-free extensions, then the ideal semantics based on $\mathcal{S}em$ ($\mathcal{S}em^{id}$) satisfies directionality*

*Proof.* Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $S \in \mathcal{US}(F)$ be an unattacked set.

(a) We have $\mathcal{AE}_{\mathcal{S}em^S}(F, S) = \{S \cap \bigcap_{E \in \mathcal{E}_{\mathcal{S}em}(F)} E\} = \{\bigcap_{E \in \mathcal{E}_{\mathcal{S}em}(F)}(S \cap E)\} = \{\bigcap_{E \in \mathcal{AE}_{\mathcal{S}em}(F,S)} E\} =^* \{\bigcap_{E \in \mathcal{E}_{\mathcal{S}em}(F\downarrow_S)} E\} = \mathcal{E}_{\mathcal{S}em^S}(F\downarrow_S)$. The starred equality comes from the fact that $\mathcal{S}em$ satisfies directionality.

(b) We will prove $\mathcal{AE}_{\mathcal{S}em^{ids}}(F, S) = \mathcal{E}_{\mathcal{S}em^{ids}}(F\downarrow_S)$ by double inclusion. First, we show that for any $E \in \mathcal{E}_{\mathcal{S}em^{ids}}(F)$ we have $E \cap S \in \mathcal{E}_{\mathcal{S}em^{ids}}(F\downarrow_S)$. As an ideal set, $E$ is admissible in $F$, so $E \cap S \in \mathcal{E}_{\mathcal{AS}}(F\downarrow_S)$ (because $\mathcal{AS}$ satisfies directionality). Furthermore, $E$ is included in all $\mathcal{S}em$ extensions of $F$, i.e. $E \subseteq \bigcap_{E' \in \mathcal{E}_{\mathcal{S}em}(F)} E'$. But we have seen at (a) that $S \cap \bigcap_{E' \in \mathcal{E}_{\mathcal{S}em}(F)} E' = \bigcap_{E' \in \mathcal{E}_{\mathcal{S}em}(F\downarrow_S)} E'$, so we have that $E \cap S$ is included in all $\mathcal{S}em$ extensions of $F\downarrow_S$, leading to $\mathcal{AE}_{\mathcal{S}em^{ids}}(F, S) \subseteq \mathcal{E}_{\mathcal{S}em^{ids}}(F\downarrow_S)$.

For the other inclusion, consider the ideal set $E' \in \mathcal{E}_{\mathcal{S}em^{ids}}(F\downarrow_S)$. As an ideal set, $E'$ is admissible in $F\downarrow_S$ and, since $S$ is unattacked, also admissible in $F$. Now, for every $\mathcal{S}em$ extension $E$ of $F$ we have $E \cap S \in \mathcal{E}_{\mathcal{S}em}(F\downarrow_S)$, because $\mathcal{S}em$ satisfies directionality. Since $E'$ is included in all $\mathcal{S}em$ extensions of $F\downarrow_S$, it follows that it is included in all $\mathcal{S}em$ extensions of $F$ as well, so $E' \in \mathcal{E}_{\mathcal{S}em^{ids}}(F)$. But $E' \cap S = E'$, which leads to $E' \in \mathcal{AE}_{\mathcal{S}em^{ids}}(F, S)$ and the desired inclusion.

(c) We need to show that $\mathcal{AE}_{\mathcal{S}em^{cids}}(F, S) = \mathcal{E}_{\mathcal{S}em^{cids}}(F\downarrow_S)$. The proof for $\mathcal{AE}_{\mathcal{S}em^{cids}}(F, S) \subseteq \mathcal{E}_{\mathcal{S}em^{cids}}(F\downarrow_S)$ is very similar to that provided for the case of ideal sets and relies on the directionality of both $\mathcal{S}em$ and $\mathcal{CO}$.

For the other inclusion note that, given $E' \in \mathcal{E}_{\mathcal{S}em^{cids}}(F\downarrow_S)$ it may be the case that $E'$ is not complete in $F$. Still, from the directionality of $\mathcal{CO}$ we have that there exists $E'' \in \mathcal{E}_{\mathcal{CO}}(F)$ such that $E'' \cap S = E'$. Furthermore, we use the fact that $E'$ is admissible in $F$ (because it is admissible in $F\downarrow_S$ and $S$ is unattacked) and we use $E$ to denote the minimal complete extension of $F$ that contains $E'$. Note that $E$ exists and is unique, according to a result in (Caminada and Pigozzi, 2011). Since $E$ is minimal, we have $E \subseteq E'' \Rightarrow E \cap S \subseteq E'' \cap S = E'$ and since $E' \subseteq E$ we have $E \cap S = E'$. We only need to show that $E$ is a complete ideal set. Let $D$ be an arbitrary $\mathcal{S}em$ extension of $F$. Since $\mathcal{S}em$ satisfies directionality, we have $D \cap S \in \mathcal{E}_{\mathcal{S}em}(F\downarrow_S)$. But then $E' \subseteq D$ so, as $D$ is complete, $E \subseteq D$. Thus, $E$ is indeed a complete ideal set.

(d) Last, but not least, we consider the case of the ideal $\mathcal{S}em$ extension. We assume that $\mathcal{S}em$ is conflict-free, so the ideal semantics is unique-status (Dvorak et al., 2011). This means that we must show that $E \cap S = E'$, where $E \in \mathcal{E}_{\mathcal{S}em^{id}}(F)$ is the ideal extension of the whole framework and $E' \in \mathcal{E}_{\mathcal{S}em^{id}}(F\downarrow_S)$ is the ideal extension of the restricted framework. First, since $E'$ is also admissible in $F$ (because $S$ is unattacked) and included in all $\mathcal{S}em$ extensions of $F$ (as we have seen in the proof for ideal sets), we conclude that $E'$ is an ideal set in $F$ as well, so we must have $E' \subseteq E$, which leads to $E' \subseteq E \cap S$. On the other hand we have that $E \cap S$ is an ideal set of the restricted framework, so it must be that $E \cap S \subseteq E'$, as $E'$ should be maximal. Thus, we proved that $E \cap S = E'$. □

Now let us see what semantics are covered by Theorem 2. We have already seen that $\mathcal{SST}$, $\mathcal{STA}$ and $\mathcal{MCF}$ do not generate any semantics satisfying directionality, which is in agreement with our result, as neither of them satisfies it either. On the other hand, both $\mathcal{PR}$ and $\mathcal{GR}^*$ satisfy directionality, so we get that $\mathcal{PR}^S$, $\mathcal{IDS}$, $\mathcal{PR}^{cids}$, $\mathcal{ID}$, $\mathcal{GR}^{*S}$, $\mathcal{GR}^{*ids}$, $\mathcal{GR}^{*cids}$ and $\mathcal{GR}^{*id}$ satisfy directionality.

Note that the resolution-based version of $\mathcal{S}em$ is not included in Theorem 2. This is consistent with the fact that we already have examples of argumentation semantics ($\mathcal{PR}$ and $\mathcal{ID}$) whose resolution-based versions fail to satisfy directionality even though the corresponding base semantics do satisfy it.

A part of the prudent semantics are covered in (Baroni and Giacomin, 2007), where it is shown that $\mathcal{GR}^P$ satisfies directionality, while $\mathcal{CO}^P$, $\mathcal{ST}^P$ and $\mathcal{PR}^P$ do not. We show in Proposition 8 that the other two ($\mathcal{CF}^P$ and $\mathcal{AS}^P$) satisfy directionality.

**Proposition 8.** *$\mathcal{CF}^P$ and $\mathcal{AS}^P$ satisfy the directionality principle.*

*Proof.* Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $S \in \mathcal{US}(F)$ be an unattacked set. Since the proofs are very similar, we only consider the case of $\mathcal{AS}^P$. For any $p$-admissible set $E$ we have that $E \cap S$ is $p$-admissible in $S$, because $\mathcal{AS}$ satisfies directionality and the prudent conflict-freeness cannot be violated by restricting the framework. Conversely, given $E' \in \mathcal{AS}^P(F{\downarrow_S})$, we have that $E'$ is also $p$-admissible in $F$, because $S$ is unattacked. □

So the only remaining semantics to consider is $\mathcal{EPS}$. But since $\mathcal{EPS}$ does not satisfy reinstatement, we can directly conclude that it does not satisfy directionality either.

### 3.1.5 Round-up

To sum up, we have presented in this section several principles for evaluating argumentation semantics. While some of them are satisfied by the vast majority of argumentation semantics (conflict-freeness, non-interference, additivity, universally defined), others are not. We have also provided significant contributions in filling the gaps in the literature for some of the principles and also by introducing the additivity principle. All results presented in this section are summarized in Table 3.3. We have underlined the results for which we have had a contribution. We have also underlined the argumentation semantics that are generally used only as helping sets for defining other semantics.

## 3.2 The Map of Argumentation Semantics

In Chapter 2 we have provided a rather comprehensive survey of the argumentation semantics that were proposed in the literature, together with other interesting sets that were just mentioned but not thoroughly studied. This section adds an original contribution to this survey by formalizing the graphical representation of the relations between semantics, which will be referred to as a map of argumentation semantics from now on. Furthermore, we distinguish between inclusion and inner inclusion relations and represent them both on the same map.

Based on the principles discussed in Section 3.1 we will identify five new argumentation semantics that will help us represent the satisfaction of these principles on our map. The cardinality properties are shown on the map via the borders of the corresponding nodes. We also introduce the concept of validation through

| Semantics | Card. | CF / adm. | $\mathcal{CF}$/weak/strong reinstatement | Non-interf./ add./dir. |
|---|---|---|---|---|
| $\mathcal{CF}$ | $\geq 1$ | Yes / No | No / No / No | Yes / Yes / Yes |
| $\mathcal{AS}$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / Yes |
| $\mathcal{CO}$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{ST}$ | $\geq 0$ | Yes / Yes | Yes / Yes / Yes | No / Yes / No |
| $\mathcal{GR}$ | $= 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{PR}$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{SST}$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{STA}$ | $\geq 1$ | Yes / No | Yes / No / No | Yes / Yes / No |
| $\mathcal{MCF}$ | $\geq 1$ | Yes / No | Yes / No / No | Yes / Yes / No |
| $\mathcal{GR}^*$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{PR}^*$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{ID}^*$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{SST}^*$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{PR}^S$ | $= 1$ | Yes / No | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{SST}^S$ | $= 1$ | Yes / No | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{STA}^S$ | $= 1$ | Yes / No | No / No / No | Yes / Yes / No |
| $\mathcal{MCF}^S$ | $= 1$ | Yes / No | No / No / No | Yes / Yes / No |
| $\mathcal{GR}^{*S}$ | $= 1$ | Yes / No | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{IDS}$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / Yes |
| $\mathcal{EAGS}$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{STA}^{ids}$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{MCF}^{ids}$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{GR}^{*ids}$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / Yes |
| $\mathcal{PR}^{cids}$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{SST}^{cids}$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{GR}^{*cids}$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{ID}$ | $= 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{EAG}$ | $= 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / No |
| $\mathcal{STA}^{id}$ | $= 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{MCF}^{id}$ | $= 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{GR}^{*id}$ | $= 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{AD}1$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{AD}2$ | $\geq 1$ | Yes / Yes | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{CF}1$ | $\geq 1$ | Yes / No | Yes / Yes / No | Yes / Yes / Yes |
| $\mathcal{CF}2$ | $\geq 1$ | Yes / No | Yes / Yes / No | Yes / Yes / Yes |
| $\mathcal{STA}2$ | $\geq 1$ | Yes / No | Yes / Yes / No | Yes / Yes / Yes |
| $\mathcal{CF}^P$ | $\geq 1$ | Yes / No | No / No / No | Yes / Yes / Yes |
| $\mathcal{AS}^P$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / Yes |
| $\mathcal{CO}^P$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{ST}^P$ | $\geq 0$ | Yes / Yes | Yes / Yes / Yes | No / Yes / No |
| $\mathcal{GR}^P$ | $= 1$ | Yes / Yes | No / No / No | Yes / Yes / Yes |
| $\mathcal{PR}^P$ | $\geq 1$ | Yes / Yes | No / No / No | Yes / Yes / No |
| $\mathcal{EPS}$ | $\geq 1$ | Yes / No | Yes / Yes / No | No / No / No |

**Table 3.3:** *Principles for evaluating semantics. Our own results are underlined.*

examples for the relations between semantics. Roughly speaking, validation consists in providing argumentation frameworks that exhibit the differences between semantics.

### 3.2.1 Graphical representation

In this subsection we introduce a graphical representation for the relations between argumentation semantics. The aim is to provide a map of argumentation semantics that is as informative as possible without cluttering it too much.

**Definition 41.** *Let $\mathcal{S}em_1$ and $\mathcal{S}em_2$ be two argumentation semantics. We say that $\mathcal{S}em_1$ is* ***included*** *in $\mathcal{S}em_2$ (notation $\mathcal{S}em_1 \subseteq \mathcal{S}em_2$) iff, for any argumentation framework $F$, $\mathcal{E}_{\mathcal{S}em_1}(F) \subseteq \mathcal{E}_{\mathcal{S}em_2}(F)$. In this case $\mathcal{S}em_1$ is said to be a* ***sub-semantics*** *of $\mathcal{S}em_2$, while $\mathcal{S}em_2$ is said to be a* ***super-semantics*** *of $\mathcal{S}em_1$. We implicitly assume that the semantics are distinct, i.e. there is at least one argumentation framework for which the inclusion is strict.*

*We say that there is an* ***inner inclusion*** *relation between $\mathcal{S}em_1$ and $\mathcal{S}em_2$ iff, for any argumentation framework $F$ and any extensions $S_1 \in \mathcal{E}_{\mathcal{S}em_1}(F)$, $S_2 \in \mathcal{E}_{\mathcal{S}em_2}(F)$ we have $S_1 \subseteq S_2$. We say that the inner inclusion is* ***tight*** *(and we write $\cup\mathcal{S}em_1 = \cap\mathcal{S}em_2$) iff, for any argumentation framework $F$, the union of the extensions given by $\mathcal{S}em_1$ is equal to the intersection of the extensions given by $\mathcal{S}em_2$. Otherwise, we say that the inner inclusion is* ***loose*** *(notation $\cup\mathcal{S}em_1 \subsetneq \cap\mathcal{S}em_2$).*

$$
\begin{aligned}
\mathcal{S}em_1 \subseteq \mathcal{S}em_2 &\Leftrightarrow \forall F(\mathcal{E}_{\mathcal{S}em_1}(F) \subseteq \mathcal{E}_{\mathcal{S}em_2}(F)) \textit{ and} \\
&\quad\ \exists F(\mathcal{E}_{\mathcal{S}em_1}(F) \subsetneq \mathcal{E}_{\mathcal{S}em_2}(F)) \\
\cup\mathcal{S}em_1 = \cap\mathcal{S}em_2 &\Leftrightarrow \forall F(\bigcup_{S\in\mathcal{E}_{\mathcal{S}em_1}(F)} S = \bigcap_{S\in\mathcal{E}_{\mathcal{S}em_2}(F)} S) \\
\cup\mathcal{S}em_1 \subseteq \cap\mathcal{S}em_2 &\Leftrightarrow \forall F(\bigcup_{S\in\mathcal{E}_{\mathcal{S}em_1}(F)} S \subseteq \bigcap_{S\in\mathcal{E}_{\mathcal{S}em_2}(F)} S) \textit{ and} \\
&\quad\ \exists F(\bigcup_{S\in\mathcal{E}_{\mathcal{S}em_1}(F)} S \subsetneq \bigcap_{S\in\mathcal{E}_{\mathcal{S}em_2}(F)} S)
\end{aligned}
\tag{3.8}
$$

Graphical representations have often been provided in the argumentation literature (Caminada, 2007; Baroni et al., 2011a; Dvorak and Gaggl, 2012) for various kinds of relations between argumentation semantics, generally providing a hierarchical view based on an undirected graph and showing a single type of relation at a time. Our aim here is to show both inclusion and inner inclusion on the same representation. We use a directed graph with several types of edges and nodes.

We use the abbreviated names of argumentation semantics as names for the nodes in the graph. The border of the node describes the cardinality properties of the corresponding semantics, as follows:

- multiple-status, not universally defined semantics will be surrounded by a dashed single line border
- unique-status, universally defined semantics will be surrounded by a solid double line border
- multiple-status, universally defined semantics will be surrounded by a solid single line border

Furthermore, we indicate the satisfaction of the non-interference and directionality principles via the background of the corresponding nodes, as follows:

- dark gray background for semantics that satisfy both directionality and non-interference
- light gray background for semantics that only satisfy non-interference
- white background for semantics that do not satisfy any of the two principles

Note that, since directionality implies non-interference, it is impossible to have argumentation semantics satisfying directionality but not non-interfence.

We have not included additivity in this convention, as it would lead to four distinct classes of semantics and thus the need for three shades of gray, which might make it too difficult to read the map. Instead, we shall use the border color for encoding the satisfaction of additivity, with gray for $\mathcal{EPS}$ (the only semantics that fails to satisfy this principle) and black for all the other semantics.

The relations between argumentation semantics will be represented by arrows connecting the corresponding nodes, as follows:

- tight inner inclusion between $\mathcal{S}em_1$ and $\mathcal{S}em_2$ – solid, double line arrow from $\mathcal{S}em_1$ to $\mathcal{S}em_2$
- loose inner inclusion between $\mathcal{S}em_1$ and $\mathcal{S}em_2$ – dashed, double line arrow from $\mathcal{S}em_1$ to $\mathcal{S}em_2$
- inclusion between $\mathcal{S}em_1$ and $\mathcal{S}em_2$ – single line arrow from $\mathcal{S}em_1$ and $\mathcal{S}em_2$. Here, we also consider a refinement: whenever $\mathcal{S}em_1$ has several super-semantics, it may happen that $\mathcal{S}em_1$ is actually equal to the intersection of its (closest) super-semantics, in which case all the arrows going away from $\mathcal{S}em_1$ will have solid lines. Otherwise, we will use dashed lines.

Two example maps are provided in Figure 3.7 (a) and (b). The second map is only provided for showing an argumentation semantics that is equal to the intersection of its immediate super-semantics ($\mathcal{PR}^{cids} = \mathcal{IDS} \cap \mathcal{CO}$). We will focus on the first map and see what information it encodes.

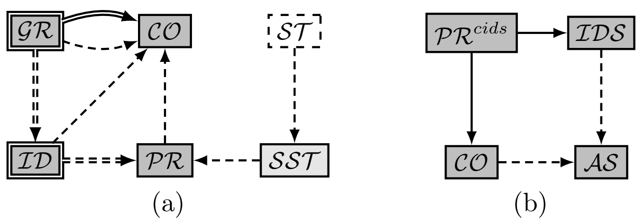



**Figure 3.7:** *Example maps for argumentation semantics.*

First, from the borders of the nodes, we see that the stable semantics is not universally defined, while the grounded and the ideal semantics are unique-status. The other semantics are multiple-status and universally defined. Furthermore, the background of the nodes encodes the fact that $\mathcal{SST}$ satisfies non-interference, but not directionality, $\mathcal{ST}$ satisfies neither of them, while the other semantics satisfy both principles. All six semantics are additive.

The single line arrows encode known inclusion relations: all stable extensions are also semi-stable, all semi-stable extensions are preferred, and so on. Note that, via the transitivity of set inclusion, we can also deduce that $\mathcal{SST} \subseteq \mathcal{CO}$ although there is no arrow between the two semantics. In fact, we will try to put as little explicit information as possible, i.e. avoid redundant arrows in the map. What this means is that for each argumentation semantics we will only provide links from its maximal sub-semantics to it and from it toward its minimal super-semantics.

The inner inclusion relations (both the tight and the loose one) are also transitive, so we can rely on this and avoid some redundant information. On the other hand, let us see that whenever $\cup\mathcal{S}em_1 \subseteq \cap\mathcal{S}em_2$, we have that for every $\mathcal{S}em'_1 \subseteq \mathcal{S}em_1$ and $\mathcal{S}em'_2 \subseteq \mathcal{S}em_2$ the same relation holds: $\cup\mathcal{S}em'_1 \subseteq \cap\mathcal{S}em'_2$ (taking out a few sets can only enlarge the union and reduce the intersection). Thus, when placing loose inner inclusions on the map, we will try to find the pairs of maximal semantics for which they hold.

Furthermore, note that whenever tight inner inclusion holds between two semantics $\cup\mathcal{S}em_1 = \cap\mathcal{S}em_2$, it may also hold for sub-semantics of $\mathcal{S}em_1$ or $\mathcal{S}em_2$ or it may become loose inner inclusion. We will specifically show on the map the pairs of minimal sub-semantics $\mathcal{S}em'_1 \subseteq \mathcal{S}em_1$, $\mathcal{S}em'_2 \subseteq \mathcal{S}em_2$ for which this happens, with the implicit assumption that for semantics that lie in between the inner inclusion remains tight.

In the map from Figure 3.7 (a) we have a tight inner inclusion between the grounded and the complete semantics, which means that we have (at least) loose inclusion between $\mathcal{GR}$ and all sub-semantics of $\mathcal{CO}$. For both of the minimal sub-semantics of $\mathcal{CO}$ shown on the map ($\mathcal{ID}$ and $\mathcal{PR}$), the relation does become loose, so we should point this out. We don't need to put an actual arrow between $\mathcal{GR}$ and $\mathcal{PR}$ because we can already deduce the relation from the transitivity of inner inclusion.

We would also like to show that a given map of argumentation semantics is complete, in the sense that all inclusion relations that are valid can be deduced from it. To this end, we introduce the validation of a map.

**Definition 42.** *We say that an argumentation framework $F$ **validates** a set theoretical formula $f$ containing semantics (notation $F : f$) iff that formula is true when replacing every semantics $\mathcal{S}em$ with $\mathcal{E}_{\mathcal{S}em}(F)$. For example, we will use formulas like $\mathcal{S}em_1 \setminus \mathcal{S}em_2 \neq \varnothing$ and $\mathcal{S}em \setminus (\mathcal{S}em_1 \cup \mathcal{S}em_2) \neq \varnothing$. We say that a family (set) of argumentation frameworks $\mathcal{V}$ **validates** a formula $f$ (notation $\mathcal{V} : f$) iff there is an argumentation framework $F \in \mathcal{V}$ that validates $f$.*

*A family of argumentation frameworks $\mathcal{V}$ **validates** a map of argumentation semantics $\mathcal{M}$ iff $\mathcal{V} : \mathcal{S}em_1 \setminus \mathcal{S}em_2 \neq \varnothing$ for any $\mathcal{S}em_1$ and $\mathcal{S}em_2$ such that there is no path between $\mathcal{S}em_1$ and $\mathcal{S}em_2$ in $\mathcal{M}$ along the single line arrows (dashed or not).*

*We say that $\mathcal{V}$ **strongly validates** $\mathcal{M}$ iff, for every argumentation semantics $\mathcal{S}em$ that is shown in $\mathcal{M}$, the following holds: $\mathcal{V} : \mathcal{S}em \setminus \bigcup \textit{non-ancestors}(\mathcal{S}em) \neq \varnothing$, where non-ancestors$(\mathcal{S}em)$ denotes the argumentation semantics $S$ from $\mathcal{M}$ such that there is not path from $\mathcal{S}em'$ to $\mathcal{S}em$ along the single line arrows of $\mathcal{M}$.*

*Given a map of argumentation semantics and an argumentation framework $F$, we use $\delta_F(\mathcal{S}em)$ to denote the extensions that **distinguish** $\mathcal{S}em$ from all non-ancestor semantics in $\mathcal{M}$:*

$$\delta_F(\mathcal{S}em) = \mathcal{E}_{\mathcal{S}em}(F) \setminus \bigcup_{\mathcal{S}em' \in \textit{non-ancestors}(\mathcal{S}em)} \mathcal{E}_{\mathcal{S}em'}(F) \tag{3.9}$$

In other words, the simple validation of a map consists of a family of argumentation frameworks that shows that the inclusion relations that cannot be deduced from the map do not hold in general. Alternatively, if the inclusion $\mathcal{S}em_1 \subseteq \mathcal{S}em_2$ can be inferred from the map, but $\mathcal{S}em_2 \subseteq \mathcal{S}em_1$ cannot be inferred, we can see the validation of $\mathcal{S}em_2 \setminus \mathcal{S}em_1 \neq \varnothing$ as a proof that the inclusion between $\mathcal{S}em_1$ and $\mathcal{S}em_2$ can be strict (the semantics are distinct).

Strong validation, on the other hand, requires that argumentation frameworks are provided for showing that each argumentation semantics is unique (distinct from all semantics that are not its super-semantics).

We will also try to use a minimal number of distinct frameworks for the validation of our maps. Strong validation will be preferred and most frameworks will also prove additional properties, such as the fact that a loose inner inclusion relation can indeed be strict. A detailed validation example is provided in the next subsection.

### 3.2.2 Principle-based semantics

While (Baroni and Giacomin, 2008) uses a principle-based approach that relies on (Baroni and Giacomin, 2007), the principles that are used there are among the more advanced ones and, to the best of our knowledge, there is no work in the argumentation literature that systematically defines argumentation semantics based on the much simpler principles of admissibility and reinstatement.

Aside from the fact that we find it interesting in itself to see how many distinct semantics can be defined solely based on the admissibility and reinstatement principles, and how they compare to existing semantics, there is also a more practical reason to it. We can use these semantics in order to show which of these principles are satisfied by each argumentation semantics. All we have to do is put these principle-based semantics on the map as well and then draw the relations between them and the other semantics.

First of all, we assume that the conflict-free principle ($cf$) is a given; all the semantics we introduce will satisfy it. Then we have the admissibility principle (denoted by $as$ in what follows) and the three kinds of reinstatement: strong ($sr$), weak ($wr$) and $\mathcal{CF}$ ($cr$). We can enforce any of these four principles on plain conflict-free sets. What we actually need to do is see how many of the 16 combinations lead to distinct semantics and how many of these semantics are already known in the argumentation literature.

First, we know that strong reinstatement implies both weak and $\mathcal{CF}$ reinstatement. So the actual distinct combinations that we can have for the reinstatement principles are: $\varnothing$, $\{wr\}$, $\{cr\}$, $\{wr, cr\}$ and $\{wr, cr, sr\}$.

Now, let us see that for admissible sets there are even fewer possible combinations for the reinstatement principles, as admissible sets that satisfy $\mathcal{CF}$-reinstatement also satisfy strong reinstatement. Indeed, suppose the set $S$ is admissible and satisfies $\mathcal{CF}$-reinstatement. Let $a$ be an argument that is defended by $S$. If $S$ attacks $a$, in order to satisfy the defense, it would have to attack one of its own arguments, which is impossible since $S$ is conflict-free. If $a$ attacks $S$, then $S$ should attack $a$ in order to defend its elements, again a contradiction. Thus, $a \not\rightarrow S$ and $S \not\rightarrow a$, the additional conditions for $\mathcal{CF}$-reinstatement, so $a$ is in $S$. Thus we have proved strong reinstatement. So, when adding reinstatement principles to admissible sets, we can rule out combinations that contain $\mathcal{CF}$-reinstatement but not strong reinstatement.

Thus, we have at most the combinations of principles covered in Definition 43.

**Definition 43.** *The use of the admissibility and reinstatement principles (weak, $\mathcal{CF}$ and strong) for defining argumentation semantics leads to the following distinct combinations:*

- $cf = \mathcal{CF}$ – *the conflict-free sets with no additional constraints*
- $cf + cr \triangleq \mathcal{CF}^{cr}$ – *conflict-free sets that satisfy $\mathcal{CF}$-reinstatement*
- $cf + wr \triangleq \mathcal{CF}^{wr}$ – *conflict-free sets that satisfy weak-reinstatement*
- $cf + wr + cr \triangleq \mathcal{CF}^{cw}$ – *conflict-free sets that satisfy $\mathcal{CF}$ and weak reinstatement*
- $cf + wr + cr + sr \triangleq \mathcal{CF}^{sr}$ – *conflict-free sets that satisfy strong reinstatement*
- $as = \mathcal{AS}$ – *admissible sets with no additional constraints*
- $as + wr \triangleq \mathcal{AS}^{wr}$ – *admissible sets that satisfy weak reinstatement*
- $as + wr + cr + sr = \mathcal{CO}$ – *admissible sets that satisfy strong reinstatement (already known in the literature as complete extensions)*

*For the five novel semantics ($\mathcal{CF}^{cr}$, $\mathcal{CF}^{wr}$, $\mathcal{CF}^{cw}$, $\mathcal{CF}^{sr}$ and $\mathcal{AS}^{wr}$) we will use the notations introduced here.*

So, from the eight distinct combinations of principles that we have identified, three have lead us to known argumentation semantics: $\mathcal{CF}$, $\mathcal{AS}$ and $\mathcal{CO}$. The other five have not been previously considered in the literature.

These semantics are not independent from one another. The relations between them come from the principles that each of them satisfies. First, let us see that all semantics provide conflict-free sets, by construction. However, as promised, we aim to find the maximal sub-semantics for each of them so that we give a minimal representation of the relations between them.

Whenever the set of principles satisfied by $\mathcal{S}em_1$ includes the set of principles satisfied by $\mathcal{S}em_2$ we will have $\mathcal{S}em_1 \subseteq \mathcal{S}em_2$. It is just like applying additional constraints on the sets provided by $\mathcal{S}em_2$. Using this observation we get the following relations between the principle-based semantics:

$$\begin{aligned}
&\mathcal{CO} \subseteq \mathcal{CF}^{sr} \subseteq \mathcal{CF}^{cw} \subseteq \mathcal{CF}^{cr} \subseteq \mathcal{CF} \\
&\mathcal{CO} \subseteq \mathcal{AS}^{wr} \subseteq \mathcal{AS} \subseteq \mathcal{CF} \\
&\mathcal{CF}^{cw} \subseteq \mathcal{CF}^{wr} \subseteq \mathcal{CF} \\
&\mathcal{AS}^{wr} \subseteq \mathcal{CF}^{wr}
\end{aligned} \tag{3.10}$$

From these relations we see for example that we have $\mathcal{CO} \subseteq \mathcal{AS}^{wr}$ and $\mathcal{CO} \subseteq \mathcal{CF}^{sr}$, so we also have $\mathcal{CO} \subseteq \mathcal{AS}^{wr} \cap \mathcal{CF}^{sr}$. The interesting question is whether this inclusion can indeed be strict or it is in fact equality. It is not difficult to see that satisfying principles from both $\mathcal{AS}^{wr}$ and $\mathcal{CF}^{sr}$ leads to complete extensions, so we have $\mathcal{CO} = \mathcal{AS}^{wr} \cap \mathcal{CF}^{sr}$. Using similar observations, the complete list of such relations is:

$$\begin{aligned}
&\mathcal{CO} = \mathcal{AS}^{wr} \cap \mathcal{CF}^{sr} \\
&\mathcal{CF}^{cw} = \mathcal{CF}^{cr} \cap \mathcal{CF}^{wr} \\
&\mathcal{AS}^{wr} = \mathcal{CF}^{wr} \cap \mathcal{AS}
\end{aligned} \tag{3.11}$$

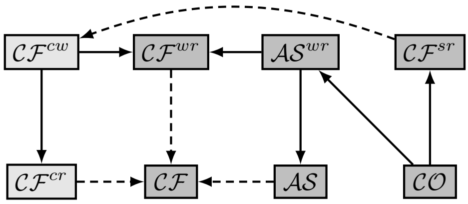



**Figure 3.8:** *The map of principle-based semantics.*

All the relations from (3.10) and (3.11) are represented in the map from Figure 3.8. Each of these semantics is clearly multiple-status and also universally defined, since the most restrictive of them, the complete semantics, satisfies these properties. The satisfaction of the conflict-free, admissibility and reinstatement principles comes from the definitions of the semantics and is encoded in the relations between them. In what follows, we discuss additivity, non-interference and directionality for the five new semantics: $\mathcal{CF}^{cr}$, $\mathcal{CF}^{wr}$, $\mathcal{CF}^{cw}$, $\mathcal{CF}^{sr}$ and $\mathcal{AS}^{wr}$ (the corresponding results for $\mathcal{CF}$, $\mathcal{AS}$ and $\mathcal{CO}$ were presented in Table 3.3).

**Proposition 9.** *The argumentation semantics $\mathcal{CF}^{wr}$, $\mathcal{CF}^{cr}$, $\mathcal{CF}^{cw}$, $\mathcal{CF}^{sr}$ and $\mathcal{AS}^{wr}$ are additive.*

*Proof.* Let $F_1 = (\mathcal{A}_1, \mathcal{R}_1)$ and $F_2 = (\mathcal{A}_2, \mathcal{R}_2)$ be two disjoint argumentation frameworks and let $F = F_1 \uplus F_2$ be their disjoint union. We consider a set of arguments $E \in \mathcal{A}_1 \cup \mathcal{A}_2$.

We start with the $\mathcal{CF}^{wr}$ semantics. Let $GR_F$, $GR_{F_1}$ and $GR_{F_2}$ be the grounded extensions of $F$, $F_1$ and $F_2$, respectively. Since the grounded semantics satisfies additivity, we have $GR_F \cap \mathcal{A}_1 = GR_{F_1}$ and $GR_F \cap \mathcal{A}_2 = GR_{F_2}$. We can use this in the definition of the $\mathcal{CF}^{wr}$ semantics and get $E \in \mathcal{E}_{\mathcal{CF}^{wr}}(F) \Leftrightarrow E \in \mathcal{E}_{\mathcal{CF}}(F)$ and $GR_F \subseteq E \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{CF}}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{CF}}(F_2)$ and $GR_{F_1} \subseteq E \cap \mathcal{A}_1$ and $GR_{F_2} \subseteq E \cap \mathcal{A}_2 \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{CF}^{wr}}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{CF}^{wr}}(F_2)$, which is the desired result. Note that we have also used the fact that $\mathcal{CF}$ is additive.

For $\mathcal{CF}^{cr}$ we need to prove that $E \in \mathcal{E}_{\mathcal{CF}^{cr}}(F) \Leftrightarrow E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{CF}^{cr}}(F_1)$ and $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{CF}^{cr}}(F_2)$. For the direct implication, suppose $E \in \mathcal{E}_{\mathcal{CF}^{cr}}(F)$. Let us consider an argument $a \in \mathcal{A}_1$ such that $(E \cap \mathcal{A}_1) \cup \{a\}$ is conflict free and $E \cap \mathcal{A}_1$ defends $a$, in $F_1$. It follows that $E \cup \{a\}$ is conflict-free in $F$ ($a$ is not in conflict with any argument from $F_2$) and also $E$ defends $a$ in $F$ ($F_2$ contains no attackers of $a$, so the defense is preserved). Then, from the definition of $\mathcal{CF}^{cr}$, it must be that $a \in E$, which leads to $a \in E \cap \mathcal{A}_1$ and, thus, $E \cap \mathcal{A}_1 \in \mathcal{E}_{\mathcal{CF}^{cr}}(F_1)$. In the same way we have that $E \cap \mathcal{A}_2 \in \mathcal{E}_{\mathcal{CF}^{cr}}(F_2)$. The proof of the converse is similar.

The proof for $\mathcal{CF}^{sr}$ follows the same approach as above, only without requiring that $E \cup \{a\}$ is conflict-free.

For $\mathcal{CF}^{cw}$ and $\mathcal{AS}^{wr}$ we use the fact that the intersection of two additive semantics is also additive (Theorem 1 (g)), as $\mathcal{CF}^{cw} = \mathcal{CF}^{cr} \cap \mathcal{CF}^{wr}$ and $\mathcal{AS}^{wr} = \mathcal{AS} \cap \mathcal{CF}^{wr}$. □

Since all the new principle-based semantics are universally defined and additive, we can use Proposition 4 to infer that they satisfy non-interference as well. Next, we discuss the directionality principle.

**Proposition 10.** *The argumentation semantics $\mathcal{CF}^{wr}$, $\mathcal{AS}^{wr}$ and $\mathcal{CF}^{sr}$ satisfy the directionality principle.*

*Proof.* Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $S \in \mathcal{US}(F)$ be an unattacked set of $F$.

We start with $\mathcal{CF}^{wr}$. We use $GR_F$ for the grounded extension of $F$ and $GR_{F\downarrow_S}$ for the grounded extension of the restricted framework. Since $\mathcal{GR}$ satisfies directionality, we have that $GR_F \cap S = GR_{F\downarrow_S}$. We use this observation and we have $\mathcal{AE}_{\mathcal{CF}^{wr}}(F, S) = \{E \cap S \mid E \in \mathcal{E}_{\mathcal{CF}}(F) \text{ and } GR_F \subseteq E\} = \{E' \in \mathcal{AE}_{\mathcal{CF}}(F, S) \mid GR_F \cap S \subseteq E'\} = \{E' \in \mathcal{E}_{\mathcal{CF}}(F\downarrow_S) \mid GR_{F\downarrow_S} \subseteq E'\} = \mathcal{E}_{\mathcal{CF}^{wr}}(F\downarrow_S)$, which proves the claim. We have also used the fact that $\mathcal{CF}$ satisfies directionality.

The proof for $\mathcal{AS}^{wr}$ goes along the same lines, using the fact that $\mathcal{AS}$ satisfies directionality.

For $\mathcal{CF}^{sr}$ we prove $\mathcal{AE}_{\mathcal{CF}^{sr}}(F, S) = \mathcal{E}_{\mathcal{CF}^{sr}}(F\downarrow_S)$ by double inclusion. First, suppose $E \in \mathcal{E}_{\mathcal{CF}^{sr}}(F)$. Then $E$ is conflict-free in $F$, which means that $E \cap S$ is conflict-free in $F\downarrow_S$, because $\mathcal{CF}$ satisfies directionality. Let $a \in S$ be an argument defended in $F\downarrow_S$ by $E \cap S$. Since $S$ is unattacked and $a \in S$, all the attackers of $a$ are in $S$, so $A$ is also defended by $E$ in $F$. But then, since $E$ is a $\mathcal{CF}^{sr}$ extension, we have that $a \in E$ and, thus, $a \in E \cap S$, which proves that $E \cap S \in \mathcal{E}_{\mathcal{CF}^{sr}}(F\downarrow_S)$ and, thus, $\mathcal{AE}_{\mathcal{CF}^{sr}}(F, S) \subseteq \mathcal{E}_{\mathcal{CF}^{sr}}(F\downarrow_S)$

For the other inclusion, suppose we have $E \in \mathcal{E}_{\mathcal{CF}^{sr}}(F\downarrow_S)$. Let $T \subseteq \mathcal{A} \setminus S$ be a maximal set (with respect to set inclusion) such that $E \cup T$ is conflict-free and $E \cup T$ defends all the arguments from $T$. First, let us see that such a set is always available. For this, note that $E \cup \varnothing$ defends all the arguments from $\varnothing$. Furthermore, we have $(E \cup T) \cap S = E$ and $E \cup T$ conflict-free from the choice of $T$. All that is left to prove is that $E \cup T$ satisfies the strong reinstatement property.

Suppose that there exists an argument $a \notin E \cup T$ such that $E \cup T$ defends $a$. Since $E \cup T$ is conflict-free and defends $a$, we have $E \cup T \not\to a$. We will also show that $a \not\to E \cup T$. First, $a \not\to T$ because, as $E \cup T$ defends all arguments from $T$, it would mean that $E \cup T \to a$ and we have already seen that this is not the case. All arguments from $S$ that are defended by $E \cup T$ are also defended by $E$ (because all their attackers are in $S$, so the defenders should be in $S$ as well) so they are already in $E$, because $E$ is a $\mathcal{CF}^{sr}$ extension of $F\downarrow_S$. So $a \notin S$. But then $a \not\to E$, as $E \subseteq S$ and $S$ is unattacked. So $E \cup T \cup \{a\}$ is a conflict-free set that defends all arguments from $T \cup \{a\}$. But this contradicts the maximality of $T$. So it must be that $E \cup T \in \mathcal{E}_{\mathcal{CF}^{sr}}(F)$, which concludes our proof. □

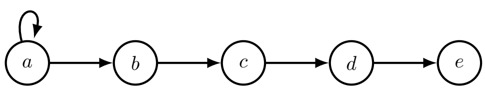



**Figure 3.9:** *$\mathcal{CF}^{cr}$ and $\mathcal{CF}^{cw}$ do not satisfy directionality.*

For $\mathcal{CF}^{cr}$ and $\mathcal{CF}^{cw}$, we provide the argumentation framework $F$ from Figure 3.9 as an example showing that these semantics do not satisfy directionality. The conflict-free sets of $F$ are $\mathcal{E}_{\mathcal{CF}}(F) = \{\varnothing, \{b\}, \{b, d\}, \{b, e\}, \{c\}, \{c, e\}, \{d\}, \{e\}\}$. Since the grounded extension is the empty set in this case, all conflict-free sets satisfy the weak reinstatement condition. Furthermore, $\{b\}$ defends argument $d$ and is not in conflict with it, thus violating $\mathcal{CF}$-reinstatement. Similarly, $\{c\}$ defends argument $e$ while not in conflict with it. All the other conflict-free sets satisfy $\mathcal{CF}$-reinstatement.

Thus, we have $\mathcal{E}_{\mathcal{CF}^{cr}}(F) = \mathcal{E}_{\mathcal{CF}^{cw}}(F) = \{\varnothing, \{b,d\}, \{b,e\}, \{c,e\}, \{d\}, \{e\}\}$. We now consider the unattacked set $S = \{a,b,c,d\}$. We have that $\mathcal{AE}_{\mathcal{CF}^{cr}}(F,S) = \mathcal{AE}_{\mathcal{CF}^{cw}}(F,S) = \{\varnothing, \{b,d\}, \{b\}, \{c\}, \{d\}\}$, but $\mathcal{E}_{\mathcal{CF}^{cr}}(F\downarrow_S) = \mathcal{E}_{\mathcal{CF}^{cw}}(F\downarrow_S) = \{\varnothing, \{b,d\}, \{c\}, \{d\}\}$, so directionality is not satisfied.

| Semantics | Card. | CF / adm. | $\mathcal{CF}$/weak/strong reinstatement | Non-interf./ add./dir. |
|---|---|---|---|---|
| $\mathcal{CF}^{cr}$ | $\geq 1$ | Yes / No | Yes / No / No | Yes / Yes / No |
| $\mathcal{CF}^{wr}$ | $\geq 1$ | Yes / No | No / Yes / No | Yes / Yes / Yes |
| $\mathcal{CF}^{cw}$ | $\geq 1$ | Yes / No | Yes / Yes / No | Yes / Yes / No |
| $\mathcal{CF}^{sr}$ | $\geq 1$ | Yes / No | Yes / Yes / Yes | Yes / Yes / Yes |
| $\mathcal{AS}^{wr}$ | $\geq 1$ | Yes / Yes | No / Yes / No | Yes / Yes / Yes |

**Table 3.4:** *Properties of principle-based semantics.*

We summarize the results with respect to principle satisfaction in Table 3.4, which is meant as a continuation of Table 3.3, leading to a total number of 48 discussed semantics. The role of the five novel semantics here is mostly representational, as they help us classify the argumentation semantics with respect to the satisfaction of the admissibility and reinstatement principles. On the other hand, we consider that they are at least as reasonable in practice as other semantics that are not based on admissibility, such as $\mathcal{CF}2$, $\mathcal{STA}$ or $\mathcal{STA}2$.

We will now show that the argumentation framework $F$ from Figure 3.10 strongly validates the map. First, let us see the extensions provided by each of semantics.

$$
\begin{aligned}
\mathcal{E}_{\mathcal{CF}}(F) =&\{\varnothing, \{a\}, \{a,d\}, \{a,d,e\}, \{a,d,f\}, \{a,e\}, \{a,f\}, \{b\},\\
&\{b,e\}, \{b,f\}, \{c\}, \{c,d\}, \{c,d,e\}, \{c,d,f\}, \{c,e\}, \{c,f\},\\
&\{d\}, \{d,e\}, \{d,f\}, \{e\}, \{f\}\}\\
\mathcal{E}_{\mathcal{CF}^{cr}}(F) =&\{\{a,d,e\}, \{a,d,f\}, \{b,e\}, \{b,f\}, \{c,d,e\}, \{c,d,f\}, \{d,e\},\\
&\{d,f\}, \{e\}, \{f\}\}\\
\mathcal{E}_{\mathcal{CF}^{wr}}(F) =&\{\{a,d,e\}, \{a,e\}, \{b,e\}, \{c,d,e\}, \{c,e\}, \{d,e\}, \{e\}\} \qquad (3.12)\\
\mathcal{E}_{\mathcal{CF}^{cw}}(F) =&\{\{a,d,e\}, \{b,e\}, \{c,d,e\}, \{d,e\}, \{e\}\}\\
\mathcal{E}_{\mathcal{CF}^{sr}}(F) =&\{\{a,d,e\}, \{d,e\}, \{e\}\}\\
\mathcal{E}_{\mathcal{AS}}(F) =&\{\varnothing, \{a\}, \{a,d\}, \{a,d,e\}, \{a,e\}, \{e\}\}\\
\mathcal{E}_{\mathcal{AS}^{wr}}(F) =&\{\{a,d,e\}, \{a,e\}, \{e\}\}\\
\mathcal{E}_{\mathcal{CO}}(F) =&\{\{a,d,e\}, \{e\}\}
\end{aligned}
$$

For strong validation, we should first be able to pick, for each semantics, at least one extension that distinguishes it from all the other semantics (except its ancestors, of course). Let us take $\mathcal{CF}^{cw}$ as an example. Its ancestors are $\mathcal{CF}^{cr}$, $\mathcal{CF}^{wr}$ and $\mathcal{CF}$, so we should find an extension that is in $\mathcal{CF}^{cw}$ but in none of $\mathcal{CF}^{sr}$, $\mathcal{AS}$, $\mathcal{AS}^{wr}$ or $\mathcal{CO}$. We have two choices for such an extension: $\{b,e\}$ and $\{c,d,e\}$.

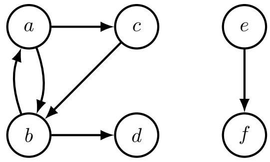



**Figure 3.10:** *Argumentation framework that strongly validates the map of principle-based argumentation semantics from Figure 3.8.*

Or, using the notation introduced in Definition 42, we have that $\delta_F(\mathcal{CF}^{cw}) = \{\{b,e\},\{c,d,e\}\}$. The distinguishing extensions for all principle-based semantics are:

$$\begin{aligned}
\delta_F(\mathcal{CF}) &= \{\{a,f\},\{b\},\{c\},\{c,d\},\{c,f\},\{d\}\} \\
\delta_F(\mathcal{CF}^{cr}) &= \{\{a,d,f\},\{b,f\},\{c,d,f\},\{d,f\},\{f\}\} \\
\delta_F(\mathcal{CF}^{wr}) &= \{\{c,e\}\} \\
\delta_F(\mathcal{CF}^{cw}) &= \{\{b,e\},\{c,d,e\}\} \\
\delta_F(\mathcal{CF}^{sr}) &= \{\{d,e\}\} \\
\delta_F(\mathcal{AS}) &= \{\varnothing,\{a\},\{a,d\}\} \\
\delta_F(\mathcal{AS}^{wr}) &= \{\{a,e\}\} \\
\delta_F(\mathcal{CO}) &= \{\{a,d,e\},\{e\}\}
\end{aligned} \tag{3.13}$$

As a side note, the argumentation framework that we have used for strong validation is minimal, in the sense that no argumentation framework with five arguments can strongly validate the map of principle-based argumentation semantics (we have tested this via a software implementation).

### 3.2.3 The classical semantics and $\mathcal{SST}$

In this subsection we consider and place on the map the remaining classical semantics ($\mathcal{GR}$, $\mathcal{PR}$ and $\mathcal{ST}$) and also the semi-stable semantics ($\mathcal{SST}$).

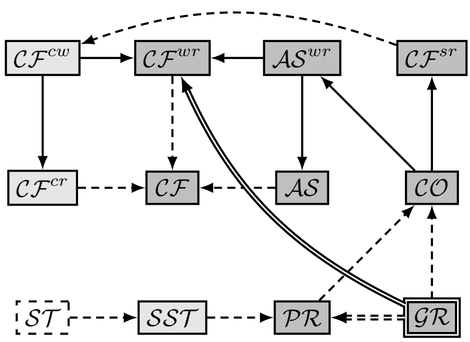



**Figure 3.11:** *Remaining classical semantics ($\mathcal{GR}$, $\mathcal{PR}$, $\mathcal{ST}$) and the semi-stable semantics $\mathcal{SST}$ added to the map.*

The grounded and preferred semantics provide complete extensions (Dung, 1995) so we have $\mathcal{GR} \subseteq \mathcal{CO}$ and $\mathcal{PR} \subseteq \mathcal{CO}$. Also, all stable extensions are semi-stable and all semi-stable extensions are preferred (Caminada, 2006b), so $\mathcal{ST} \subseteq \mathcal{SST} \subseteq \mathcal{PR}$.

Furthermore, the grounded extension is included in all complete extensions (Dung, 1995). But, in fact, the grounded extension is included in all conflict-free sets that satisfy weak reinstatement ($\mathcal{CF}^{wr}$). Since the grounded extension is itself a member of $\mathcal{CF}^{wr}$, the inner inclusion is tight. This inner inclusion propagates to all sub-semantics of $\mathcal{CF}^{wr}$ – by definition. On the other hand, we should point out where the inner inclusion becomes loose. In our case, we have that the grounded extension may be strictly included in the intersection of all preferred extensions, so we put this loose inner inclusion on the map.

The strong validation of the map can no longer be done using a single argumentation framework, because of the behavior of the stable semantics. Indeed, the strong validation for $\mathcal{ST}$ is $F : \mathcal{ST} \smallsetminus \mathcal{GR} \neq \varnothing$, while for the semi-stable semantics it is $F : \mathcal{SST} \smallsetminus (\mathcal{ST} \cup \mathcal{GR}) \neq \varnothing$. However, $\mathcal{SST}$ is distinct from $\mathcal{ST}$ only when $\mathcal{ST}$ provides no extension (Caminada, 2006b). Thus, the two conditions cannot be simultaneously satisfied by an argumentation framework. We will provide two frameworks for the strong validation of our map: $F_1$ from Figure 3.12 and $F_2$ from Figure 3.13.

The first framework satisfies the validation condition for $\mathcal{SST}$, so implicitly $\mathcal{E}_{\mathcal{ST}}(F_1) = \varnothing$. The conditions corresponding to all the other semantics are satisfied. In fact, $F_1$ can also be regarded as the argumentation framework that strongly validates the map without $\mathcal{ST}$. The second framework satisfies the validation condition for $\mathcal{ST}$, so implicitly $\mathcal{E}_{\mathcal{ST}}(F_2) \neq \varnothing$ which leads to $\mathcal{E}_{\mathcal{SST}}(F_2) = \mathcal{E}_{\mathcal{ST}}(F_2)$. On the other hand $F_2$ satisfies the validation conditions for all the other semantics, acting as a framework that strongly validates the map without $\mathcal{SST}$. Taken together, the two frameworks provide the strong validation for the whole map.

We start with framework $F_1$, the one from Figure 3.12. To see that it does indeed strongly validate the map without $\mathcal{ST}$, we provide again the extensions that distinguish each semantics from all the others (except super-semantics).

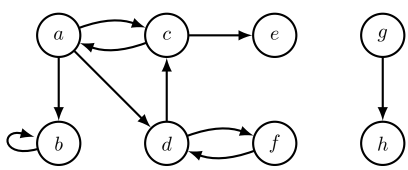



**Figure 3.12:** *Argumentation framework $F_1$ that validates the map from Figure 3.11 when there are no stable extensions: $\mathcal{E}_{\mathcal{ST}}(F_1) = \varnothing$.*

$$\begin{aligned}
&\delta_{F_1}(\mathcal{CF}) = \{\{a,e,h\},\{a,f,h\},\{a,h\},\{c\},\{d\},\{d,e\},\{d,h\},\{e\},\{e,f\}\}\\
&\delta_{F_1}(\mathcal{CF}^{cr}) = \{\{a,e,f,h\},\{c,f,h\},\{c,h\},\{d,e,h\},\{e,f,h\},\{e,h\},\{f,h\},\{h\}\}\\
&\delta_{F_1}(\mathcal{CF}^{wr}) = \{\{d,g\}\}\\
&\delta_{F_1}(\mathcal{CF}^{cw}) = \{\{d,e,g\}\}\\
&\delta_{F_1}(\mathcal{CF}^{sr}) = \{\{c,g\},\{e,f,g\},\{e,g\}\}\\
&\delta_{F_1}(\mathcal{AS}) = \{\varnothing,\{a\},\{a,e\},\{a,e,f\},\{a,f\},\{c,f\},\{f\}\}\\
&\delta_{F_1}(\mathcal{AS}^{wr}) = \{\{a,e,g\},\{a,f,g\},\{a,g\}\}\\
&\delta_{F_1}(\mathcal{CO}) = \{\{f,g\}\}\\
&\delta_{F_1}(\mathcal{GR}) = \{\{g\}\}\\
&\delta_{F_1}(\mathcal{PR}) = \{\{c,f,g\}\}\\
&\delta_{F_1}(\mathcal{SST}) = \{\{a,e,f,g\}\}
\end{aligned} \tag{3.14}$$

The interesting information is related to the semantics that we have added in this subsection. We have that $\{f,g\}$ is a complete extension that is neither grounded, nor preferred. Furthermore, we have $\{c,f,g\}$ as a preferred extension that is not semi-stable and $\{a,e,f,g\}$ as a semi-stable extension that is not stable.

Furthermore, note that our example framework also demonstrates the loose inner inclusion between $\mathcal{GR}$ and $\mathcal{PR}$, as $\mathcal{E}_{\mathcal{GR}}(F) = \{g\}$ and $\mathcal{E}_{\mathcal{PR}}(F) = \{\{c,f,g\},\{a,e,f,g\}\}$, so the intersection of all preferred extensions is $\{f,g\}$.

Next, we consider the case when there are stable extensions, so $\mathcal{SST}$ coincides with $\mathcal{ST}$. For this case, the strong validation is provided by the framework from Figure 3.13.

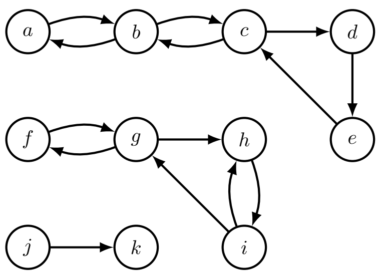



**Figure 3.13:** *Argumentation framework $F_2$ that validates the map from figure 3.11 when there is at least one stable extension: $\mathcal{E}_{\mathcal{ST}}(F_2) = \mathcal{E}_{\mathcal{SST}}(F_2) \neq \varnothing$.*

We provide the distinguishing extensions below. Note that whenever there are too many such extensions, we only provide one of them and the total number available. Again the framework also exhibits the distinction between the grounded extension and the intersection of all preferred extensions.

$$
\begin{aligned}
&\delta_{F_2}(\mathcal{CF}) \ni \{a,c\}, \qquad |\delta_{F_2}(\mathcal{CF})| = 80\\
&\delta_{F_2}(\mathcal{CF}^{cr}) \ni \{a,c,f,h,k\}, \qquad |\delta_{F_2}(\mathcal{CF}^{cr})| = 54\\
&\delta_{F_2}(\mathcal{CF}^{wr}) \ni \{a,c,i,j\}, \qquad |\delta_{F_2}(\mathcal{CF}^{wr})| = 15\\
&\delta_{F_2}(\mathcal{CF}^{cw}) \ni \{a,c,f,h,j\}, \qquad |\delta_{F_2}(\mathcal{CF}^{cw})| = 34\\
&\delta_{F_2}(\mathcal{CF}^{sr}) \ni \{a,h,j\}, \qquad |\delta_{F_2}(\mathcal{CF}^{sr})| = 8\\
&\delta_{F_2}(\mathcal{AS}) \ni \{a,f\}, \qquad |\delta_{F_2}(\mathcal{AS})| = 20\\
&\delta_{F_2}(\mathcal{AS}^{wr}) \ni \{a,i,j\}, \qquad |\delta_{F_2}(\mathcal{AS}^{wr})| = 8\\
&\delta_{F_2}(\mathcal{CO}) = \{\{a,f,j\},\{a,j\},\{b,d,f,j\},\{b,d,j\},\{f,h,j\},\{f,i,j\},\{f,j\}\}\\
&\delta_{F_2}(\mathcal{GR}) = \{\{j\}\}\\
&\delta_{F_2}(\mathcal{PR}) = \{\{a,f,h,j\},\{a,f,i,j\}\}\\
&\delta_{F_2}(\mathcal{ST}) = \{\{b,d,f,h,j\},\{b,d,f,i,j\}\}
\end{aligned}
\tag{3.15}
$$

### 3.2.4 The compact map of argumentation semantics

In this subsection we provide a map that covers several well-established argumentation semantics: naive ($\mathcal{MCF}$), stage ($\mathcal{STA}$), $\mathcal{CF}2$, ideal ($\mathcal{ID}$) and resolution-based grounded ($\mathcal{GR}^*$). We also include several newer semantics that are yet not present in many argumentation surveys: eager ($\mathcal{EAG}$), *stage*2 ($\mathcal{STA}2$) and enhanced preferred ($\mathcal{EPS}$). We have left out the prudent semantics, three of the SCC-recursive semantics ($\mathcal{AD}1$, $\mathcal{AD}2$ and $\mathcal{CF}1$) and the parameterized ideal semantics (together with all related sets). We will consider these for the extended map in the next subsection.

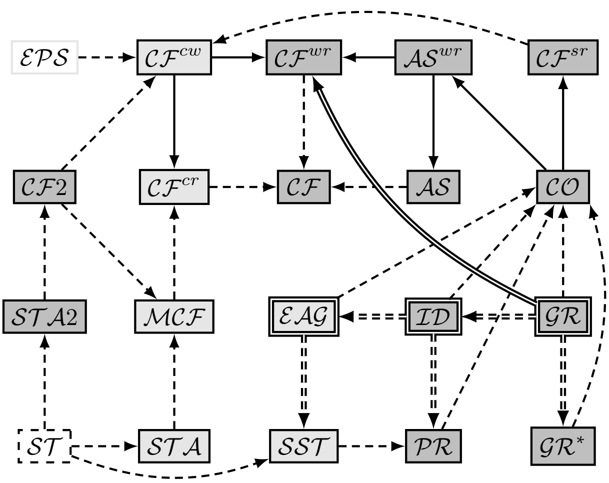



**Figure 3.14:** *Compact map of argumentation semantics.*

The map from Figure 3.14 is an extension of that presented in Figure 3.11. In what follows, we discuss the new relations and semantics included in the map.

A nice overview of naive-based semantics is presented in (Dvorak and Gaggl, 2012) and covers $\mathcal{MCF}$, $\mathcal{STA}$, $\mathcal{STA}2$ and $\mathcal{CF}2$. We use the following results: $\mathcal{MCF}$ satisfies $\mathcal{CF}$-reinstatement ($\mathcal{MCF} \subseteq \mathcal{CF}^{cr}$), all $\mathcal{CF}2$ extensions are also naive ($\mathcal{CF}2 \subseteq \mathcal{MCF}$), $\mathcal{CF}2$ satisfies both $\mathcal{CF}$ and weak reinstatement ($\mathcal{CF}2 \subseteq \mathcal{CF}^{cw}$), all stage extensions are also naive ($\mathcal{STA} \subseteq \mathcal{MCF}$), all *stage*2 extensions are also $\mathcal{CF}2$ extensions ($\mathcal{STA}2 \subseteq \mathcal{CF}2$), the stable semantics is included in both the stage and *stage*2 semantics ($\mathcal{ST} \subseteq \mathcal{STA}$ and $\mathcal{ST} \subseteq \mathcal{STA}2$). Some of these results are also presented in earlier works, such as (Baroni and Giacomin, 2007).

The ideal extension is complete ($\mathcal{ID} \subseteq \mathcal{CO}$) and, by definition, included in all preferred extensions ($\cup\mathcal{ID} \subseteq \cap\mathcal{PR}$); it also includes the grounded extension ($\cup\mathcal{GR} \subseteq \cap\mathcal{ID}$) (Dung et al., 2007). The eager extension is complete as well ($\mathcal{EAG} \subseteq \mathcal{CO}$) and included in all semi-stable extensions ($\cup\mathcal{EAG} \subseteq \cap\mathcal{SST}$); additionally, the ideal extension is included in the eager extension ($\cup\mathcal{ID} \subseteq \cap\mathcal{EAG}$) (Caminada, 2007).

Resolution-based grounded extensions are complete and, hence, also contain the grounded extension (Baroni and Giacomin, 2008) – $\mathcal{GR}^* \subseteq \mathcal{CO}$ and $\cup\mathcal{GR} \subseteq \cap\mathcal{GR}^*$. As far as $\mathcal{EPS}$ is concerned, we have seen in the previous section that it satisfies both $\mathcal{CF}$ and weak reinstatement, so we have $\mathcal{EPS} \subseteq \mathcal{CF}^{cw}$.

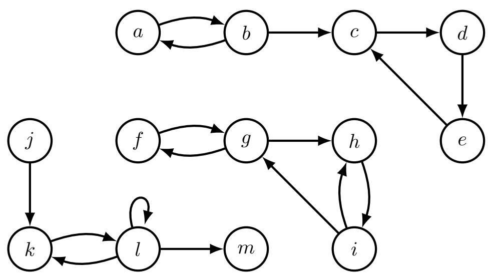



**Figure 3.15:** *Argumentation framework $F_1$ that validates the map from Figure 3.14 when there is at least one non-empty admissible set and there is no stable extension: $\mathcal{E}_{\mathcal{EPS}}(F_1) = \mathcal{E}_{\mathcal{PR}}(F_1) \neq \{\emptyset\}$, $\mathcal{E}_{\mathcal{ST}}(F_1) = \emptyset$.*

As far as the strong validation of the map is concerned, we already know that a single framework is not enough, because of the stable semantics. However, in this case we will need to also consider the behavior of the enhanced preferred semantics. Indeed, whenever $\mathcal{EPS}$ differs from $\mathcal{PR}$, we would only have $\emptyset$ as an admissible set (Zhang and Lin, 2010), which would make several semantics coincide. Indeed, all sub-semantics of $\mathcal{AS}$ would only give the empty set (or be void themselves, for the case of the stable semantics). Furthermore, we would have $\mathcal{CF}^{wr} = \mathcal{CF}$ and $\mathcal{CF}^{cw} = \mathcal{CF}^{cr}$. Thus, whenever we want to have $\mathcal{EPS}$ distinct from $\mathcal{PR}$, accounting for all the semantics that are trivialized leads to the simpler map from Figure 3.16.

First, we provide the framework from Figure 3.15 and show that it strongly validates the map from Figure 3.14, but without $\mathcal{EPS}$, which coincides in this case with $\mathcal{PR}$ and without $\mathcal{ST}$, which is void.

$$
\begin{aligned}
&\delta_{F_1}(\mathcal{CF}) \ni \{a,c\}, \quad |\delta_{F_1}(\mathcal{CF})| = 228\\
&\delta_{F_1}(\mathcal{CF}^{cr}) \ni \{a,c,f,k,m\}, \quad |\delta_{F_1}(\mathcal{CF}^{cr})| = 45\\
&\delta_{F_1}(\mathcal{CF}^{wr}) \ni \{a,c,i,j\}, \quad |\delta_{F_1}(\mathcal{CF}^{wr})| = 26\\
&\delta_{F_1}(\mathcal{CF}^{cw}) \ni \{a,c,f,h,j\}, \quad |\delta_{F_1}(\mathcal{CF}^{cw})| = 67\\
&\delta_{F_1}(\mathcal{CF}^{sr}) \ni \{a,f,h,j,m\}, \quad |\delta_{F_1}(\mathcal{CF}^{sr})| = 26\\
&\delta_{F_1}(\mathcal{AS}) \ni \{a,f\}, \quad |\delta_{F_1}(\mathcal{AS})| = 20\\
&\delta_{F_1}(\mathcal{AS}^{wr}) \ni \{a,i,j\}, \quad |\delta_{F_1}(\mathcal{AS}^{wr})| = 8\\
&\delta_{F_1}(\mathcal{CO}) = \{\{a,f,j\},\{f,h,j\},\{f,i,j\}\}\\
&\delta_{F_1}(\mathcal{GR}) = \{\{j\}\}\\
&\delta_{F_1}(\mathcal{PR}) = \{\{a,f,h,j\},\{a,f,i,j\}\}\\
&\delta_{F_1}(\mathcal{SST}) = \{\{b,d,f,h,j\},\{b,d,f,i,j\}\}\\
&\delta_{F_1}(\mathcal{ID}) = \{\{f,j\}\}\\
&\delta_{F_1}(\mathcal{EAG}) = \{\{b,d,f,j\}\}\\
&\delta_{F_1}(\mathcal{MCF}) \ni \{a,c,f,h,k,m\}, \quad |\delta_{F_1}(\mathcal{MCF})| = 13\\
&\delta_{F_1}(\mathcal{STA}) = \{\{b,d,f,h,k,m\},\{b,d,f,i,k,m\}\}\\
&\delta_{F_1}(\mathcal{CF}2) = \{\{a,c,g,j,m\},\{a,d,g,j,m\},\{a,e,g,j,m\},\{b,d,g,j,m\}\}\\
&\delta_{F_1}(\mathcal{STA}2) \ni \{a,c,f,h,j,m\}, \quad |\delta_{F_1}(\mathcal{STA}2)| = 6\\
&\delta_{F_1}(\mathcal{GR}^*) = \{\{a,j\},\{b,d,j\}\}
\end{aligned}
\tag{3.16}
$$

Additionally, note that the framework also validates some of the loose inner inclusions depicted on the map. Indeed, the grounded extension is $\{j\}$, the ideal extension is $\{f,j\}$ and the eager extension is $\{b,d,f,j\}$, so $\mathcal{GR} \subsetneq \mathcal{ID} \subsetneq \mathcal{EAG}$ in this case.

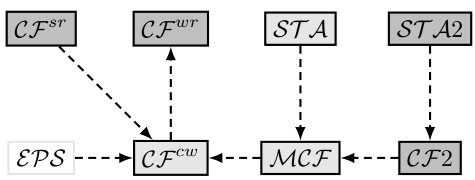



**Figure 3.16:** *Map of argumentation semantics that can be distinct and non-void when there is no non-empty admissible set.*

For the case when the stable semantics is non-void, an argumentation framework very similar to the one in Figure 3.13 can be used. We skip the details here.

Last, but not least, the simplified map corresponding to $\mathcal{E}_{\mathcal{AS}}(F) = \{\varnothing\}$, shown in Figure 3.16, is strongly validated by the framework from Figure 3.17. The distinguishing extensions in this case are:

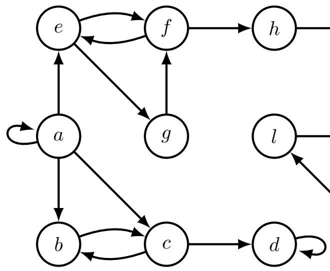



**Figure 3.17:** *Argumentation framework $F_2$ that validates the framework from Figure 3.14 when there is no non-empty admissible set. In fact $F_2$ strongly validates the map from Figure 3.16*

$$
\begin{aligned}
&\delta_{F_2}(\mathcal{CF}) \ni \{b,e\}, \qquad |\delta_{F_2}(\mathcal{CF})| = 45\\
&\delta_{F_2}(\mathcal{CF}^{cr}) \ni \{b,h,j\}, \qquad |\delta_{F_2}(\mathcal{CF}^{cr})| = 29\\
&\delta_{F_2}(\mathcal{CF}^{sr}) \ni \{b,g,h\}, \qquad |\delta_{F_2}(\mathcal{CF}^{sr})| = 15\\
&\delta_{F_2}(\mathcal{MCF}) \ni \{b,e,i,k\}, \qquad |\delta_{F_2}(\mathcal{MCF})| = 11\\
&\delta_{F_2}(\mathcal{STA}) = \{\{c,e,i,k\}\}\\
&\delta_{F_2}(\mathcal{CF}2) \ni \{b,f,i,k\}, \qquad |\delta_{F_2}(\mathcal{CF}2)| = 7\\
&\delta_{F_2}(\mathcal{STA}2) = \{\{b,e,h,j\},\{b,e,h,k\},\{b,e,h,l\}\}\\
&\delta_{F_2}(\mathcal{EPS}) = \{\{f,i,k\},\{l\}\}
\end{aligned}
\tag{3.17}
$$

Note that the framework from figure 3.17 consists of a single connected component, in contrast with the other examples we have provided so far for validation. This is because to the fact that the enhanced preferred semantics does not satisfy additivity.

### 3.2.5 Extended map of argumentation semantics

In this subsection we provide an extended map of argumentation semantics, covering all the 48 semantics that we have introduced. The properties of these semantics are presented in Table 3.3 and Table 3.4. In what follows we discuss the semantics that were not included in the compact map from the previous subsection.

The SCC-recursive semantics $\mathcal{AD}1$ and $\mathcal{AD}2$ give complete extensions (Baroni et al., 2005), so $\mathcal{AD}1 \subseteq \mathcal{CO}$ and $\mathcal{AD}2 \subseteq \mathcal{CO}$. As far as $\mathcal{CF}1$ is concerned, we have seen that it satisfies both $\mathcal{CF}$ and weak reinstatement, so $\mathcal{CF}1 \subseteq \mathcal{CF}^{cw}$.

From the definition of the prudent semantics we have $\mathcal{GR}^P \subseteq \mathcal{CO}^P \subseteq \mathcal{AS}^P \subseteq \mathcal{CF}^P$ and $\mathcal{PR}^P \subseteq \mathcal{AS}^P$. Furthermore, it is shown in (Coste-Marquis et al., 2005) that the prudent stable extensions are also prudent preferred, so $\mathcal{ST}^P \subseteq \mathcal{PR}^P$. In addition, we can also deduce from the definitions that $\mathcal{CF}^P \subseteq \mathcal{CF}$, $\mathcal{AS} \subseteq \mathcal{AS}^P$ and $\mathcal{ST}^P \subseteq \mathcal{ST}^P$.

Although not mentioned in (Coste-Marquis et al., 2005), all prudent preferred extensions are also prudent complete ($\mathcal{PR}^P \subseteq \mathcal{CO}^P$). Indeed, suppose that $S$ is a prudent preferred extension and $a$ is an argument such that $S$ defends $a$ and

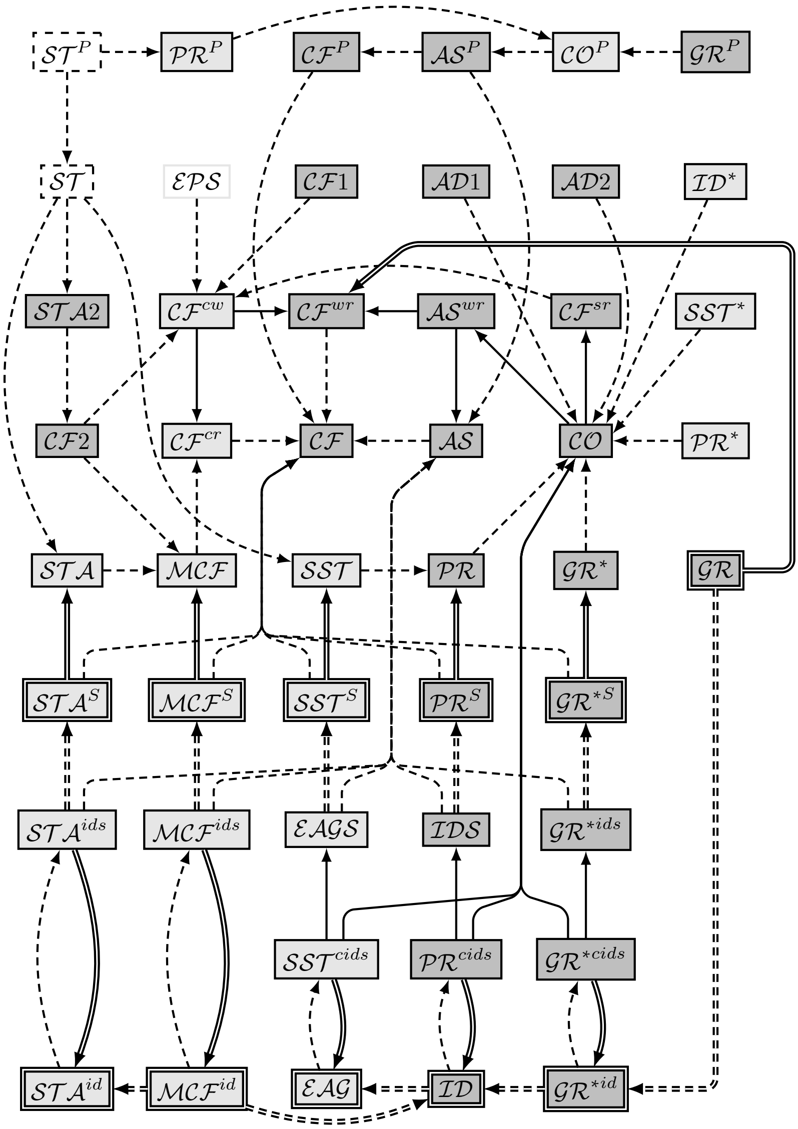



**Figure 3.18:** *Extended map of argumentation semantics.*

$S \cup \{a\}$ is without indirect conflicts. Then $S \cup \{a\}$ is also a $p$-admissible set. Since $S$ is a maximal $p$-admissible set, it must be that $S \cup \{a\} = S$, which leads to $a \in S$ and shows that $S$ is a complete $p$-extension.

Another aspect worth mentioning is the fact that we can see the definition of prudent complete semantics as enforcing a new kind of reinstatement on admissible sets. Note, however that this reinstatement is strictly weaker than $\mathcal{CF}$-reinstatement. Indeed, if $S \cup \{a\}$ is without indirect conflict, then $S \cup \{a\}$ is conflict-free.

The resolution-based semantics $\mathcal{PR}^*$, $\mathcal{SST}^*$ and $\mathcal{ID}^*$ give complete extensions, because the corresponding base semantics give complete extensions (Baroni et al., 2011b). Thus, we have $\mathcal{PR}^* \subseteq \mathcal{CO}$, $\mathcal{SST}^* \subseteq \mathcal{CO}$ and $\mathcal{ID}^* \subseteq \mathcal{CO}$.

For the remaining semantics we focus on the definitions and the most important properties. As the map is quite large, we no longer aim to provide all possible relations. We also allow a certain degree of redundancy for the sake of clarity and symmetry.

For any base semantics $\mathcal{S}em$, the skeptical semantics are, by definition, the intersection of all the extensions of $\mathcal{S}em$, so $\cup\mathcal{S}em^S = \cap\mathcal{S}em$. Furthermore, we have proved that $\mathcal{S}em^S \subseteq \mathcal{CF}$ and we show this on the map by grouping all the arrows towards $\mathcal{CF}$. This helps keep the map readable and also shows that those relations are similar.

For the ideal sets $\mathcal{S}em^{ids}$ we know that they are included in all $\mathcal{S}em$ extensions and, thus, also in $\mathcal{S}em^S$: $\cup\mathcal{S}em^{ids} \subseteq \cap\mathcal{S}em^S$. Also, we show on the map that $\mathcal{S}em^{ids} \subseteq \mathcal{AS}$, by definition. Complete ideal sets are ideal sets that are also complete, so we have $\mathcal{S}em^{cids} \subseteq \mathcal{S}em^{ids}$ and $\mathcal{S}em^{cids} \subseteq \mathcal{CO}$.

The ideal extension is the maximal ideal set (or even complete ideal set, for $\mathcal{PR}$, $\mathcal{SST}$ and $\mathcal{GR}^*$) corresponding to the base semantics. This leads to both $\mathcal{S}em^{id} \subseteq \mathcal{S}em^{ids}$ and $\cup\mathcal{S}em^{ids} \subseteq \cap\mathcal{S}em^{id}$ (or $\mathcal{S}em^{id} \subseteq \mathcal{S}em^{cids}$ and $\cup\mathcal{S}em^{cids} \subseteq \cap\mathcal{S}em^{id}$). The inner inclusion relations between ideal semantics: $\mathcal{GR} \subseteq \mathcal{GR}^{*id} \subseteq \mathcal{ID} \subseteq \mathcal{EAG}$, $\mathcal{MCF}^{id} \subseteq \mathcal{STA}^{id}$ and $\mathcal{MCF}^{id} \subseteq \mathcal{ID}$ are taken from (Dvorak et al., 2011).

Due to the large number of semantics included in the map and also to the fact that most of the argumentation semantics that are included here but not in the compact map have received a limited amount of attention in the argumentation literature (or they are very new), we will not pursue the strong validation in this case. This aspect will be considered in future research.

### 3.2.6 Discussion

This section has focused on the relations between argumentation frameworks and on the graphical representation of their properties. Here we discuss a few possible applications of our approach.

First of all, having a map that shows all semantics together, their most basic properties and the relations between them gives a good overview for researchers that are interested in using argumentation semantics or defining new ones. Suppose that a new argumentation semantics $\mathcal{S}em$ is proposed, possibly using an intricate algorithm that hides some of its properties. Comparing $\mathcal{S}em$ to all the argumentation semantics from the literature in order to make sure that it is indeed novel, or simply for showing how it relates to each of them, might be tedious. Instead, the map that we have proposed can be used as a guide for comparisons.

The first comparison to be performed is the one against the most general semantics, $\mathcal{CF}$ and the result should be that $\mathcal{S}em$ is included in $\mathcal{CF}$ – to the best of our knowledge, no semantics that provides non-conflict-free extensions has been proposed in the argumentation literature so far. From here on, one needs to compare with the immediate sub-semantics of $\mathcal{CF}$, thus establishing the satisfaction of some principles. Whenever it is found that $\mathcal{S}em$ is included in some semantics $\mathcal{S}em'$, comparisons can be performed against the immediate sub-semantics of

$\mathcal{S}em'$, until all minimal super-semantics of $\mathcal{S}em$ are identified and $\mathcal{S}em$ can be placed on the map as well.

Furthermore, the strong validation of a map provides examples that demonstrate its completeness, in the sense that all inclusion relations that cannot be inferred from the map do not generally hold. In addition, the strong validation frameworks can be used as a tool for recognizing argumentation semantics or for choosing a suitable semantics for an application domain. We discuss each aspect separately.

First, consider a multi-agent scenario based on argumentation, where each agent has its knowledge base represented as an argumentation framework and uses some argumentation semantics for taking decisions. It might be important for the interaction between agents to have information about the actual semantics used by each agent. In a heterogeneous environment the agents may also use different names for semantics, making it difficult to share this information with others, even if they are willing to do it. On the other hand, the frameworks that strongly validate maps of argumentation semantics can be used in this context. An agent can send such a framework to another agent and ask whether certain sets are seen as extensions by that agent. If extensions are picked from those that distinguish semantics from one another, this approach can lead to discovering the argumentation semantics that is used by the other agent, without the need for providing or knowing a name for it.

For the task of choosing the right argumentation semantics for a given application domain, one can consider the argumentation frameworks that strongly validate the map of the semantics that are considered as viable alternatives, and then see the distinguishing extensions. Based on the characteristics of the domain, some of those extensions may be desirable or not. This analysis may be pursued at the abstract level (more difficult) or may require the additional step of modeling an example within the application domain such that its corresponding framework is the one that strongly validates the map.

## 3.3 Chapter Summary

In this chapter we provided an overview of the satisfaction of several evaluation principles for argumentation semantics, filling in the gaps in the current literature where needed. We have also provided a graphical representation encoding all these properties, as well as the relations between semantics. The unique traits of each semantics were emphasized via argumentation frameworks strongly validating our maps.

Several of the most basic principles for evaluating argumentation semantics were presented in Section 3.1: cardinality, conflict-freeness, admissibility, reinstatement, directionality, non-interference. We have also discussed the satisfaction of these principles for each of the 43 argumentation semantics from Chapter 2, either by referring to existing results from the literature, where they are available, or by providing the proofs ourselves. We have also introduced a novel principle, additivity, which is closely related to non-interference, yet distinct from it. We proved that almost all argumentation semantics satisfy this principle. The most important immediate benefit of this property is the fact that for large frameworks the extensions can be computed separately for connected components, then aggregated to form the extensions of the whole framework. Table 3.3 from page 42 summarizes the results discussed in the section.

In Section 3.2 we offered a general overview of the argumentation semantics we have introduced, by providing a map representation that includes both the characteristics of each semantics and the relations between them. In order to capture the satisfaction of admissibility and reinstatement, we have introduced five novel semantics and also discussed their properties. Table 3.4 from page 50 summarizes the properties of these new semantics.

An extended map of all the 48 semantics, showing their properties and the relations between them, was presented in Figure 3.18 from page 58. Simpler maps covering just the more commonly used semantics were discussed as well. For those maps we have also provided argumentation frameworks that exhibit the distinguishing features of each semantics. This information can be applied for recognizing or choosing suitable argumentation semantics based on their behavior.

*Chapter 4*

# Normal Forms of Modal Formulas

In this chapter we discuss normal forms of modal formulas and their satisfaction. The most important results are related to the construction of models for the satisfiable modal formulas from $Ml9\Diamond, \mathbf{E}$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. We start by providing an introduction to modal languages, satisfaction and bisimulations in Section 4.1. We then discuss the negation normal form in Section 4.2. Furthermore, we cover the disjunctive normal form for $ML(\Diamond)$ and $ML(\Diamond, \overline{\Diamond})$ in sections 4.3 and 4.4, respectively. The normal form and the satisfaction of formulas from $ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ are presented in Section 4.5. We conclude with a summary of the chapter in Section 4.6.

## 4.1 Preliminaries

This section aims to provide an initial modal logic background, for this chapter but also for the rest of the thesis. We follow the approach in (Blackburn et al., 2001), but we only introduce here the concepts that are relevant for the results in this chapter and for a solid formal connection between modal logic and argumentation, to be discussed in Chapter 5.

### 4.1.1 Modal languages and satisfaction

In this subsection we discuss several modal language and their satisfaction. We start with the basic modal language.

**Definition 44.** *The* ***basic modal language*** $ML(\Diamond)$ *is defined by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \rightarrow \phi \mid \phi \leftrightarrow \phi \mid \Diamond\phi \mid \Box\phi \tag{4.1}$$

*where $p$ ranges over a finite set of propositional symbols $\mathcal{P}rop$.*

Note that in Definition 44 we have used an extended set of operators, as follows:
- constant truth / top / validity: $\top$
- constant falsum / bottom / inconsistency: $\bot$
- conjunction: $\phi \wedge \psi$ = "$\phi$ and $\psi$"
- disjunction: $\phi \vee \psi$ = "$\phi$ or $\psi$"

- implication: $\phi \rightarrow \psi$ = "if $\phi$, then $\psi$"
- bi-implication / equivalence: $\phi \leftrightarrow \psi$ = "$\phi$ if and only if $\psi$"
- diamond: $\Diamond\phi$ = "$\phi$ is possible"
- box: $\Box\phi$ = "$\phi$ is necessary"

For example, if we take $\mathcal{P}rop = \{p, q\}$, we have that $p \wedge \Diamond q$, $\Box(\Diamond p \rightarrow q)$ and $\Diamond\phi \leftrightarrow (q \wedge \Diamond q)$ are syntactically well-formed formulas from $ML(\Diamond)$. Unary operators have the highest precedence, followed (in order) by conjunction, disjunction, implication and equivalence. We will use parentheses whenever needed.

**Definition 45.** *A* ***Kripke model*** *is a tuple $\mathfrak{M} = (W, R, V)$, where $W$ is a non-empty set of* ***worlds*** *(or* ***states****), $R \subseteq W \times W$ is a binary* ***accessibility relation*** *between worlds and $V : \mathcal{P}rop \rightarrow \mathcal{P}(W)$ is a* ***valuation function*** *that gives, for each propositional symbol $p \in \mathcal{P}rop$, the set of worlds $V(p) \subseteq W$ where $p$* ***holds*** *(or is* ***satisfied*** */* ***true****).*

*Given a model $\mathfrak{M} = (W, R, V)$ and a world $w \in W$, the pair $(\mathfrak{M}, w)$ is referred to as a* ***pointed model****.*

An example Kripke model is presented in Figure 4.1. The implied set of propositional symbols is $\mathcal{P}rop = \{p, q\}$. For each world $w$ and each propositional symbol $x \in \mathcal{P}rop$, we put $x$ next to $w$ if $x$ is true at $w$ and $\neg x$ otherwise. Thus, the model from the figure is given by:

$$\begin{aligned} W &= \{w_1, w_2, w_3, w_4, w_5, w_6\} \\ R &= \{(w_1, w_2), (w_1, w_3), (w_3, w_4), (w_3, w_5), (w_5, w_6)\} \\ V(p) &= \{w_1, w_2, w_4, w_6\} \\ V(q) &= \{w_1, w_2, w_3\} \end{aligned} \tag{4.2}$$

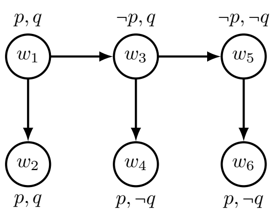



**Figure 4.1:** *Example Kripke model*

We can also define the disjoint union of two models, which is similar to the disjoint union of two argumentation frameworks, discussed in Chapter 3.

**Definition 46.** *Two models $\mathfrak{M}_1 = (W_1, R_1, V_1)$ and $\mathfrak{M}_2 = (W_2, R_2, V_2)$ are* ***disjoint*** *if $W_1 \cap W_2 = \varnothing$. The* ***disjoint union*** *of $\mathfrak{M}_1$ and $\mathfrak{M}_2$ (notation $\mathfrak{M}_1 \uplus \mathfrak{M}_2$) is the model $\mathfrak{M} = (W, R, V)$ given by:*

$$\begin{aligned} W &= W_1 \cup W_2 \\ R &= R_1 \cup R_2 \\ V(p) &= V_1(p) \cup V_2(p), \quad \text{for all } p \in \mathcal{P}rop \end{aligned} \tag{4.3}$$

We see that the valuation function tells us whether propositional symbols are true (or satisfied) at a world in a model. This idea can be extended to more complex formulas.

**Definition 47.** *For any Kripke model $\mathfrak{M} = (W, R, V)$ and any world $w \in W$, we recursively define the* ***satisfaction*** *of a formula $\phi$ at $w$ in $\mathfrak{M}$ (notation $\mathfrak{M}, w \Vdash \phi$) as follows:*

$$\begin{aligned}
&\mathfrak{M}, w \Vdash \top \\
&\mathfrak{M}, w \nVdash \bot \\
&\mathfrak{M}, w \Vdash p \Leftrightarrow w \in V(p) \\
&\mathfrak{M}, w \Vdash \neg\phi \Leftrightarrow \mathfrak{M}, w \nVdash \phi \\
&\mathfrak{M} w \Vdash \phi \wedge \psi \Leftrightarrow \mathfrak{M}, w \Vdash \phi \texttt{ and } \mathfrak{M}, w \Vdash \psi \\
&\mathfrak{M}, w \Vdash \phi \vee \psi \Leftrightarrow \mathfrak{M}, w \Vdash \phi \texttt{ or } \mathfrak{M}, w \Vdash \psi \\
&\mathfrak{M}, w \Vdash \phi \rightarrow \psi \Leftrightarrow (\mathfrak{M}, w \Vdash \phi \Rightarrow \mathfrak{M}, w \Vdash \psi) \\
&\mathfrak{M}, w \Vdash \phi \leftrightarrow \psi \Leftrightarrow (\mathfrak{M}, w \Vdash \phi \Leftrightarrow \mathfrak{M}, w \Vdash \psi) \\
&\mathfrak{M}, w \Vdash \Diamond\phi \Leftrightarrow \exists w'(w' \in W \texttt{ and } (w, w') \in R \texttt{ and } \mathfrak{M}, w' \Vdash \phi) \\
&\mathfrak{M}, w \Vdash \Box\phi \Leftrightarrow \forall w'(w' \in W \texttt{ and } (w, w') \in R \Rightarrow \mathfrak{M}, w' \Vdash \phi)
\end{aligned} \tag{4.4}$$

*Whenever $\mathfrak{M}, w \Vdash \phi$, we say that $\phi$ is* ***satisfied*** *(or* ***true****) at $w$ in $\mathfrak{M}$. A formula $\phi$ is* ***satisfiable*** *iff there exists a pointed model that satisfies it.*

For example, let us see again the model from Figure 4.1 and show that $\mathfrak{M}, w_1 \Vdash q \wedge \Box\Diamond\neg p$. First, we have $\mathfrak{M}, w_1 \Vdash q$ because $w_1 \in V(q)$. Furthermore, there exists a world accessible from $w_1$, namely $w_3$, such that all worlds accessible from it ($w_4$ and $w_5$) satisfy $\neg q$. Thus, $\mathfrak{M}, w_1 \Vdash \Diamond\Box\neg q$, which leads to the desired result. Also note that, because $w_2$ has no successors, we have $\mathfrak{M}, w_2 \Vdash \Box\phi$ and $\mathfrak{M}, w_2 \nVdash \Diamond\phi$, for any formula $\phi$.

Note that we have used `and`, `or`, $\Rightarrow$ and $\Leftrightarrow$ instead of $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$ respectively, in order to distinguish the meta-language from the actual modal language. We will stick to this convention throughout the chapter.

**Definition 48.** *Given a model $\mathfrak{M} = (W, R, V)$, we define the* ***extended valuation function*** *$V^*$ as follows:*

$$V^*(\phi) = \{w \in W \mid \mathfrak{M}, w \Vdash \phi\} \tag{4.5}$$

In words, the extended valuation function gives the set of worlds that satisfy a given formula. Note that we have not specified a particular modal language, as this is applicable to any of the languages that will be used here. As an example, for the model from Figure 4.1, we can write that $V^*(q \wedge \Diamond\Box\neg q) = \{w_1, w_3\}$.

Note that the basic modal language presented in (Blackburn et al., 2001) contains several modalities, of various arities, each of them interpreted with respect to a different relation. Since we are interested in the use of modal logic for argumentation, we only need one binary relation that we can link to the attack relation from argumentation. On the other hand, it is also possible to add new modalities based on the same relation or using a distinct but fixed relation. In what follows, we introduce the converse and the global modalities, the corresponding languages and the satisfiability of these operators.

**Definition 49.** *The* ***converse modal language*** $ML(\Diamond, \overline{\Diamond})$ *is defined by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \rightarrow \phi \mid \phi \leftrightarrow \phi \mid \Diamond\phi \mid \Box\phi \mid \overline{\Diamond}\phi \mid \overline{\Box}\phi \tag{4.6}$$

*where* $p$ *ranges over a finite set of propositional symbols* $\mathcal{P}rop$*.*

What is new with respect to the basic modal language are the converse diamond $\overline{\Diamond}$ and the converse box $\overline{\Box}$. Examples of well-formed formulas from the converse modal language include: $p \wedge \overline{\Diamond}\Diamond q$, $\overline{\Box}(p \wedge \Diamond q)$ and $\neg\Diamond p \rightarrow \overline{\Diamond}\neg p$.

Modal satisfaction of formulas from $ML(\Diamond, \overline{\Diamond})$ is similar to the one we have seen for the basic modal language, all we need to do is add the rules for the converse modalities.

**Definition 50.** *Modal satisfaction of* $ML(\Diamond, \overline{\Diamond})$ *formulas is recursively defined by the rules provided in Definition 47 plus the following rules for the converse modalities:*

$$\begin{aligned} \mathfrak{M}, w \Vdash \overline{\Diamond}\phi &\Leftrightarrow \exists w'(w' \in W \textit{ and } (w', w) \in R \textit{ and } \mathfrak{M}, w' \Vdash \phi) \\ \mathfrak{M}, w \Vdash \overline{\Box}\phi &\Leftrightarrow \forall w'(w' \in W \textit{ and } (w', w) \in R \Rightarrow \mathfrak{M}, w' \Vdash \phi) \end{aligned} \tag{4.7}$$

What we see from Definition 50 is that the converse modalities are quite similar to the basic ones, but they look "backwards" at the accessibility relation $R$. For the model from Figure 4.1 we can see that $\mathfrak{M}, w_5 \Vdash \neg q \wedge \overline{\Diamond}q$.

**Definition 51.** *The* ***global modal language*** $ML(\Diamond, \mathbf{E})$ *is defined by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \rightarrow \phi \mid \phi \leftrightarrow \phi \mid \Diamond\phi \mid \Box\phi \mid \mathbf{E}\phi \mid \mathbf{A}\phi \tag{4.8}$$

*where* $p$ *ranges over a finite set of propositional symbols* $\mathcal{P}rop$*.*

Examples of well-formed formulas using the global modalities: $\mathbf{A}p \vee \mathbf{E}\neg q$, $\mathbf{E}(p \wedge \Diamond q)$, $\mathbf{E}(p \wedge \mathbf{A}\neg q)$. The meaning of these operators is revealed in Definition 52.

**Definition 52.** *Modal satisfaction of* $ML(\Diamond, \mathbf{E})$ *formulas is recursively defined by the rules provided in Definition 47 plus the following rules for the global modalities:*

$$\begin{aligned} \mathfrak{M}, w \Vdash \mathbf{E}\phi &\Leftrightarrow \exists w'(w' \in W \textit{ and } \mathfrak{M}, w' \Vdash \phi) \\ \mathfrak{M}, w \Vdash \mathbf{A}\phi &\Leftrightarrow \forall w'(w' \in W \Rightarrow \mathfrak{M}, w' \Vdash \phi) \end{aligned} \tag{4.9}$$

$\mathbf{E}$ is called the existential modality, while $\mathbf{A}$ is the universal modality. Note that the evaluation of $\mathbf{E}$ and $\mathbf{A}$ does not use the accessibility relation $R$, hence justifying the name "global". Alternatively, we can say that the global modalities are always evaluated with respect to the relation $W \times W$. Let us see that $\mathfrak{M}, w_1 \Vdash \mathbf{A}(p \vee \Diamond\neg q)$, where $\mathfrak{M}$ is the model from Figure 4.1. For this, we need to show that $\mathfrak{M}, w \Vdash p \vee \Diamond\neg q$ for all $w \in W$. This is true for $w_1$, $w_2$, $w_4$ and $w_6$, as they satisfy $p$. For $w_3$ and $w_5$ the result follows from the fact that $\Diamond\neg q$ holds at these worlds.

Since the global and the converse modalities are independent from each other, we can add them both to $ML(\Diamond)$ and get an even more expressive modal language.

**Definition 53.** *The* ***global converse modal language*** $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ *is defined by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \to \phi \mid \phi \leftrightarrow \phi \mid \Diamond\phi \mid \Box\phi \mid \overline{\Diamond}\phi \mid \overline{\Box}\phi \mid \mathbf{E}\phi \mid \mathbf{A}\phi \quad (4.10)$$

*where $p$ ranges over a set of propositional symbols $\mathcal{P}rop$.*

**Definition 54.** *Modal satisfaction of $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ formulas is recursively defined by combining the rules provided in Definition 47, Definition 50 and Definition 52.*

For example, the formula $\mathbf{E}(\Box p \wedge \overline{\Box} q)$ holds iff there exists a world $w$ such that all predecessors of $w$ satisfy $q$ and all successors of $w$ satisfy $p$.

We have seen that the evaluation of modal formulas is based on Kripke models. There are also some formulas, called validities, that are satisfied by any model.

**Definition 55.** *A formula $\phi$ is* ***valid*** *(alternatively, $\phi$ is a* ***validity****) iff, for any model $\mathfrak{M} = (W, R, V)$ and any world $w \in W$ it holds that $\mathfrak{M}, w \Vdash \phi$. We will write this as $\Vdash \phi$. The set of all valid formulas from a given language is called the* ***logic*** *of that language.*

The introduction we have provided so far is largely semantical. We are not interested in the axiomatization of modal logics here, as we will only use semantical proofs for our results. For such details, the reader may consult (Blackburn et al., 2001).

The syntactical aspects that we are going to use are validities of the form $\phi \leftrightarrow \psi$, showing that two formulas $\phi$ and $\psi$ behave the same. This will be useful for transforming modal formulas into more desirable forms. For now, we provide the relations between dual operators as examples of such equivalences.

The logical operators (and constants) that we have seen so far can be grouped into the following pairs: $(\bot, \top)$, $(\wedge, \vee)$, $(\Diamond, \Box)$, $(\overline{\Diamond}, \overline{\Box})$ and $(\mathbf{E}, \mathbf{A})$. In order to provide more concise presentations in the rest of this chapter, we will use the following notations:

- $atom \subseteq \{\bot, \top, p, \neg p\}$ – a generic atomic construct from the language
- $\bigcirc \subseteq \{\neg, \Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\}$ – a generic unary operator
- $\oplus \subseteq \{\wedge, \vee, \to, \leftrightarrow\}$ – a generic binary operator

Furthermore, $atom_{dual}$, $\bigcirc_{dual}$ and $\oplus_{dual}$ will denote the corresponding dual. The following equivalences are valid:

$$\begin{aligned}
& atom \leftrightarrow \neg atom_{dual}, && \text{for } atom \in \{\bot, \top\} \\
& \bigcirc\phi \leftrightarrow \neg \bigcirc_{dual} \neg\phi, && \text{for } \bigcirc \in \{\Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\} \\
& \phi \oplus \psi \leftrightarrow \neg(\neg\phi \oplus_{dual} \neg\psi), && \text{for } \oplus \in \{\wedge, \vee\} \\
& (\phi \to \psi) \leftrightarrow (\neg\phi \vee \psi) \\
& (\phi \leftrightarrow \psi) \leftrightarrow ((\phi \to \psi) \wedge (\psi \to \phi))
\end{aligned} \quad (4.11)$$

Intuitively, the evaluation of $\Diamond\overline{\Box}\Box p$ is more complex than that of $\Diamond p$, as we need to test the formulas satisfied at worlds in a larger neighborhood of the current world. This idea is captured by the modal degree of a formula.

**Definition 56.** *The **modal depth** (or **modal degree**) md of a global converse modal formula is recursively defined as follows:*

$$\begin{aligned}
&md(atom) = 0, \qquad \textit{for } atom \in \{\bot, \top, p\}\\
&md(\neg\phi) = md(\phi)\\
&md(\phi \oplus \psi) = \max(md(\phi), md(\psi)), \qquad \textit{for } \oplus \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}\\
&md(\bigcirc\phi) = 1 + md(\phi), \qquad \textit{for } \bigcirc \in \{\Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\}
\end{aligned} \tag{4.12}$$

For example, we have that $md(\mathbf{E}(\Diamond p \wedge \overline{\Box}\Box\neg q)) = 3$ and $md(\Diamond p \rightarrow q) = 1$. The modal degree will be used in the discussion of disjunctive normal forms.

Other modalities can be added, as well as operators for combining modalities, leading to a great variety of modal logics. In this chapter, however, we only focus on the modal languages that we have already introduced. We will see a few other languages in the first section of Chapter 5, in the context of describing argumentation semantics.

### 4.1.2 Bisimulations

We have seen that there exist valid equivalences $\phi \leftrightarrow \psi$, which means that, for every pointed model $(\mathfrak{M}, w)$, we have that $\mathfrak{M}, w \Vdash \phi \Leftrightarrow \mathfrak{M}, w \Vdash \psi$. In what follows, we discuss bisimulation, a property that leads to the equivalence of models with respect to modal satisfaction.

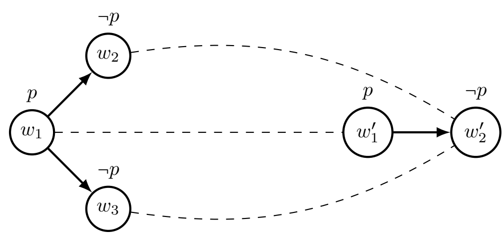



**Figure 4.2:** *Bisimulation example.*

Let us look at the two models from Figure 4.2, call them $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$. It is easy to see that $\mathfrak{M}, w_1 \Vdash \Box\neg p$ and $\mathfrak{M}', w_1' \Vdash \Box\neg p$. Also, $\mathfrak{M}, w_2 \Vdash \overline{\Diamond} p$ and $\mathfrak{M}', w_2' \Vdash \overline{\Diamond} p$. Similarly, $\mathfrak{M}, w_3' \Vdash \mathbf{A}(\neg p \vee \Diamond\neg p)$ and $\mathfrak{M}', w_2' \Vdash \mathbf{A}(\neg p \vee \Diamond\neg p)$. In fact, for any two worlds $w \in W$ and $w' \in W'$ that are connected by a dashed line and for any formula $\phi \in ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ we have that $\mathfrak{M}, w \Vdash \phi \Leftrightarrow \mathfrak{M}', w' \Vdash \phi$. The dashed lines from the figure encode what is introduced in Definition 57 as total converse bisimulation.

**Definition 57.** *Let $\mathfrak{M}_1 = (W_1, R_1, V_1)$ and $\mathfrak{M}_2 = (W_2, R_2, V_2)$ be two models and let $Z \subseteq W_1 \times W_2$ be a binary relation between the worlds of the two models. We consider the following properties:*

*(a) for any $(w_1, w_2) \in Z$ and for any $p \in \mathcal{P}rop$, $w_1 \in V_1(p) \Leftrightarrow w_2 \in V_2(p)$*

*(b) for any $w_1, w_1' \in W_1$ and $w_2 \in W_2$, if $(w_1, w_2) \in Z$ and $(w_1, w_1') \in R_1$ then there exists a world $w_2' \in W_2$ such that $(w_2, w_2') \in R_2$ and $(w_1', w_2') \in Z$ (zig for the basic modality)*

(c) *for any* $w_1 \in W_1$ *and* $w_2, w_2' \in W_2$*, if* $(w_1, w_2) \in Z$ *and* $(w_2, w_2') \in R_2$ *then there exists a world* $w_1' \in W_1$ *such that* $(w_1, w_1') \in R_1$ *and* $(w_1', w_2') \in Z$ *(zag for the basic modality)*

(d) *for any* $w_1, w_1' \in W_1$ *and* $w_2' \in W_2$*, if* $(w_1', w_2') \in Z$ *and* $(w_1, w_1') \in R_1$ *then there exists a world* $w_2 \in W_2$ *such that* $(w_2, w_2') \in R_2$ *and* $(w_1, w_2) \in Z$ *(zig for converse)*

(e) *for any* $w_1' \in W_1$ *and* $w_2, w_2' \in W_2$*, if* $(w_1', w_2') \in Z$ *and* $(w_2, w_2') \in R_2$ *then there exists a world* $w_1 \in W_1$ *such that* $(w_1, w_1') \in R_1$ *and* $(w_1, w_2) \in Z$ *(zag for converse)*

(f) *for any world* $w_1 \in W_1$ *there exists a world* $w_2 \in W_2$ *such that* $(w_1, w_2) \in Z$ *(zig for the global modality)*

(g) *for any world* $w_2 \in W_2$ *there exists a world* $w_1 \in W_1$ *such that* $(w_1, w_2) \in Z$ *(zag for the global modality)*

*Based on these properties, we define the following:*

- $Z$ *is a* ***(basic) bisimulation*** *between* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$ *iff* $Z$ *satisfies (a), (b) and (c)*
- $Z$ *is a* ***converse bisimulation*** *between* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$ *iff* $Z$ *satisfies (a), (b), (c), (d) and (e)*
- $Z$ *is a* ***total bisimulation*** *between* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$ *iff* $Z$ *satisfies (a), (b), (c), (f) and (g)*
- $Z$ *is a* ***total converse bisimulation*** *between* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$ *iff* $Z$ *satisfies all properties*

*If there exists a* $\beta$*-bisimulation* $Z$ *between two models* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$*,* $\beta \in \{$ *basic, converse, total, total converse* $\}$*, we say that the models are* $\beta$*-bisimilar and write this as* $\mathfrak{M}_1 \leftrightarrows \mathfrak{M}_2$ *(for the basic bisimulation, respectively* $\leftrightarrows_c$*,* $\leftrightarrows_t$ *and* $\leftrightarrows_{tc}$ *for the other three). If* $w_1$ *is a world from* $\mathfrak{M}_1$ *and* $w_2$ *is a world from* $\mathfrak{M}_2$ *such that* $(w_1, w_2) \in Z$*, we say that the two pointed models are* $\beta$*-bisimilar and we write this as* $\mathfrak{M}_1, w_1 \leftrightarrows \mathfrak{M}_2, w_2$ *(with subscripts for bisimulations other than the basic one).*

As shown in (Blackburn et al., 2001), bisimilar models cannot be distinguished by modal logics. We present this result formally in Proposition 11.

**Proposition 11.** *Modal satisfaction is invariant under bisimulation, in the sense that, for every two pointed models* $(\mathfrak{M}_1, w_1)$ *and* $(\mathfrak{M}_2, w_2)$ *we have:*

$$\begin{aligned}
&\mathfrak{M}_1, w_1 \leftrightarrows \mathfrak{M}_2, w_2 \Rightarrow \forall\phi(\phi \in ML(\Diamond) \Rightarrow (\mathfrak{M}_1, w_1 \Vdash \phi \Leftrightarrow \mathfrak{M}_2, w_2 \Vdash \phi))\\
&\mathfrak{M}_1, w_1 \leftrightarrows_c \mathfrak{M}_2, w_2 \Rightarrow \forall\phi(\phi \in ML(\Diamond, \overline{\Diamond}) \Rightarrow (\mathfrak{M}_1, w_1 \Vdash \phi \Leftrightarrow \mathfrak{M}_2, w_2 \Vdash \phi))\\
&\mathfrak{M}_1, w_1 \leftrightarrows_t \mathfrak{M}_2, w_2 \Rightarrow \forall\phi(\phi \in ML(\Diamond, \mathbf{E}) \Rightarrow (\mathfrak{M}_1, w_1 \Vdash \phi \Leftrightarrow \mathfrak{M}_2, w_2 \Vdash \phi))\\
&\mathfrak{M}_1, w_1 \leftrightarrows_{tc} \mathfrak{M}_2, w_2 \Rightarrow \forall\phi(\phi \in ML(\Diamond, \overline{\Diamond}, \mathbf{E}) \Rightarrow (\mathfrak{M}_1, w_1 \Vdash \phi \Leftrightarrow \mathfrak{M}_2, w_2 \Vdash \phi))
\end{aligned} \tag{4.13}$$

In words, there is a correspondence between the types of bisimulation from Definition 57 and the modal languages introduced in this subsection. Whenever two pointed models are bisimilar, any formula from the corresponding modal language is either satisfied by both models or not satisfied by either of them. Alternatively, we say that the modal formulas of that language cannot distinguish the two models from one another.

It is rather easy to see that the relation indicated by the dashed lines in Figure 4.2 is a total converse bisimulation between the two models. For us, bisimulation is useful in the following sense: we know that formulas from $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ cannot distinguish the two models. Thus, any function or algorithm that does distinguish them cannot be characterized within $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. As an example, consider computing for $w_1$ and $w'_1$ the number of successors that satisfy $\neg p$. The results will be distinct so we can conclude that $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ is not expressive enough for distinguishing the number of successors that satisfy a certain formula. We will apply a similar reasoning for showing that certain modal logics lack the expressive power for describing certain argumentation semantics.

### 4.1.3 Notations

We end this subsection with a few notations that will help us write long formulas more compactly:

$$\begin{aligned} \bigvee \Phi &\triangleq \bigvee_{\phi \in \Phi} \phi \\ \bigwedge \Phi &\triangleq \bigwedge_{\phi \in \Phi} \phi \\ \odot\, \Phi &\triangleq (\bigwedge_{\phi \in \Phi} \Diamond \phi) \wedge \Box (\bigvee_{\phi \in \Phi} \phi) \\ \overline{\odot} \Phi &\triangleq (\bigwedge_{\phi \in \Phi} \overline{\Diamond} \phi) \wedge \overline{\Box} (\bigvee_{\phi \in \Phi} \phi) \end{aligned} \tag{4.14}$$

for any set of formulas $\Phi$. We will also use

$$pr(\Phi) = \bigwedge_{p \in \Phi} p \wedge \bigwedge_{p \in \mathcal{P}rop \setminus \Phi} \neg p \tag{4.15}$$

where $\Phi \subseteq \mathcal{P}rop$. We will use lowercase Greek letters for modal formulas and uppercase Greek letters for sets of formulas.

These notations are to be understood just as abbreviations of the longer formulas they represent. For example, $\odot\{p, \neg q\}$ still stands for $\Diamond p \wedge \Diamond \neg q \wedge \Box (p \vee \neg q)$, so we do not talk about a modal language that includes $\odot$.

## 4.2 Negation Normal Form

One normal form that is widely used in logic, often without being given a name, is the negation normal form. Having a formula in this form generally simplifies proofs, as fewer operators need to be considered. We provide the definition of the negation normal form for $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. For modal logics using only s subset of the modalities, the corresponding terms from the BNF are to be removed.

**Definition 58.** *A modal formula from $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ is in negation normal form iff it is generated by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg p \mid \phi \wedge \phi \mid \phi \vee \phi \mid \Diamond \phi \mid \Box \phi \mid \overline{\Diamond} \phi \mid \overline{\Box} \phi \mid \mathbf{E} \phi \mid \mathbf{A} \phi \tag{4.16}$$

*where $p$ ranges over the set $\mathcal{P}rop$ of propostional symbols. We denote the corresponding language with $ML(\Diamond, \overline{\Diamond}, \mathbf{E})_{NNF}$.*

The distinguishing feature of the negation normal form, aside from not using $\rightarrow$ and $\leftrightarrow$, is that the negation only appears next to propositional symbols. A normal form is most useful if there is also a possibility for translating arbitrary formulas to their equivalent normal form. In this case, the translation is fairly straightforward. We provide it here nevertheless for having a formal grasp on things, with the use of a uniform notation.

The first step for reaching the negation normal form is to replace all occurrences of the equivalence operator $\leftrightarrow$ by using the validity $(\phi \leftrightarrow \psi) \leftrightarrow (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$. More formally, what we need to do is apply the following translation function:

$$\begin{aligned}
&tr_{eq}(atom) = atom, \quad \textit{for atom} \in \{\top, \bot, p\}\\
&tr_{eq}(\bigcirc\phi) = \bigcirc tr_{eq}(\phi), \quad \textit{for } \bigcirc \in \{\neg, \Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\}\\
&tr_{eq}(\phi \oplus \psi) = tr_{eq}(\phi) \oplus tr_{eq}(\psi), \quad \textit{for } \oplus \in \{\wedge, \vee, \rightarrow\}\\
&tr_{eq}(\phi \leftrightarrow \psi) = (tr_{eq}(\phi) \rightarrow tr_{eq}(\psi)) \wedge (tr_{eq}(\psi) \rightarrow tr_{eq}(\psi))
\end{aligned} \tag{4.17}$$

Next, we also replace the implication operator using $\phi \rightarrow \psi \leftrightarrow \neg\phi \vee \psi$, which leads to the following translation function:

$$\begin{aligned}
&tr_{imp}(atom) = atom, \quad \textit{for atom} \in \{\top, \bot, p\}\\
&tr_{imp}(\bigcirc\phi) = \bigcirc tr_{imp}(\phi), \quad \textit{for } \bigcirc \in \{\neg, \Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\}\\
&tr_{imp}(\phi \oplus \psi) = tr_{imp}(\phi) \oplus tr_{imp}(\psi), \quad \textit{for } \oplus \in \{\wedge, \vee\}\\
&tr_{imp}(\phi \rightarrow \psi) = \neg tr_{imp}(\phi) \vee tr_{imp}(\psi)
\end{aligned} \tag{4.18}$$

Last, but not least, we need to move the negations toward the propositional symbols. To this end, we use equivalences that involve dual operators and negation. For the binary operators $\wedge$ and $\vee$, these are the well known De Morgan laws:

$$\begin{aligned}
\neg(\phi \wedge \psi) &\leftrightarrow \neg\phi \vee \neg\psi\\
\neg(\phi \vee \psi) &\leftrightarrow \neg\phi \wedge \neg\psi
\end{aligned} \tag{4.19}$$

For the negation itself, we use $\neg\neg\phi \leftrightarrow \phi$. We also have the following equivalences involving the dual modalities:

$$\begin{aligned}
\neg\Diamond\phi \leftrightarrow \Box\neg\phi &\quad \neg\Box\phi \leftrightarrow \Diamond\neg\phi\\
\neg\overline{\Diamond}\phi \leftrightarrow \overline{\Box}\neg\phi &\quad \neg\overline{\Box}\phi \leftrightarrow \overline{\Diamond}\neg\phi\\
\neg\mathbf{E}\phi \leftrightarrow \mathbf{A}\neg\phi &\quad \neg\mathbf{A}\phi \leftrightarrow \mathbf{E}\neg\phi
\end{aligned} \tag{4.20}$$

The translation function that uses the above equivalences in order to obtain the negation normal form of a formula is given by:

$$\begin{aligned}
&tr_{neg}(atom) = atom, \quad \textit{for } atom \in \{\top, \bot, p\} \\
&tr_{neg}(\bigcirc\phi) = \bigcirc tr_{neg}(\phi), \quad \textit{for } \bigcirc \in \{\Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\} \\
&tr_{neg}(\phi \oplus \psi) = tr_{neg}(\phi) \oplus tr_{neg}(\psi), \quad \textit{for } \oplus \in \{\lor, \land\} \\
&tr_{neg}(\neg atom) = atom_{dual}, \quad \textit{for } atom \in \{\top, \bot\} \\
&tr_{neg}(\neg p) = \neg p, \quad \textit{for } p \in \mathcal{P}rop \\
&tr_{neg}(\neg\neg\phi) = tr_{neg}(\phi) \\
&tr_{neg}(\neg \bigcirc \phi) = \bigcirc_{dual} tr_{neg}(\neg\phi), \quad \textit{for } \bigcirc \in \{\Diamond, \Box, \overline{\Diamond}, \overline{\Box}, \mathbf{E}, \mathbf{A}\} \\
&tr_{neg}(\neg(\phi \oplus \psi)) = tr_{neg}(\neg\phi) \oplus_{dual} tr_{neg}(\neg\psi), \quad \textit{for } \oplus \in \{\lor, \land\}
\end{aligned} \tag{4.21}$$

**Definition 59.** *The function $tr_{NNF}$ that translates an arbitrary formula from $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ to a negation normal form is defined as:*

$$tr_{NNF} = tr_{neg} \circ tr_{imp} \circ tr_{eq} \tag{4.22}$$

*where $tr_{eq}$, $tr_{imp}$ and $tr_{neg}$ are the translations given in (4.17), (4.18) and (4.21), respectively.*

**Proposition 12.** *For every formula $\phi \in ML(\Diamond, \overline{\Diamond}, \mathbf{E})$, it holds that $tr_{NNF}(\phi) \in ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ and $\Vdash \phi \leftrightarrow tr_{NNF}(\phi)$.*

In other words, Proposition 12 states that $tr_{NNF}(\phi)$ does indeed provide a formula that is in negation normal form and, even more, it is also equivalent to $\phi$. The proof is a rather straightforward induction using the fact that all the performed translations are based on equivalences, as we have seen already. We end the presentation of NNF with an example. We consider the formula $\varphi = p \leftrightarrow \Box\Diamond p$. Its translation is:

$$\begin{aligned}
tr_{NNF}(\varphi) &= tr_{neg}(tr_{imp}(tr_{eq}(\varphi))) \\
&= tr_{neg}(tr_{imp}(tr_{eq}(p \leftrightarrow \Box\Diamond p))) \\
&= tr_{neg}(tr_{imp}((p \rightarrow \Box\Diamond p) \land (\Box\Diamond p \rightarrow p))) \\
&= tr_{neg}((\neg p \lor \Box\Diamond p) \land (\neg\Box\Diamond p \lor p)) \\
&= (\neg p \lor \Box\Diamond p) \land (\Diamond\Box\neg p \lor p)
\end{aligned} \tag{4.23}$$

For the other normal form translations in this chapter we will assume that the initial formula is already in negation normal form, in order to simplify the formal proofs.

## 4.3 Disjunctive Normal Form for $ML(\Diamond)$

In this section we rely on the disjunctive normal form (DNF) for $ML(\Diamond)$ that was introduced in (Fine, 1975) and provide an equivalent form that allows a more compact representation using the notations from (4.14).

### 4.3.1 Modal minterms and their satisfaction

We start by introducing modal minterms, the building blocks for the disjunctive normal form.

**Definition 60.** *Let $\mathcal{P}rop = \{p_1, p_2, \dots, p_n\}$ be the set of propositional symbols from the basic modal language $ML(\Diamond)$. The set of **modal minterms of degree** $n$ (notation: $\mathcal{F}_n$) is recursively defined as follows:*

$$\begin{aligned} \mathcal{F}_0 &= \{\bigwedge_{k=1}^{n} \widetilde{p_k} \mid \widetilde{p_k} = p_k \text{ or } \widetilde{p_k} = \neg p_k\} \\ \mathcal{F}_{n+1} &= \{\pi \wedge \bigwedge_{\phi \in \mathcal{F}_n} \widetilde{\Diamond\phi} \mid \pi \in \mathcal{F}_0, \widetilde{\Diamond\phi} = \Diamond\phi \text{ or } \widetilde{\Diamond\phi} = \neg\Diamond\phi\}\}, \quad \text{for all } n \geq 0 \end{aligned} \tag{4.24}$$

For example, if $\mathcal{P}rop = \{p\}$, then we have:

$$\begin{aligned} \mathcal{F}_0 &= \{p, \neg p\} \\ \mathcal{F}_1 &= \{p \wedge \Diamond p \wedge \Diamond\neg p, p \wedge \Diamond p \wedge \neg\Diamond\neg p, p \wedge \neg\Diamond p \wedge \Diamond\neg p, p \wedge \neg\Diamond p \wedge \neg\Diamond\neg p, \\ &\quad \neg p \wedge \Diamond p \wedge \Diamond\neg p, \neg p \wedge \Diamond p \wedge \neg\Diamond\neg p, \neg p \wedge \neg\Diamond p \wedge \Diamond\neg p, \neg p \wedge \neg\Diamond p \wedge \neg\Diamond\neg p\} \end{aligned} \tag{4.25}$$

Let us see that the number of modal minterms increases very rapidly with $n$: $|\mathcal{F}_0| = 2^{|\mathcal{P}rop|}$ and $|\mathcal{F}_{n+1}| = 2^{|\mathcal{P}rop|+|F_n|}$. This means that such minterms are not very practical when an exhaustive analysis is needed. However, in formal proofs and also for small values of $n$ they will be very useful.

**Proposition 13.** *Let $\mathfrak{M} = (W, R, V)$ be a model and $w \in W$ a world in $\mathfrak{M}$. Then, for any $n$, there exists exactly one modal minterm of degree $n$ that is satisfied at $w$ in $\mathfrak{M}$.*

Intuitively, the conjuncts of the modal minterms of degree $n$ are the propositional symbols and the diamond of each modal minterm of degree $n-1$, taken in all possible combinations. Since each such conjunct must be either true or false at a given world, it follows that exactly one such combination will hold. For a formal proof, see (Fine, 1975).

**Definition 61.** *Let $n$ be an integer. The modal minterm of degree $n$ that is satisfied at a world $w$ in a model $\mathfrak{M}$ is denoted by $\sigma_n(w)$:*

$$\sigma_n(w) \in \mathcal{F}_n \text{ and } \mathfrak{M}, w \Vdash \sigma_n(w) \tag{4.26}$$

*Such a minterm exists and is unique according to Proposition 13.*

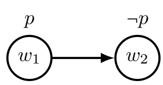



**Figure 4.3:** *Kripke model to exemplify the modal minterm satisfied at a given world.*

As an example. for the simple model from Figure 4.3 we have that $\sigma_1(w_1) = p \wedge \neg\Diamond p \wedge \Diamond\neg p$. To compute $\sigma_i(w)$ in the general case, one needs to test the satisfaction of propositional symbols at $w$ and the satisfaction of $\mathcal{F}_{i-1}$ minterms for successors of $w$ (in other words, compute the value of $\sigma_{i-1}$ for all successors).

Let us now discuss the actual meaning of a modal minterm of degree $n$. If we group the positive and the negative conjuncts together, we can write such a minterm as

$$\psi = \bigwedge_{p\in\Phi_0} p \wedge \bigwedge_{p\in\mathcal{P}rop\setminus\Phi_0} \neg p \wedge \bigwedge_{\phi\in\Phi_1} \Diamond\phi \wedge \bigwedge_{\phi\in\mathcal{F}_{n-1}\setminus\Phi_1} \neg\Diamond\phi \tag{4.27}$$

for some $\Phi_0 \subseteq \mathcal{P}rop$ and $\Phi_1 \subseteq \mathcal{F}_{n-1}$.

The meaning of the modal part is that the minterms from $\Phi_1$, and only those, are satisfied at worlds accessible from the current one. The same idea can be conveyed by the shorthand notation $\diamonddot\Phi_1 = (\bigwedge_{\phi\in\Phi_1}\Diamond\phi)\wedge\Box(\bigvee_{\phi\in\Phi_1}\phi)$. We formalize this intuition in Proposition 14.

**Proposition 14.** *Let $n \geq 0$ be an integer and let $\Phi \subseteq \mathcal{F}_n$ be a set of modal minterms of degree $n$. Then we have that:*

$$\Box(\bigvee_{\phi\in\Phi}\phi) \leftrightarrow \bigwedge_{\phi\in\mathcal{F}_n\setminus\Phi} \neg\Diamond\phi \tag{4.28}$$

*Proof.* We provide a semantical argument. If $\Box(\bigvee_{\phi\in\Phi}\phi)$ holds at a world $w$ in a model, then all worlds accessible from $w$ will satisfy one minterm from $\Phi$. But then none of the minterms from $\mathcal{F}_n \setminus \Phi$ will be satisfied. Conversely, if no minterm from $\mathcal{F}_n \setminus \Phi$ is satisfied by worlds accessible from $w$, then $\Box\bigvee_{\phi\in\Phi}\phi$ holds at $w$. □

With this, any modal minterm of degree $n$ is equivalent to a formula that can be written compactly using the shorter notations $\diamonddot$ and $pr$. We redefine modal minterms based on this observation.

**Definition 62.** *Let $\mathcal{P}rop = \{p_1, p_2, \ldots, p_n\}$ be the set of propositional symbols from the basic modal language $ML(\Diamond)$. The set of* ***modal minterms of degree*** *$n$ (notation: $\mathcal{F}_n$) is recursively defined as follows:*

$$\begin{aligned}\mathcal{F}_0 &= \{pr(\Phi_0) \mid \Phi_0 \subseteq \mathcal{P}rop\}\\ \mathcal{F}_{n+1} &= \{pr(\Phi_0) \wedge \diamonddot\Phi_1 \mid \Phi_0 \subseteq \mathcal{P}rop, \Phi_1 \subseteq \mathcal{F}_n\}, \quad \textit{for all } n \geq 0\\ \mathcal{F} &= \bigcup_{n\geq 0}\mathcal{F}_n\end{aligned} \tag{4.29}$$

Using this notation, the modal minterms of degree 0 and 1 for $\mathcal{P}rop = \{p\}$ can be written as:

$$\begin{aligned}\mathcal{F}_0 &= \{p, \neg p\}\\ \mathcal{F}_1 &= \{p\wedge\diamonddot\{p,\neg p\}, p\wedge\diamonddot\{p\}, p\wedge\diamonddot\{\neg p\}, p\wedge\diamonddot\varnothing,\\ &\quad \neg p\wedge\diamonddot\{p,\neg p\}, \neg p\wedge\diamonddot\{p\}, \neg p\wedge\diamonddot\{\neg p\}, \neg p\wedge\diamonddot\varnothing\}\end{aligned} \tag{4.30}$$

Note that there was no need to use the $pr$ notation, as $p$ is shorter than $pr(\{p\})$, but in the general case $pr$ can also contribute to shortening the representation of the minterm.

As an example, let us see that the rule for recursively computing $\sigma_i(w)$ can be written as: $\sigma_i(w) = pr(\{p \in \mathcal{P}rop \mid w \in V(p)\}) \wedge \diamonddot\{\sigma_{i-1}(w') \mid (w, w') \in R\}$, which is quite intuitive in this form. For the model from Figure 4.3 this leads to $\sigma_i(w_0) = p \wedge \diamonddot\{\neg p\}$.

**Proposition 15.** *For every modal minterm of degree $n$, $\phi \in \mathcal{F}_n$, there exists a model $\mathfrak{M} = (W, R, V)$ and a world $w \in W$ such that $\mathfrak{M}, w \Vdash \phi$.*

*Proof.* We will use $\mathcal{M}(\phi)$ to refer to the pointed model $(\mathfrak{M}, w)$ that we are looking for. We define $\mathcal{M}$ as a recursive function and use induction to show that the provided pointed model does indeed satisfy $\phi$. For the base case, $n = 0$, we define $\mathcal{M}$ as follows:

$$\mathcal{M}(pr(\Phi_0)) = ((\{w\}, \varnothing, V), w), \quad V(p) = \begin{cases} \{w\}, & \textit{for } p \in \Phi_0 \\ \varnothing, & \textit{for } p \in \mathcal{P}rop \setminus \Phi_0 \end{cases} \tag{4.31}$$

where $\Phi_0 \subseteq \mathcal{P}rop$. It is fairly obvious in this case that $\mathcal{M}(\phi) \Vdash \phi$.

For the induction step, we use the values of $\mathcal{M}(\phi)$ for $\phi \in \mathcal{F}_{n-1}$ in order to construct the models for modal minterms of degree $n$. For any $\Phi_0 \subseteq \mathcal{P}rop$ and any $\Phi_1 \subseteq \mathcal{F}_{n-1}$ we have:

$$\mathcal{M}(pr(\Phi_0) \wedge \diamond\Phi_1) = ((W_0 \cup \bigcup_{\phi \in \Phi_1} W_\phi, R_0 \cup \bigcup_{\phi \in \Phi_1} (R_\phi \cup \{(w_0, w_\phi)\}), V), w_0) \tag{4.32}$$

where we have used $\mathcal{M}(\phi) = ((W_\phi, R_\phi, V_\phi), w_\phi)$ for $\phi \in \Phi_1$ and $\mathcal{M}(pr(\Phi_0)) = ((W_0, R_0, V_0), w_0)$. The valuation function $V$ is given by:

$$V(p) = V_0(p) \cup \bigcup_{\phi \in \Phi_1} V_\phi(p) \tag{4.33}$$

In words, what we do is take the disjoint union of the models for $pr(\Phi_0)$ and $\phi \in \Phi_1$ (these exist based on the induction hypothesis), then extend the accessibility relation so that all $w_\phi$'s are accessible from $w_0$. Doing this does not affect the satisfaction of $\phi$ at $w_\phi$. Furthermore, with the extended accessibility relation, we have that $\mathcal{M}(pr(\Phi_0) \wedge \diamond\Phi_1) \Vdash pr(\Phi_0) \wedge \diamond\Phi_1$, which completes the induction. □

In other words, we have shown that every modal minterm from $\mathcal{F}$ is satisfiable. We will provide an example for the construction of a model that satisfies a given modal minterm in the next subsection, where we will also have to deal with the converse modality.

### 4.3.2 Translation to normal form

Based on the modal minterms introduced in the previous subsection and also on the satisfiability results presented there, we are ready to introduce the disjunctive normal form and to discuss the translation of an arbitrary formula from the basic modal language to an equivalent DNF.

**Definition 63.** *A modal formula $\phi \in ML(\Diamond)$ is in **disjunctive normal form** (DNF) iff it is the disjunction of a (possibly empty) set of minterms of degree $n$, for some integer $n$. We will denote the corresponding language with $ML(\Diamond)_{DNF}$.*

$$ML(\Diamond)_{DNF} = \{\bigvee \Phi \mid \exists n(n \geq 0 \wedge \Phi \subseteq \mathcal{F}_n)\} \tag{4.34}$$

Next, we discuss the translation of an arbitrary formula $\phi \in ML(\Diamond)$ to an equivalent disjunctive normal form. We assume that the formula is first converted to negation normal form using the translation function $tr_{NNF}$ discussed in the previous subsection. We will use the translation function $tr_n$, defined only for modal formulas of modal degree $\le n$. For the atomic elements of $ML(\Diamond)_{NNF}$, we have:

$$\begin{aligned}
tr_n(\bot) &= \bigvee \varnothing \\
tr_n(\top) &= \bigvee \mathcal{F}_n \\
tr_0(p) &= \bigvee \{pr(\Phi_0) \mid \Phi_0 \subseteq \mathcal{P}rop, p \in \Phi_0\} \\
tr_n(p) &= \bigvee \{pr(\Phi_0) \wedge \diamondcdot \Phi_1 \mid \Phi_0 \subseteq \mathcal{P}rop, p \in \Phi_0, \Phi_1 \subseteq \mathcal{F}_{n-1}\} \\
tr_0(\neg p) &= \bigvee \{pr(\Phi_0) \mid \Phi_0 \subseteq \mathcal{P}rop, p \notin \Phi_0\} \\
tr_n(\neg p) &= \bigvee \{pr(\Phi_0) \wedge \diamondcdot \Phi_1 \mid \Phi_0 \subseteq \mathcal{P}rop, p \notin \Phi_0, \Phi_1 \subseteq \mathcal{F}_{n-1}\}
\end{aligned} \tag{4.35}$$

For conjunction and disjunction, we assume that we already have $tr_n(\phi) = \bigvee \Phi$ and $tr_n(\psi) = \bigvee \Psi$, with $\Phi \subseteq \mathcal{F}_n$ and $\Psi \subseteq \mathcal{F}_n$. Then:

$$\begin{aligned}
tr_n(\phi \vee \psi) &= \bigvee (\Phi \cup \Psi) \\
tr_n(\phi \wedge \psi) &= \bigvee (\Phi \cap \Psi)
\end{aligned} \tag{4.36}$$

Last, but not least, for the modal operators, we assume we know $tr_{n-1}(\phi) = \bigvee \Phi$ and we get:

$$\begin{aligned}
tr_n(\Diamond \phi) &= \bigvee \{pr(\Phi_0) \wedge \diamondcdot \Phi_1 \mid \Phi_0 \subseteq \mathcal{P}rop, \Phi_1 \subseteq \mathcal{F}_{n-1}, \Phi_1 \cap \Phi \neq \varnothing\} \\
tr_n(\Box \phi) &= \bigvee \{pr(\Phi_0) \wedge \diamondcdot \Phi_1 \mid \Phi_0 \subseteq \mathcal{P}rop, \Phi_1 \subseteq \Phi\}
\end{aligned} \tag{4.37}$$

With these, we can define the translation of a modal formula to its disjunctive normal form.

**Definition 64.** *For every formula $\phi \in ML(\Diamond)$ and for every $n \ge md(\phi)$ we define the translation to disjunctive normal form of degree $n$ as:*

$$tr_{DNF,n}(\phi) = tr_n(tr_{NNF}(\phi)) \tag{4.38}$$

*where $tr_n$ is the translation function defined by relations* (4.35), (4.36) *and* (4.37).

Note that given a formula $\phi$ of modal degree $n$, we can convert it to its equivalent normal form for any $k \ge n$. In practice we will prefer the normal form of the same modal degree as the formula, especially because of the exponential blowup of the number of minterms with respect to $n$.

**Proposition 16.** *For every formula $\phi \in ML(\Diamond)$ and every $n \ge md(\phi)$ it holds that $tr_{DNF,n}(\phi) \in ML(\Diamond)_{DNF}$ and $\phi \leftrightarrow tr_{DNF,n}(\phi)$ is a validity.*

*Proof.* We already know from Proposition 12 that $tr_{NNF}$ preserves modal satisfaction. All we need to do is prove that the same holds for $tr_n$. To this end, let us first see that this is the case for the relations in (4.35) and (4.36), by using Proposition

13 and Proposition 15. For (4.37) we need to prove that, for any $\Psi_1 \subseteq \Psi_2 \subseteq \mathcal{F}_n$, the following holds:

$$\bigwedge_{\psi \in \Psi_1} \Diamond\psi \wedge \Box\Big(\bigvee_{\psi \in \Psi_2} \psi\Big) \leftrightarrow \bigvee\{\Diamonddot\Psi \mid \Psi_1 \subseteq \Psi \subseteq \Psi_2\} \tag{4.39}$$

We provide a semantical proof. What we need is the set of all $\Diamonddot\Psi$'s that are satisfiable iff the left hand side of the formula is satisfiable. Since $\Diamond\psi$ must hold for all $\psi \in \Psi_1$, it must be that $\Psi_1 \subseteq \Psi$. Furthermore, $\Box(\bigvee_{\psi \in \Psi_2} \psi)$ is satisfiable iff all accessible worlds satisfy one of the formulas from $\Psi_2$. This means that we must also have $\Psi \subseteq \Psi_2$. This proves the $\rightarrow$ part of the equivalence.

For the $\leftarrow$ part, it is easy to see that whenever $\Diamonddot\Psi$ holds, for $\Psi_1 \subseteq \Psi \subseteq \Psi_2$, the left hand side of the formula holds as well. All that we need to do is remember that $\Diamonddot\Psi$ stands for $\bigwedge_{\psi \in \Psi} \Diamond\psi \wedge \Box \bigvee_{\psi \in \Psi} \psi$. □

As an example, let us compute the 1-st degree normal form of $\varphi = p \wedge \Diamond\neg p$. We have:

$$\begin{aligned}
tr_{DNF,1}(p) &= \bigvee\{p \wedge \Diamonddot\varnothing, p \wedge \Diamonddot\{p\}, p \wedge \Diamonddot\{\neg p\}, p \wedge \Diamonddot\{p, \neg p\}\} \\
tr_{DNF,0}(\neg p) &= \bigvee\{\neg p\} \\
tr_{DNF,1}(\Diamond\neg p) &= \bigvee\{p \wedge \Diamonddot\{\neg p\}, p \wedge \Diamonddot\{p, \neg p\}, \neg p \wedge \Diamonddot\{\neg p\}, \neg p \wedge \Diamonddot\{p, \neg p\}\} \\
tr_{DNF,1}(p \wedge \Diamond\neg p) &= \bigvee\{p \wedge \Diamonddot\{\neg p\}, p \wedge \Diamonddot\{p, \neg p\}\}
\end{aligned} \tag{4.40}$$

If we substitute the abbreviations with the formulas they stand for, we get the actual normal form:

$$tr_{DNF,1}(p \wedge \Diamond\neg p) = p \wedge \Diamond\neg p \wedge \Box\neg p \vee p \wedge \Diamond p \wedge \Diamond\neg p \wedge \Box(p \vee \neg p) \tag{4.41}$$

## 4.4 Disjunctive normal form for $ML(\Diamond, \overline{\Diamond})$

In this section we discuss the disjunctive normal form for $ML(\Diamond, \overline{\Diamond})$. We shall see that the ideas from the previous section need several adjustments in order to fit this more expressive language, especially because both the basic modalities and the converse ones are evaluated with respect to the same accessibility relation.

### 4.4.1 Satisfiable modal minterms

We expect modal minterms of the converse modal language to be similar to those of the basic modal language, so we start with the following tentative form:

$$\begin{aligned}
\mathcal{G}_0 &= \{pr(\Phi) \mid \Phi \subseteq \mathcal{P}rop\} \\
\mathcal{G}_{n+1} &= \{pr(\Phi_0) \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \mid \Phi_0 \subseteq \mathcal{P}rop, \Phi_1 \subseteq \mathcal{G}_n, \Phi_2 \subseteq \mathcal{G}_n\}, \quad \textit{for all } n \geq 0 \\
\mathcal{G} &= \bigcup_{n \geq 0} \mathcal{G}_n \\
\mathcal{G}^* &= \bigcup_{n \geq 1} \mathcal{G}_n
\end{aligned} \tag{4.42}$$

Note that the exponential blowup in the number of minterms for this case is even worse than in the case of $ML(\Diamond)$, as we have $|\mathcal{G}_0| = 2^{|\mathcal{P}rop|}$ and $|\mathcal{G}_{n+1}| = 2^{|\mathcal{P}rop|+2|\mathcal{G}_n|}$.

On the other hand, let us see that in the presence of the converse modality not all such minterms are satisfiable. This happens because both modalities work with the same relation and, thus, some of the combinations are inconsistent. Indeed, consider the formula $\xi = p \wedge \Diamonddot\{p \wedge \Diamonddot\varnothing \wedge \overline{\Diamonddot}\{\neg p\}\} \wedge \overline{\Diamonddot}\varnothing \in \mathcal{G}_2$. This formula is satisfied at $w$ in a model iff $w \in V(p)$ and there is a world $w' \in W$ such that $(w, w') \in R$ and $\mathfrak{M}, w' \Vdash \overline{\Diamonddot}\{\neg p\}$. But then it would mean that $p$ should not hold at $w$, which is a contradiction.

In order to preserve some of the nice properties that we have seen for the modal minterms of $ML(\Diamond)$, such as the fact that each minterm is satisfiable, and also for reducing the exponential blowup to a minimum, we need to filter out all the unsatisfiable minterms. Given a modal minterm $pr(\Phi_0) \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \in \mathcal{G}_n$, $n \geq 1$, we have a part that looks at the current world – $\Phi_0$, a part that looks forward – $\Phi_1$, and a part that looks backward – $\Phi_2$. Since the elements of $\Phi_1$ and $\Phi_2$ are minterms as well, they also look forward and backward for $n \geq 2$. Thus, we need to make sure that the current world is consistent with the information seen by forward terms when looking backward and also with information seen by backward terms when looking forward. In what follows, we aim to formalize this intuition. First, we define the reduction of a minterm as the minterm that we obtain by reducing its degree with 1.

**Definition 65.** *Let $\phi = pr(\Phi_0) \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \in G^*$ be a modal minterm. The **reduction** of $\phi$ is defined as:*

$$\overline{\rho}(\phi) = \begin{cases} pr(\Phi_0), & md(\phi) = 1 \\ pr(\Phi_0) \wedge \Diamonddot\{\overline{\rho}(\phi_1) \mid \phi_1 \in \Phi_1\} \wedge \overline{\Diamonddot}\{\overline{\rho}(\phi_2) \mid \phi_2 \in \Phi_2\}, & md(\phi) > 1 \end{cases} \tag{4.43}$$

Note that we have used $\overline{\rho}$ to denote reduction. This is because we will also use the reduction function for basic modal formulas later on and we want to distinguish between the two. We will generally use a bar above the name of functions that work with the converse modal language, especially when we also use similar functions with the basic modal language as well.

In order to identify the satisfiable minterms from $\mathcal{G}_n$ we will also use the following straightforward result:

**Proposition 17.** *If $\mathfrak{M}, w \Vdash \phi$, with $\phi \in \mathcal{G}^*$, then $\mathfrak{M}, w \Vdash \overline{\rho}(\phi)$.*

Now, suppose the minterm $\phi = pr(\Phi_0) \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2$, with $\Phi_0 \subseteq \mathcal{P}rop$, $\Phi_1 \subseteq \mathcal{G}_n$ and $\Phi_2 \subseteq \mathcal{G}_n$, is satisfied at a world $w$ in a model $\mathfrak{M}$ and let $\psi \in \Phi_1$. Then there is a world $w'$, accessible from $w$, such that $\mathfrak{M}, w' \Vdash \psi$. On the other hand, $\psi$ is itself a minterm, of degree $n-1$, so we can write it as $\psi = pr(\Psi_0) \wedge \Diamonddot\Psi_1 \wedge \overline{\Diamonddot}\Psi_2$, with $\Psi_0 \subseteq \mathcal{P}rop$, $\Psi_1 \subseteq \mathcal{G}_{n-2}$ and $\Psi_2 \subseteq \mathcal{G}_{n-2}$. But then this means that any predecessor of $w'$, including $w$, must satisfy one formula from $\Psi_2$. But it is quite intuitive that the modal minterm of degree $n-2$ satisfied at $w$ is $\overline{\rho}(\overline{\rho}(\phi))$. Thus, in order for $\phi$ to be satisfiable, it should hold that $\overline{\rho}(\overline{\rho}(\phi)) \in \Psi_2$. This should happen for all minterms $\psi \in \Phi_1$ and a similar condition can be formulated for the elements of $\Phi_2$.

**Definition 66.** *The* ***satisfiable modal minterms of degree*** $n$ *for* $ML(\Diamond, \overline{\Diamond})$ *are recursively defined as follows:*

$$
\begin{aligned}
\overline{\mathcal{F}}_0 =& \{pr(\Phi_0) \mid \Phi \subseteq \mathcal{P}rop\} \\
\overline{\mathcal{F}}_1 =& \{pr(\Phi_0) \wedge \diamonddot\Phi_1 \wedge \overline{\diamonddot}\Phi_2 \mid \Phi_0 \subseteq \mathcal{P}rop, \Phi_1 \subseteq \overline{\mathcal{F}}_0, \Phi_2 \subseteq \overline{\mathcal{F}}_0\} \\
\overline{\mathcal{F}}_{n+1} =& \{\phi = pr(\Phi_0) \wedge \diamonddot\Phi_1 \wedge \overline{\diamonddot}\Phi_2 \mid \Phi_0 \subseteq \mathcal{P}rop, \\
& \Phi_1 \subseteq \overline{\mathcal{F}}_n, \forall\psi(\psi \in \Phi_1 \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\overline{\diamonddot}), \\
& \Phi_2 \subseteq \overline{\mathcal{F}}_n, \forall\psi(\psi \in \Phi_2 \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\diamonddot)\}, \quad \textit{for all } n \geq 1 \\
\overline{\mathcal{F}} =& \bigcup_{n \geq 0} \overline{\mathcal{F}}_n
\end{aligned}
\tag{4.44}
$$

*Where, for any minterm* $\phi = pr(\Phi_0) \wedge \diamonddot\Phi_1 \wedge \overline{\diamonddot}\Phi_2$*, we define* $\phi/\diamonddot \triangleq \Phi_1$ *and* $\phi/\overline{\diamonddot} \triangleq \Phi_2$*.*

**Proposition 18.** *Let* $\phi$ *be a modal minterm from* $\overline{\mathcal{F}}_n$*, with* $n \geq 1$*. Then* $\overline{\rho}(\phi)$ *is in* $\overline{\mathcal{F}}_{n-1}$*.*

*Proof.* Clearly we have that $\overline{\rho}(\phi) \in \mathcal{G}_{n-1}$. What we need to check is whether the satisfiability constraints imposed by $\overline{\mathcal{F}}_{n-1}$ are satisfied. From $\forall\psi(\psi \in \phi/\diamonddot \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\overline{\diamonddot})$ we can deduce that $\forall\psi(\psi \in \overline{\rho}(\phi)/\diamonddot \Rightarrow \overline{\rho}(\overline{\rho}(\overline{\rho}(\phi))) \in \psi/\overline{\diamonddot})$, which is the corresponding condition for $\overline{\rho}(\phi)$. The second condition can be deduced similarly. ☐

Since for $n = 0$ and $n = 1$ there is actually no restriction on the minterms ($\overline{\mathcal{F}}_0 = \mathcal{G}_0$ and $\overline{\mathcal{F}}_1 = \mathcal{G}_1$), we will discuss 2-nd degree examples. We take $\mathcal{P}rop = \{p\}$.

Let us first consider the formula $\xi = p \wedge \diamonddot\{p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\neg p\}\} \wedge \overline{\diamonddot}\varnothing$, which we know is not satisfiable. We need to test whether $\overline{\rho}(\overline{\rho}(\phi)) \in p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\neg p\}/\overline{\diamonddot}$. It is easy to see that this leads to $p \in \{\neg p\}$, so the condition is not satisfied.

On the other hand let us see that the formula $\phi = p \wedge \diamonddot\{\neg p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{p\}\} \wedge \overline{\diamonddot}\{\neg p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{p\}\}$ does belong to $\overline{\mathcal{F}}_2$. We will see shortly that this formula is satisfiable.

**Proposition 19.** *Let* $\mathfrak{M} = (W, R, V)$ *be a model and let* $w \in W$ *be a world in* $\mathfrak{M}$*. Then, for any* $n \geq 0$*, there exists a unique minterm* $\phi \in \overline{\mathcal{F}}_n$ *such that* $\mathfrak{M}, w \Vdash \phi$*.*

*Proof.* The uniqueness follows from the choice of the minterms, as no two can be satisfied together. For the existence we proceed by induction on $n$. For the base case, $n = 0$, we have that $\mathcal{M}, w \Vdash pr(\{p \in \mathcal{P}rop \mid w \in V(p)\})$.

For the induction step, we use the result for $i < n$ and prove it for $n$. We know that for every world $v \in W$ there exists a unique $\phi_v \in \overline{\mathcal{F}}_{n-1}$ that holds at $v$. Then, the formula $\phi = pr(\{p \in \mathcal{P}rop \mid w \in V(p)\}) \wedge \diamonddot\{\phi_v \mid v \in W, (w, v) \in R\} \wedge \overline{\diamonddot}\{\phi_v \mid v \in W, (v, w) \in R\}$ is satisfied at $w$. We must show that $\phi \in \overline{\mathcal{F}}_n$.

Let $v$ be an arbitrary world accessible from $w$. This implies $\phi_v \in \phi/\diamonddot$. Since $\phi$ holds at $w$, $\overline{\rho}(\overline{\rho}(\phi))$ also holds at $w$. We also know from Proposition 18 that $\overline{\rho}(\overline{\rho}(\phi))$ is in $\overline{\mathcal{F}}$, more precisely in $\overline{\mathcal{F}}_{n-2}$. But then, according to the way we construct the satisfied minterms, it must be that $\overline{\rho}(\overline{\rho}(\phi)) \in \phi_v/\overline{\diamonddot}$. Similarly, it can be shown that $\overline{\rho}(\overline{\rho}(\phi)) \in \phi_v/\diamonddot$ for all $\phi_v \in \phi/\overline{\diamonddot}$. Thus, we have $\phi \in \overline{\mathcal{F}}$. This concludes our proof. ☐

**Definition 67.** *Let $n$ be an integer. The modal minterm of degree $n$ that is satisfied at a world $w$ in a model $\mathfrak{M}$ is denoted by $\overline{\sigma}_n(w)$:*

$$\overline{\sigma}_n(w) \in \overline{\mathcal{F}}_n \textit{ and } \mathfrak{M}, w \Vdash \overline{\sigma}_n(w) \tag{4.45}$$

*Such a minterm exists and is unique according to Proposition 19.*

Recall that in Definition 66 we have called the elements of $\overline{\mathcal{F}}_n$ satisfiable minterms. We are now ready to justify this name.

**Proposition 20.** *For every satisfiable modal minterm of degree $n$, $\phi \in \overline{\mathcal{F}}_n$, there exists a model $\mathfrak{M} = (W, R, V)$ and a world $w \in W$ such that $\mathfrak{M}, w \Vdash \phi$.*

*Proof.* We use $\overline{\mathcal{M}}(\phi)$ to refer to the pointed model $(\mathfrak{M}, w)$ that satisfies $\phi$. Our proof provides an algorithm for recursively constructing this model.

We proceed by induction on the modal degree of the minterms. For the base case, we consider $n = 0$, i.e. the minterms from $\overline{\mathcal{F}}_0$. For every minterm $pr(\Phi) \in \overline{\mathcal{F}}_0$, with $\Phi \subseteq \mathcal{P}rop$, we can build the model $\mathfrak{M}_\Phi = (\{w\}, \varnothing, V_\Phi)$, where $V_\Phi(p) = \{w\}$ if $p \in \Phi$ and $V_\Phi(p) = \varnothing$ otherwise. Clearly we have $\mathfrak{M}_\Phi, w \Vdash pr(\Phi)$.

For the induction step, we assume that the claim holds for $i < n$ and we prove it for $n$. Thus, given $\phi = pr(\Phi_0) \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2$, we use the models $\overline{\mathcal{M}}(\psi) = ((W_\psi, R_\psi, V_\psi), w_\psi)$, for $\psi \in \phi/\Diamonddot \cup \phi/\overline{\Diamonddot}$. We also use the model $\overline{\mathcal{M}}(pr(\Phi_0)) = ((W_0, R_0, V_0), w_0)$ We construct the model for $\phi$ as in the proof of Proposition 15, i.e. $\mathfrak{M} = (W, R, V)$ given by:

$$\begin{aligned} W &= W_0 \cup \bigcup_{\psi \in \phi/\Diamonddot} W_\psi \cup \bigcup_{\psi \in \phi/\overline{\Diamonddot}} W_\psi \\ R &= R_0 \cup \bigcup_{\psi \in \phi/\Diamonddot} (R_\psi \cup \{(w_0, w_\psi)\}) \cup \bigcup_{\psi \in \phi/\overline{\Diamonddot}} (R_\psi \cup \{(w_\psi, w_0)\}) \\ V(p) &= V_0(p) \cup \bigcup_{\psi \in \phi/\Diamonddot} V_\psi(p) \cup \bigcup_{\psi \in \phi/\overline{\Diamonddot}} V_\psi(p) \end{aligned} \tag{4.46}$$

Let us see that, for all $\psi \in \phi/\Diamonddot$, we have $\mathfrak{M}, w_\psi \Vdash \psi$. The only thing that might violate this is the addition of $(w_0, w_\psi)$ to the accessibility relation. However, this is not a problem, since $\overline{\rho}(\overline{\rho}(\phi)) \in \psi/\overline{\Diamonddot}$ based on Definition 66. Similarly, $\mathfrak{M}, w_\psi \Vdash \psi$ for all $\psi \in \phi/\overline{\Diamonddot}$. Thus, we can conclude that $\mathfrak{M}, w_0 \Vdash \phi$, which ends our proof. □

As promised, we will now show that the formula $\phi = p \wedge \Diamonddot\{\neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}\} \wedge \overline{\Diamonddot}\{\neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}\}$ is satisfiable, by constructing a model for it, according to the proof of Proposition 20. The final model is presented in Figure 4.4. We will not go into all the details, but only discuss the last step of the process. We will use $\mathfrak{M}{\downarrow}_S$ to refer to the restriction of $\mathfrak{M}$ to the worlds in $S$. We have that $\mathfrak{M}{\downarrow}_{\{w_1, w_2, w_3\}}, w_2 \Vdash \neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}$. The key observation is that the link between $w_2$ and $w_4$ does not change the 1-st degree modal minterm satisfied at $w_2$. This is a direct consequence of the fact that our formula satisfies $\overline{\rho}(\overline{\rho}(\phi)) = p \in \neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}/\Diamonddot$. So we also have $\mathfrak{M}, w_2 \Vdash \neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}$. Similarly we have $\mathfrak{M}, w_6 \Vdash \neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}$. Coupled with $\mathfrak{M}, w_4 \Vdash p$, we reach the desired conclusion that $\mathfrak{M}, w_4 \Vdash \phi$.

Let us also emphasize the distinction between the pointed models that we have constructed for the basic modal language, which were in fact rooted trees, and

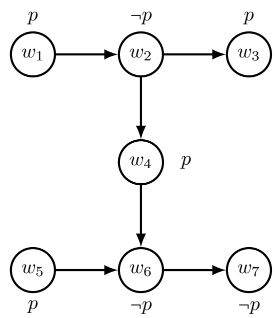



**Figure 4.4:** *Example for the satisfaction of the modal minterms of* $ML(\Diamond, \overline{\Diamond})$.

the models we build here. These models are no longer trees because of the in edges that we add in order to satisfy the converse diamond terms. However, these models can be interpreted as trees if the direction of the arrows is ignored, so they are a special kind of directed acyclic graphs.

Constructing models that satisfy a given formula is important for us because it will be a crucial point in the most important results of Chapter 5. We will explore this aspect in more detail, for $ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$, in the next section.

### 4.4.2 Translation to normal form

We continue by formalizing the disjunctive normal form for the converse modal language.

**Definition 68.** *A modal formula* $\phi \in ML(\Diamond, \overline{\Diamond})$ *is in* ***disjunctive normal form*** *(DNF) iff it is a (possibly degenerate) disjunction of distinct satisfiable modal minterms of the same degree. We use* $ML(\Diamond, \overline{\Diamond})_{DNF}$ *to refer to the corresponding language.*

$$ML(\Diamond, \overline{\Diamond})_{DNF} = \{\bigvee \Phi \mid \exists n (n \geq 0 \wedge \Phi \subseteq \overline{\mathcal{F}}_n)\} \tag{4.47}$$

Now we discuss the translation of arbitrary modal formulas $\phi \in ML(\Diamond, \overline{\Diamond})$ to an equivalent disjunctive normal form. The procedure is similar to that used for the DNF of $ML(\Diamond)$. We denote the translation with $\overline{tr}_n$. We will use the translation for formulas of modal depth $\geq n$. We assume that the formulas are in negation normal form. To simplify notation, we introduce the following additional abbreviation: for $\phi = pr(\Phi_0)$ or $\phi = pr(\Phi_0) \wedge \diamonddot\Phi_1 \wedge \overline{\diamonddot}\Phi_2$, we will use $\phi/pr$ to refer to $\Phi_0$. For atomic formulas we have the following translations.

$$\begin{aligned} \overline{tr}_n(\bot) &= \bigvee \varnothing \\ \overline{tr}_n(\top) &= \bigvee \overline{\mathcal{F}}_n \\ \overline{tr}_n(p) &= \bigvee \{\phi \in \overline{\mathcal{F}}_n \mid p \in \phi/pr\} \\ \overline{tr}_n(\neg p) &= \bigvee \{\phi \in \overline{\mathcal{F}}_n \mid p \notin \phi/pr\} \end{aligned} \tag{4.48}$$

For disjunction and conjunction, the translation is the same as for the basic modal language. If we have $\overline{tr}_n(\phi) = \bigvee \Phi$ and $\overline{tr}_n(\psi) = \bigvee \Psi$ then we can write:

$$\begin{aligned}\overline{tr}_n(\phi \vee \psi) &= \bigvee(\Phi \cup \Psi) \\ \overline{tr}_n(\phi \wedge \psi) &= \bigvee(\Phi \cap \Psi)\end{aligned} \tag{4.49}$$

For the translation of the modal operators we assume that we know $\overline{tr}_{n-1}(\phi) = \bigvee \Phi$. We have:

$$\begin{aligned}\overline{tr}_n(\Diamond\phi) &= \bigvee\{\psi \in \overline{\mathcal{F}}_n \mid \psi/\,\diamonddot \cap \Phi \neq \varnothing\} \\ \overline{tr}_n(\Box\phi) &= \bigvee\{\psi \in \overline{\mathcal{F}}_n \mid \psi/\diamonddot \subseteq \Phi\} \\ \overline{tr}_n(\overline{\Diamond}\phi) &= \bigvee\{\psi \in \overline{\mathcal{F}}_n \mid \psi/\overline{\diamonddot} \cap \Phi \neq \varnothing\} \\ \overline{tr}_n(\overline{\Box}\phi) &= \bigvee\{\psi \in \overline{\mathcal{F}}_n \mid \psi/\overline{\diamonddot} \subseteq \Phi\}\end{aligned} \tag{4.50}$$

**Definition 69.** *For any formula $\phi \in ML(\Diamond, \overline{\Diamond})$ and for any $n \geq md(\phi)$, we define the translation of $\phi$ to n-th degree disjunctive normal form as:*

$$\overline{tr}_{DNF,n}(\phi) = \overline{tr}_n(tr_{NNF}(\phi)) \tag{4.51}$$

*where $\overline{tr}_n$ is the translation function defined by the relations (4.48), (4.49) and (4.50).*

**Proposition 21.** *For every formula $\phi \in ML(\Diamond, \overline{\Diamond})$ and for any $n \geq md(\phi)$, it holds that $\overline{tr}_{DNF,n}(\phi) \in ML(\Diamond, \overline{\Diamond})_{DNF}$ and $\phi \leftrightarrow \overline{tr}_{DNF,n}(\phi)$ is a validity.*

*Proof.* The fact that the translated formula is in disjunctive normal form follows directly from the construction. For the equivalence, we need to check that $\overline{tr}_n$ preserves modal satisfaction in all cases. For the translations presented in (4.48) and (4.49) the equivalence between the formula and its translation follows from Proposition 19 and Proposition 20. For the translations from equation (4.50), we can use the following validity

$$\begin{aligned}&\bigwedge_{\phi\in\Phi_1} \Diamond\phi \wedge \Box(\bigvee_{\phi\in\Phi_2} \phi) \wedge \bigwedge_{\phi\in\Phi_1'} \overline{\Diamond}\phi \wedge \overline{\Box}(\bigvee_{\phi\in\Phi_2'} \phi) \\ \leftrightarrow &\bigvee\{\diamonddot\Phi \wedge \overline{\diamonddot}\Phi' \mid \Phi_1 \subseteq \Phi \subseteq \Phi_2, \Phi_1' \subseteq \Phi' \subseteq \Phi_2'\}\end{aligned} \tag{4.52}$$

which is similar to the one we have used in the proof of Proposition 16. □

As an example, we consider the translation of the formula $\phi = p \wedge \Diamond p \wedge \overline{\Diamond}\neg p$.

We assume $\mathcal{P}rop = \{p\}$. We have:

$$
\begin{aligned}
\overline{tr}_{DNF,0}(p) &= \bigvee\{p\} \\
\overline{tr}_{DNF,0}(\neg p) &= \bigvee\{\neg p\} \\
\overline{tr}_{DNF,1}(p) &= \bigvee\{\phi \in \overline{\mathcal{F}}_1 \mid p \in \phi/pr\} \\
\overline{tr}_{DNF,1}(\Diamond p) &= \bigvee\{\phi \in \overline{\mathcal{F}}_1 \mid p \in \phi/\Diamonddot\} \\
\overline{tr}_{DNF,1}(\overline{\Diamond}\neg p) &= \bigvee\{\phi \in \overline{\mathcal{F}}_1 \mid \neg p \in \phi/\overline{\Diamonddot}\} \\
\overline{tr}_{DNF,1}(p \wedge \Diamond p \wedge \overline{\Diamond}\neg p) &= \bigvee\{\phi \in \overline{\mathcal{F}}_1 \mid p \in \phi/pr, p \in \phi/\Diamonddot, \neg p \in \phi/\overline{\Diamonddot}\} \\
&= \bigvee\{p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{\neg p\}, p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p, \neg p\}, \\
&\qquad p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\{\neg p\}, p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\{p, \neg p\}\}
\end{aligned}
\tag{4.53}
$$

Written without the use of abbreviations, this normal form becomes:

$$
\begin{aligned}
\overline{tr}_{DNF,1}(p \wedge \Diamond p \wedge \overline{\Diamond}\neg p) = {} & p \wedge \Diamond p \wedge \Box p \wedge \overline{\Diamond}\neg p \wedge \overline{\Box}\neg p \\
& \vee p \wedge \Diamond p \wedge \Box p \wedge \overline{\Diamond} p \wedge \overline{\Diamond}\neg p \wedge \overline{\Box}(p \vee \neg p) \\
& \vee p \wedge \Diamond p \wedge \Diamond\neg p \wedge \Box(p \vee \neg p) \wedge \overline{\Diamond}\neg p \wedge \overline{\Box}\neg p \\
& \vee p \wedge \Diamond p \wedge \Diamond\neg p \wedge \Box(p \vee \neg p) \wedge \overline{\Diamond} p \wedge \overline{\Diamond}\neg p \wedge \overline{\Box}(p \vee \neg p)
\end{aligned}
\tag{4.54}
$$

## 4.5 Normal form for $ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$

In this section we discuss the satisfiability of modal formulas that use the global modalities ($ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$) with the use of normal forms.

### 4.5.1 Extracted normal form

In this subsection we discuss a disjunctive normal form related to the global modalities. This normal form is more similar to NNF than to the DNF's that we have seen for $ML(\Diamond)$ and $ML(\Diamond, \overline{\Diamond})$, in the sense that it does not rely on minterms and a very strict form for the formulas, but just provides definite requirements for the global modalities. We start from the approach presented in (Areces and Gorin, 2010) for extracting the global modalities from within the scope of other modalities, then use some simple transformations to reach what we shall call the global normal form (GNF). The most important results of this section are related to the satisfiability of global formulas.

The satisfaction of the global modalities does not depend on the actual world where the evaluation is performed. This means that a formula $\mathbf{A}\phi$ or $\mathbf{E}\phi$ that is a part of a larger formula can be evaluated just once, globally, then the result can be used for deciding the satisfaction of the larger formula. This intuition is captured in the following validities taken from (Areces and Gorin, 2010):

$$
\begin{aligned}
\phi(\mathbf{E}\psi) &\leftrightarrow (\mathbf{E}\psi \to \phi(\mathbf{E}\psi/\top)) \wedge (\neg\mathbf{E}\psi \to \phi(\mathbf{E}\psi/\bot)) \\
\phi(\mathbf{A}\psi) &\leftrightarrow (\mathbf{A}\psi \to \phi(\mathbf{A}\psi/\top)) \wedge (\neg\mathbf{A}\psi \to \phi(\mathbf{A}\psi/\bot))
\end{aligned}
\tag{4.55}
$$

where the notation $\phi(\alpha)$ means that $\alpha$ is a subformula of $\phi$ (and may even occur more than once in $\phi$) and $\varphi(\alpha/\beta)$ stands for the formula obtained by replacing all occurrences of $\alpha$ with $\beta$. The intuition is that, for example, when $\mathbf{E}\psi$ is evaluated for a model and is true, the evaluation of any formula that contains $\mathbf{E}\psi$ can use $\top$ instead of all occurrences of $\mathbf{E}\psi$.

Intuitively, the repeated application of (4.55) should enable us to bring all the global modalities from a formula at top level (with respect to other modalities, but not necessarily with respect to negation). What we want is actually even more: we want all global modalities at top level (so not in the scope of negations either) and we want the formulas to be also in negation normal form. We start by defining this kind of formulas as extracted normal forms.

**Definition 70.** *A global modal formula is in **extracted normal form** (ENF) iff it is defined by the following BNF:*

$$\begin{aligned} \phi &::= \psi \mid \mathbf{E}\psi \mid \mathbf{A}\psi \mid \phi \wedge \phi \mid \phi \vee \phi \\ \psi &::= \top \mid \bot \mid p \mid \neg p \mid \psi \wedge \psi \mid \psi \vee \psi \mid \Diamond\psi \mid \Box\psi \end{aligned} \tag{4.56}$$

*where $p$ ranges over a set of propositional symbols $\mathcal{P}rop$. We will denote the corresponding language with $ML(\Diamond, \mathbf{E})_{ENF}$.*

**Proposition 22.** *There exists a translation function $tr_{ENF} : ML(\Diamond, \mathbf{E})_{NNF} \to ML(\Diamond, \mathbf{E})_{ENF}$ such that, for every formula $\phi \in ML(\Diamond, \mathbf{E})_{NNF}$ it holds that $\phi \leftrightarrow tr_{ENF}(\phi)$ is valid.*

*Proof.* First, we provide an equivalent formulation of (4.55) such that the negation normal form is preserved.

$$\begin{aligned} \varphi(\mathbf{E}\psi) &\leftrightarrow \mathbf{E}\psi \wedge \varphi(\mathbf{E}\psi/\top) \vee \mathbf{A}tr_{NNF}(\neg\psi) \wedge \varphi(\mathbf{E}\psi/\bot) \\ \varphi(\mathbf{A}\psi) &\leftrightarrow \mathbf{A}\psi \wedge \varphi(\mathbf{A}\psi/\top) \vee \mathbf{E}tr_{NNF}(\neg\psi) \wedge \varphi(\mathbf{A}\psi/\bot) \end{aligned} \tag{4.57}$$

It is clear in 4.57 that whenever the left hand side is in negation normal form, so will be the right hand side. Now we proceed to the proof of the proposition, by induction on the number of global modalities that are under the scope of other unary operators.

For the base case, when there is no such global modality, the formula is already in extracted normal form, so there is nothing to prove, we simply take $tr_{ENF}(\phi) = \phi$.

For the induction step, we assume that all formulas that have $i < n$ global modalities that are not at top level can be translated to ENF and we prove the claim for $n$. Of the $n$ global modalities that are not at top level, we pick one such that the subformula it encloses does not contain another global modality. We can always make such a choice, since the formulas are finite. Suppose that $\mathbf{E}\psi$ is the corresponding subformula. Then we can use the first relation from (4.57) and write the translation as:

$$tr_{ENF}(\varphi(\mathbf{E}\psi)) = \mathbf{E}\psi \wedge tr_{ENF}(\varphi(\mathbf{E}\psi/\top)) \vee \mathbf{A}tr_{NNF}(\neg\psi) \wedge tr_{ENF}(\varphi(\mathbf{E}\psi/\bot)) \tag{4.58}$$

where the values for $tr_{ENF}$ on the right hand side can be computed using the induction hypothesis, since the number of global modalities that are not at top level must have been reduced by at least one after the substitution of $\mathbf{E}\psi$ with $\top$ or $\bot$. If the chosen global modality is $\mathbf{A}$, then we use the second relation from (4.57) and define the translation as:

$$tr_{ENF}(\varphi(\mathbf{A}\psi)) = \mathbf{A}\psi \wedge tr_{ENF}(\varphi(\mathbf{A}\psi/\top)) \vee \mathbf{E}tr_{NNF}(\neg\psi) \wedge tr_{ENF}(\varphi(\mathbf{A}\psi/\bot)) \tag{4.59}$$

□

As an example, we translate the formula $\phi = \Box p \wedge \mathbf{A}(\Diamond p \vee \mathbf{E}\neg p)$ to an equivalent extracted normal form. We have:

$$\begin{aligned} &tr_{ENF}(\phi) = tr_{ENF}(\Box p \wedge \mathbf{A}(\Diamond p \vee \mathbf{E}\neg p)) \\ &= \mathbf{E}\neg p \wedge tr_{ENF}(\Box p \wedge \mathbf{A}(\Diamond p \vee \top)) \vee \mathbf{A}tr_{NNF}(\neg\neg p) \wedge tr_{ENF}(\Box p \wedge \mathbf{A}(\Diamond p \vee \bot)) \\ &= \mathbf{E}\neg p \wedge \Box p \wedge \mathbf{A}(\Diamond p \vee \top) \vee \mathbf{A}p \wedge \Box p \wedge \mathbf{A}(\Diamond p \vee \bot) \end{aligned} \tag{4.60}$$

### 4.5.2 Global normal form

Let us see that in (4.60) we can also use the commutativity and distributivity properties for conjunction and disjunction in order to rearrange and group terms. Furthermore, we can exploit the validity $\mathbf{A}(\phi \wedge \psi) \leftrightarrow \mathbf{A}\phi \wedge \mathbf{A}\psi$.

$$\begin{aligned} \phi &\leftrightarrow \mathbf{E}\neg p \wedge \Box p \wedge \mathbf{A}(\Diamond p \vee \top) \vee \mathbf{A}p \wedge \Box p \wedge \mathbf{A}(\Diamond p \vee \bot) \\ &\leftrightarrow \Box p \wedge \mathbf{E}\neg p \wedge \mathbf{A}(\Diamond p \vee \top) \vee \Box p \wedge \mathbf{A}(p \wedge (\Diamond p \vee \bot)) \end{aligned} \tag{4.61}$$

This leads us to the global normal form, which is in fact a kind of disjunctive normal form that only imposes restrictions with respect to the global modalities.

**Definition 71.** *A formula from $ML(\Diamond,\mathbf{E})$ is said to be in* ***global normal form*** *(GNF) iff it is defined by the following BNF:*

$$\begin{aligned} \phi &::= \psi \wedge \mathbf{A}\psi \mid \psi \wedge \xi \wedge \mathbf{A}\psi \mid \phi \vee \phi \\ \xi &::= \mathbf{E}\psi \mid \mathbf{E}\psi \vee \xi \\ \psi &::= \top \mid \bot \mid p \mid \neg p \mid \psi \wedge \psi \mid \psi \vee \psi \mid \Diamond\psi \mid \Box\psi \end{aligned} \tag{4.62}$$

*Where $p$ ranges over the set of propositional symbols $\mathcal{P}rop$. We will denote the corresponding language with $ML(\Diamond,\mathbf{E})_{GNF}$.*

In fact, the BNF from Definition 71 describes formulas that can be written as:

$$\phi = \bigvee_{i=1}^{n} \left( \phi_{i0} \wedge \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi'_i \right) \tag{4.63}$$

where the formulas $\phi_{ij}$ and $\phi'_i$ are from $ML(\Diamond)$ for all $i$ from 1 to $n$ and for all $j$ from 0 to $n_i$. Note that it is possible to have $n_i = 0$ for some values of $i$. In such cases, the corresponding conjunction vanishes from the formula.

**Proposition 23.** *There exists a translation function* $tr_{E2G} : ML(\Diamond, \mathbf{E})_{ENF} \to ML(\Diamond, \mathbf{E})_{GNF}$ *such that, for all* $\phi \in ML(\Diamond, \mathbf{E})_{ENF}$ *it holds that* $\phi \leftrightarrow tr_{E2G}(\phi)$ *is valid.*

*Proof.* We proceed by induction using the BNF of the extracted normal form. For the base case, we have the following translations, each of them corresponding to a validity, for any $\psi \in ML(\Diamond)$:

$$\begin{aligned} tr_{E2G}(\psi) &= \psi \wedge \mathbf{A}\top \\ tr_{E2G}(\mathbf{E}\psi) &= \top \wedge \mathbf{E}\psi \wedge \mathbf{A}\top \\ tr_{E2G}(\mathbf{A}\psi) &= \top \wedge \mathbf{A}\psi \end{aligned} \tag{4.64}$$

For the induction step, we need to discuss conjunction and disjunction, assuming that both operands can be translated to their global normal form. For disjunction we have directly:

$$tr_{E2G}(\phi \vee \psi) = tr_{E2G}(\phi) \vee tr_{E2G}(\psi) \tag{4.65}$$

For the conjunction, we need to use commutativity and distributivity properties of conjunction and disjunction and also $\mathbf{A}(\phi \wedge \psi) \leftrightarrow \mathbf{A}\phi \wedge \mathbf{A}\psi$, just as we did for the example translation in (4.61):

$$\begin{aligned} tr_{E2G}(\phi) &= \bigvee_{i=1}^{m} \left( \phi_{i0} \wedge \bigwedge_{k=1}^{m_i} \mathbf{E}\phi_{ik} \wedge \mathbf{A}\phi'_i \right) \\ tr_{E2G}(\psi) &= \bigvee_{j=1}^{n} \left( \psi_{j0} \wedge \bigwedge_{k=1}^{n_j} \mathbf{E}\psi_{jk} \wedge \mathbf{A}\psi'_j \right) \\ \Rightarrow tr_{E2G}(\phi \wedge \psi) &= \bigvee_{\substack{1 \le i \le m \\ 1 \le j \le n}} \left( (\phi_{i0} \wedge \psi_{j0}) \wedge (\bigwedge_{k=1}^{m_i} \mathbf{E}\phi_{ik} \wedge \bigwedge_{k=1}^{n_j} \mathbf{E}\psi_{jk}) \wedge \mathbf{A}(\phi'_i \wedge \psi'_j) \right) \end{aligned} \tag{4.66}$$

□

We are now ready to define the translation of arbitrary global modal formulas to their equivalent GNF.

**Definition 72.** *For any global modal formula* $\phi \in ML(\Diamond, \mathbf{E})$*, the translation to its equivalent normal form is given by:*

$$tr_{GNF}(\phi) = tr_{E2G}(tr_{ENF}(tr_{NNF}(\phi))) \tag{4.67}$$

**Proposition 24.** *For any global modal formula* $\phi \in ML(\Diamond, \mathbf{E})$ *it holds that* $\phi \leftrightarrow tr_{GNF}(\phi)$ *is valid.*

*Proof.* The result follows directly from the similar results proved for $tr_{NNF}$, $tr_{ENF}$ and $tr_{E2G}$. □

Although we have discussed the global normal form only for $ML(\Diamond, \mathbf{E})$, note that extending the previous results to $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ is fairly straightforward, as all that needs to be done is to change the BNF's of both ENF and GNF to include the converse modalities as well. In fact, the global normal form can be defined for other modal languages as well.

### 4.5.3 Satisfaction of $ML(\Diamond,\mathbf{E})$ formulas

In what follows we are going to discuss the satisfiability of global formulas. While the fact that $ML(\Diamond,\mathbf{E})$ has the finite model property is known in the literature and can be proved via filtrations (Blackburn et al., 2001), here we are interested in the actual construction of models for global formulas, as we will need such models in order to show how modal logic relates to argumentation principles.

Our aim is to provide an algorithm for checking the satisfiability of a global formula $\phi$ syntactically (using normal forms) and also, whenever $\phi$ is satisfiable, to construct a model for it. To the best of our knowledge, such an approach is not available in the literature. We will first consider $ML(\Diamond,\mathbf{E})$.

**Lemma 1.** *Let $\phi \in ML(\Diamond,\mathbf{E})_{GNF}$ be a global modal formula in global normal form, i.e.*

$$\phi = \bigvee_{i=1}^{n}\left(\phi_{i0} \wedge \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi_i'\right) \tag{4.68}$$

*Then $\phi$ is satisfiable iff there exists an $i$ such that $\phi_{i0}\wedge\bigwedge_{j=1}^{n_i}\mathbf{E}\phi_{ij}\wedge\mathbf{A}\phi_i'$ is satisfiable.*

Lemma 1 is quite straightforward, as it follows directly from the modal satisfaction of disjunction. Its role is to simplify the formulas we need to deal with for deciding satisfiability.

**Lemma 2.** *Let $\phi = \phi_0\wedge\bigwedge_{i=1}^{n}\mathbf{E}\phi_i\wedge\mathbf{A}\phi' \in ML(\Diamond,\mathbf{E})$ be a global modal formula such that $\phi_i$ and $\phi'$ are from $ML(\Diamond)$. Then $\phi$ is satisfiable iff $\phi_i \wedge \mathbf{A}\phi'$ is satisfiable for any $i$ from 0 to $n$.*

*Proof.* For the direct implication, suppose $\phi$ is satisfiable. Then there exists a pointed model $(\mathfrak{M}, w)$ that satisfies it. But then, based on the definition of satisfiability for the existential modality, for any $i$ from 0 to $n$ there must exist a world $w_i \in W$ such that $\mathfrak{M}, w_i \Vdash \phi_i$. From this and from the satisfaction of $\phi$, we can easily deduce that $\mathfrak{M}, w_i \Vdash \mathbf{A}\phi'$. Thus, we also have $\mathfrak{M}, w_i \Vdash \phi_i \wedge \mathbf{A}\phi'$.

For the converse, we assume that for every $i$ from 0 to $n$ we can construct a pointed model $(\mathfrak{M}_i, w_i)$ that satisfies $\phi_i \wedge \mathbf{A}\phi'$. Then the disjoint union of these models satisfies $\phi$ at $w_0$:

$$\biguplus_{i=0}^{n}\mathfrak{M}_i, w_0 \Vdash \phi_0 \wedge \bigwedge_{i=1}^{n}\mathbf{E}\phi_i \wedge \mathbf{A}\phi' \tag{4.69}$$

□

Lemma 2 is important because, coupled with Lemma 1, it allows us to reduce the satisfiability problem for global modal formulas (in global normal form) to the satisfiability of $\phi \wedge \mathbf{A}\psi$ formulas, where $\phi, \psi \in ML(\Diamond)$. We can simplify the problem further by considering that $\phi$ is in disjunctive normal form.

**Lemma 3.** *Let $\phi$ and $\psi$ be two basic modal language formulas. If $\phi$ is in disjunctive normal form, i.e. $\phi = \bigvee\Phi$ with $\Phi \subseteq \mathcal{F}_{md(\phi)}$, then $\phi\wedge\mathbf{A}\psi$ is satisfiable iff there exists $\phi' \in \Phi$ such that $\phi' \wedge \mathbf{A}\psi$ is satisfiable.*

The result presented in Lemma 3 follows directly from the modal satisfaction of disjunction. Based on it, we can now focus only on formulas $\phi \wedge \mathbf{A}\psi$ where $\phi$ is a modal minterm. If $\psi$ is a disjunctive normal form of the same degree as $\phi$, i.e. $\phi \in \mathcal{F}_n$ and $\psi = \bigvee \Psi$ with $\Psi \subseteq \mathcal{F}_n$, then it is easy to see that a necessary condition for the satisfiability of $\phi \wedge \mathbf{A}\psi$ is that $\phi \in \Psi$. On the other hand, this condition is in general not sufficient. Indeed, consider the following example: $\phi = p \wedge \diamond\{\neg p\}$ and $\psi = p \wedge \diamond\{p\} \vee p \wedge \diamond\{\neg p\}$. Then $\phi \wedge \mathbf{A}\psi$ cannot be satisfied, because no world from a model that satisfies $\mathbf{A}\psi$ can satisfy $\neg p$, which means that $\phi$ cannot hold.

Furthermore, note that the formula $\mathbf{A}(p \wedge \diamond\{\neg p\})$ is not satisfiable. So the modal minterms of the basic modal language are not always satisfiable if within the scope of the universal modality. This happens because the satisfaction of a given modal minterm requires the existence of accessible worlds in the model such that other formulas are satisfied. If those formulas are not "compatible" to what needs to be satisfied globally, then the modal minterm cannot hold either.

Note that, at least in theory, if we have a modal formula $\phi \wedge \mathbf{A}\psi$, where $\phi$ and $\psi$ are arbirary basic modal language formulas, we can convert them to equivalent disjunctive normal forms of the same degree $n$ (the maximum between $md(\phi)$ and $md(\psi)$), i.e. $tr_{DNF,n}(\phi) = \bigvee \Phi$ and $tr_{DNF,n}(\psi) = \bigvee \Psi$. Then, using Lemma 3, the satisfiability of $\phi \wedge \mathbf{A}\psi$ can be decided by considering $\phi' \wedge \mathbf{A}(\bigvee \Psi)$, for all $\phi' \in \Phi$, with the advantage that $\phi'$ and $\bigvee \Psi$ have the same modal degree and we can use at least the previously mentioned necessary condition for satisfiability.

However, such an approach may be highly impractical because of the exponential blowup in the number of minterms with respect to the modal degree. Thus, we will prefer to use normal forms of minimal degree possible, especially for the formula that is within the scope of the universal modality. To this end, we will make use of the result from Lemma 4.

**Definition 73.** *Let $\phi \in \mathcal{F}_n$ be a modal minterm of degree $n$. The **restriction** of $\phi$ to a modal minterm of degree $i \leq n$ (notation $\tau_i(\phi)$) is defined as:*

$$\begin{aligned} \tau_0(\phi) &= \phi/pr \\ \tau_{i+1}(\phi) &= \phi/pr \wedge \diamond\{\tau_i(\alpha) \mid \alpha \in \phi/\diamond\} \end{aligned} \tag{4.70}$$

*Equivalently, we can define the restriction as $\tau_i(\phi) = \rho^{(n-i)}(\phi)$, where $\rho^{(k)}$ denotes the repeated application of reduction $k$ times (we assume that $\rho$ is the straightforward adaptation of the reduction function $\overline{\rho}$, introduced in Definition 65, to work with basic modal formulas).*

**Lemma 4.** *Let $\phi = pr(\Phi_0) \wedge \diamond\Phi_1$ be a modal minterm of degree $m$ and let $\psi = \bigvee \Psi$ be a disjunctive normal form of degree $n$, with $m \geq n$. Then $\phi \wedge \mathbf{A}\psi$ is satisfiable iff $\tau_n(\phi) \in \Psi$ and $\alpha \wedge \mathbf{A}\psi$ is satisfiable for all $\alpha \in \Phi_1$.*

*Proof.* For the direct implication, suppose that there exists a pointed model $(\mathfrak{M}, w)$ that satisfies $\phi \wedge \mathbf{A}\psi$. Then let us first see that the modal minterm of degree $n$ that is satisfied at $w$ is $\sigma_n(w) = \tau_n(\phi)$. Since $\psi$ must hold at all worlds, we deduce that $\tau_n(\phi) \in \Psi$. Furthermore, for every $\alpha \in \Phi_1$ there must be a world $w_\alpha$ in $\mathfrak{M}$, accessible from $w$, such that $\mathfrak{M}, w_\alpha \Vdash \alpha$. Since $\psi$ is satisfied globally, we get that $\mathfrak{M}, w_\alpha \Vdash \alpha \wedge \mathbf{A}\psi$.

For the converse, suppose that for every $\alpha \in \Phi_1$ we have a pointed model $(\mathfrak{M}_\alpha, w_\alpha)$ that satisfies $\alpha \wedge \mathbf{A}\psi$, where $\mathfrak{M}_\alpha = (W_\alpha, R_\alpha, V_\alpha)$. We also consider the model $(\mathfrak{M}_0, w_0)$ that satisfies $pr(\Phi_0)$, with $\mathfrak{M}_0 = (W_0, R_0, V_0)$. We construct the model $\mathfrak{M} = (W, R, V)$ given by:

$$\begin{aligned} W &= W_0 \cup \bigcup_{\alpha \in \Phi_1} W_\alpha \\ R &= R_0 \cup \bigcup_{\alpha \in \Phi_1} (R_\alpha \cup \{(w_0, w_\alpha)\}) \\ V(p) &= V_0(p) \cup \bigcup_{\alpha \in \Phi_1} V_\alpha(p), \quad \textit{for all } p \in \mathcal{P}rop \end{aligned} \tag{4.71}$$

Note that this is the same construction that we have used for the models of the basic modal language in the proof of Proposition 15. Thus, it is not surprising that $\mathfrak{M}, w_0 \Vdash \phi$. Let us see that we also have $\mathfrak{M}, w_0 \Vdash \psi$. Indeed, the modal minterm of degree $n$ that holds at $w_0$ is $\sigma_n(w_0) = \tau_n(\phi)$, which we know is in $\Psi$. Using this and the fact that $\mathfrak{M}_\alpha, w_\alpha \Vdash \mathbf{A}\psi$ for all $\alpha \in \Phi_1$, we can conclude that $\mathfrak{M}, w_0 \Vdash \phi \wedge \mathbf{A}\psi$, which is the desired result. □

Let us discuss the implications of Lemma 4. Suppose we have a modal minterm $\phi \in \mathcal{F}_m$ and a disjunctive normal form $\psi = \bigvee \Psi$, with $\Psi \subseteq \mathcal{F}_n$. If $m \geq n$, the lemma allows us to reduce the satisfiability of $\phi \wedge \mathbf{A}\psi$ to the satisfiability of $\alpha \wedge \mathbf{A}\psi$, where $\alpha$ are modal minterms of degree $m-1$. If $m-1 \geq n$, we can apply the lemma again. If we do this recursively, we eventually reduce the problem to the satisfiability of $\alpha \wedge \mathbf{A}\psi$ where $\alpha$ are modal minterms of degree $n-1$. Furthermore, if $m < n$, we can say that $\phi \wedge \mathbf{A}\psi$ is satisfiable iff there exists a modal minterm $\beta$, of degree $n-1$ such that $\beta \wedge \mathbf{A}\psi$ is satisfiable and $\phi = \tau_m(\beta)$. Thus, we can always reduce the satisfiability problem to the case $m = n-1$.

Let $\psi = \bigvee \Psi$, with $\Psi \subseteq \mathcal{F}_n$, be a disjunctive normal form and $\phi \in \mathcal{F}_n$ a modal minterm. We have seen that $\phi \wedge \mathbf{A}\psi$ can be satisfiable only if $\phi \in \Psi$. On the other hand, let us note that, if $\phi = pr(\Phi_0) \wedge \diamond\Phi_1$, we must also have $\alpha \wedge \mathbf{A}\psi$ satisfiable for all $\alpha \in \Phi_1$, according to Lemma 4. But in order for this to hold, there must exist a modal minterm $\beta$, of degree $n$, such that $\beta \wedge \mathbf{A}\psi$ is satisfiable and $\tau_{n-1}(\beta) = \rho(\beta) = \alpha$. So this $\beta$ should be an element of $\Psi$ as well.

Based on this intuition, we define the following function: $sat(\Psi) = \{\phi \in \Psi \mid \forall\alpha(\alpha \in \phi/\diamond \Rightarrow \exists\beta(\beta \in \Psi \texttt{ and } \rho(\beta) = \alpha))\}$. This function filters out the modal minterms that cannot be satisfied because some of the formulas they require to be satisfied at accessible worlds cannot be provided by the minterms from $\Psi$. Let us consider the following example: $\Psi = \{p \wedge \diamond\{p \wedge \diamond\{p\}\}, p \wedge \diamond\{\neg p \wedge \diamond\{\neg p\}\}, \neg p \wedge \diamond\{\neg p \wedge \diamond\{p\}\}\}$. Let us see that $\neg p \wedge \diamond\{\neg p \wedge \diamond\{p\}\}$ cannot be satisfied at the same time as $\mathbf{A}(\bigvee \Psi)$, since none of the minterms from $\Psi$ reduces to the required $\neg p \wedge \diamond\{p\}$. The other two minterms do satisfy the requirement, so we can write $\Psi_1 = sat(\Psi) = \{p \wedge \diamond\{p \wedge \diamond\{p\}\}, p \wedge \diamond\{\neg p \wedge \diamond\{\neg p\}\}\}$. Using the same reasoning we get $\Psi_2 = sat(\Psi_1) = \{p \wedge \diamond\{p \wedge \diamond\{p\}\}\}$ and also $\Psi_{i+1} = sat(\Psi_i) = \{p \wedge \diamond\{p \wedge \diamond\{p\}\}\}$ for all $i \geq 2$.

Since in the general case we have $sat(\Psi) \subseteq \Psi$ by definition and also we only work on finite sets $\Psi$, repeatedly applying the $sat$ function eventually leads to a stationary value. We will refer to the elements of the corresponding set as minterms that are possibly satisfiable together.

**Definition 74.** *Let $\Psi \subseteq \mathcal{F}_n$ be a set of modal minterms. The set of* ***minterms possibly satisfiable together*** *(notation $sat^*(\Psi)$) is defined as the stationary value of the sequence $(\Psi_n)_{n\geq 0}$, where $\Psi_0 = \Psi$ and, for $n \geq 0$, $\Psi_{n+1} = sat(\Psi_n)$ with sat is given by:*

$$sat(\Psi) = \{\phi \in \Psi \mid \forall\alpha(\alpha \in \phi/\diamond \Rightarrow \exists\beta(\beta \in \Psi \wedge \rho(\beta) = \alpha))\} \tag{4.72}$$

*Furthermore, we define $sat^*_{-1}(\Psi) = \{\rho(\psi) \mid \psi \in sat^*(\Psi)\}$, the set of modal minterms of degree $n-1$ that are possibly satisfiable with $\bigvee\Psi$.*

In other words, the set $sat^*_{-1}(\Psi)$ contains all the modal minterms $\phi$ of degree $md(\bigvee\Psi)-1$ for which $\phi \wedge \mathbf{A}(\bigvee\Psi)$ cannot be proved unsatisfiable by the repeated application of Lemma 4. The good news is that all these minterms are in fact satisfiable and we can also construct a model that satisfies all of them (at different worlds).

**Lemma 5.** *Let $\Psi \subseteq \mathcal{F}_n$ be a set of modal minterms of degree $n$ and let $\phi \in \mathcal{F}_{n-1}$. Then $\phi \wedge \mathbf{A}(\bigvee\Psi)$ is satisfiable iff $\phi \in sat^*_{-1}(\Psi)$.*

*Proof.* We start with the direct implication. If $\phi \wedge \mathbf{A}(\bigvee\Psi)$ is satisfiable, i.e. there exists a pointed model $(\mathfrak{M}, w)$ that satisfies it, then, for the $n$-th degree modal minterm satisfied at $w$ – $\sigma_n(w)$ – we have that $\mathfrak{M}, w \Vdash \sigma_n(w) \wedge \mathbf{A}(\bigvee\Psi)$. Thus, using the definitions of $sat$ and $sat^*$, it must be that $\sigma_n(w) \in sat^*(\Psi)$. Furthermore, since $\rho(\sigma_n(w)) = \phi$, we can deduce that $\phi \in sat^*_{-1}(\Phi)$.

Next, we prove the converse. For each minterm $\alpha \in sat^*_{-1}(\Psi)$, we choose one minterm $\phi_\alpha \in sat^*(\Psi)$ such that $\rho(\phi_\alpha) = \alpha$. Then we construct the model $\mathfrak{M} = (W, R, V)$ given by:

$$\begin{aligned} &W = \{w_\alpha \mid \alpha \in sat^*_{-1}(\Psi)\} \\ &R = \{(w_\alpha, w_\beta) \in W \times W \mid \beta \in \phi_\alpha/\diamond\} \\ &V(p) = \{w_\alpha \in W \mid p \in \phi_\alpha/pr\}, \quad \textit{for all } p \in \mathcal{P}rop \end{aligned} \tag{4.73}$$

We can now prove by induction on $i$ that, for every $\alpha \in sat^*_{-1}(\Psi)$, $\sigma_i(w_\alpha) = \tau_i(\phi_\alpha)$ and, thus, $\mathfrak{M}, w_\alpha \Vdash \alpha$. For the base case, $i = 0$, we need to show that $\sigma_0(w_\alpha) = \phi_\alpha/pr$, which follows from the choice of the valuation function $V$.

For the induction step we have $\sigma_{i+1}(w_\alpha) = \sigma_0(w_\alpha) \wedge \diamond\{\sigma_i(w_\beta) \mid (w_\alpha, w_\beta) \in R\} = \phi_\alpha/pr \wedge \diamond\{\tau_i(\phi_\beta) \mid \beta \in \phi_\alpha/\diamond\} = \tau_{i+1}(\phi_\alpha)$, which is the desired result. This concludes our proof. ☐

Note that in general it is possible to construct several models for the minterms $\alpha \in sat^*_{-1}(\Psi)$, based on the choice of each $\phi_\alpha$. This may be important in practical applications, as it allows us to look for models that satisfy certain additional properties. We will discuss this aspect in the next section. We are now ready to put everything together and provide one of the important results of this chapter.

**Theorem 3.** *Let $\phi \in ML(\Diamond, \mathbf{E})$ be a global modal formula. The satisfiability of $\phi$ can be verified syntactically and, whenever $\phi$ is satisfiable, a model that satisfies it can be constructed recursively.*

*Proof.* The claim follows from lemmas 1, 2, 3 and 5. The algorithm proceeds as follows:

`1. convert` $\phi$ `to global normal form:`

$$\phi = \bigvee_{i=1}^{n} \left( \phi_{i0} \wedge \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi'_i \right) \tag{4.74}$$

`2. for each` $i$ `from` $1$ `to` $n$ `do`
`2.1. consider the formula` $\phi_i = \phi_{i0} \wedge \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi'_i$
`2.2. for each` $j$ `from` $0$ `to` $n_i$ `do`
`2.2.1. consider the formula` $\phi_{ij} \wedge \mathbf{A}\phi'_i$
`2.2.2. consider` $m = max(md(\phi_{ij}), md(\phi'_i) - 1) + 1$
`2.2.3. convert` $\phi'_i$ `to a disjunctive normal form of degree` $m$`:`

$$tr_{DNF,m}(\phi'_i) = \bigvee \Phi' \tag{4.75}$$

`2.2.4. compute` $sat^*_{-1}(\Phi')$
`2.2.5. convert` $\phi_{ij}$ `to disjunctive normal form of degree` $m{-}1$`:`

$$tr_{DNF,m-1}(\phi_{ij}) = \bigvee \Phi \tag{4.76}$$

`2.2.6. if` $\Phi \cap sat^*_{-1}(\Phi') = \varnothing$`, then` $\phi_{ij} \wedge \mathbf{A}\phi'_i$ `is not satisfiable`
`2.2.7. else` $\phi_{ij} \wedge \mathbf{A}\phi'_i$ `is satisfiable + use Lemma 5 to construct the pointed model` $(\mathfrak{M}_{ij}, w_{ij})$ `such that`

$$\mathfrak{M}_{ij}, w_{ij} \Vdash \phi_{ij} \wedge \mathbf{A}\phi'_i \tag{4.77}$$

`2.3. if there exists` $j$ `such that` $\phi_{ij} \wedge \mathbf{A}\phi'_i$ `is not satisfiable, then` $\phi_i$ `is not satisfiable`
`2.4. else` $\phi_i$ `is satisfiable + use the pointed models` $(\mathfrak{M}_{ij}, w_{ij})$ `and Lemma 2 to construct the pointed model` $(\mathfrak{M}_i, w_i)$ `such that`

$$\mathfrak{M}_i, w_i \Vdash \phi_i \tag{4.78}$$

`3. if there exists` $i$ `such that` $\phi_i$ `is satisfiable, then the formula` $\phi$ `is satisfiable and` $\mathfrak{M}_i, w_i \Vdash \phi$
`4. else` $\phi$ `is not satisfiable`

□

Note that the algorithm provided in the proof does not make use of Lemma 4. We have presented the algorithm like this because it is simpler and from a theoretical point of view there is no problem. However, in practice, as mentioned previously, the lemma is quite useful, as it allows us to use a normal form of minimal degree for the formula that is within the scope of the universal modality.

### 4.5.4 Satisfaction of $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ formulas

We will now discuss the translation of the satisfiability results from $ML(\Diamond, \mathbf{E})$ to $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. Lemmas 1, 2 and 3 can be reused simply by replacing the modal language, as the converse modality has no impact on the results. Lemma 4 relies on the model construction provided in the proof of Proposition 20. For Lemma 5, on the other hand, we need to provide a different proof. We start by defining the corresponding notion of restriction of a modal minterm.

**Definition 75.** *Let $\phi \in \overline{\mathcal{F}}_n$ be a modal minterm of degree $n$. The* ***restriction*** *of $\phi$ to a modal minterm of degree $i \le n$ (notation $\overline{\tau}_i(\phi)$) is defined as:*

$$\begin{aligned} &\overline{\tau}_0(\phi) = \phi/pr \\ &\overline{\tau}_{i+1}(\phi) = \phi/pr \wedge \diamond\{\overline{\tau}_i(\alpha) \mid \alpha \in \phi/\diamond\} \wedge \overline{\diamond}\{\overline{\tau}_i(\alpha) \mid \alpha \in \phi/\overline{\diamond}\} \end{aligned} \tag{4.79}$$

*Equivalently, we can define the restriction as $\overline{\tau}_i(\phi) = \overline{\rho}^{(n-i)}(\phi)$, where $\overline{\rho}^{(k)}$ denotes the repeated application of reduction $k$ times.*

We also need to define the counterpart of the $sat$ function for $ML(\diamond, \overline{\diamond}, \mathbf{E})$. The $\overline{sat}$ function aims to remove from a set of minterms $\Psi \subseteq \overline{\mathcal{F}}_n$ the minterms $\phi$ for which we can prove that $\phi \wedge \mathbf{A}(\bigvee \Psi)$ is not satisfiable. Suppose that $\phi \wedge \mathbf{A}(\bigvee \Psi)$ is satisfiable, so there exists a pointed model $(\mathfrak{M}, w)$ that satisfies it. Then for all $\alpha \in \phi/\diamond$ there exists a world $w_\alpha$, with $(w, w_\alpha) \in R$, such that $\mathfrak{M}, w_\alpha \Vdash \alpha \wedge \mathbf{A}(\bigvee \Psi)$. Furthermore, we also have $\mathfrak{M}, w_\alpha \Vdash \overline{\sigma}_n(w_\alpha) \wedge \mathbf{A}(\bigvee \Psi)$, so it must be that $\overline{\sigma}_n(w_\alpha) \in \Psi$. Also we have $\overline{\rho}(\overline{\sigma}_n(w_\alpha)) = \alpha$. So far, this is the same as for $ML(\diamond, \mathbf{E})$. Let us see that in this case there is another property that must hold, namely $\rho(\phi) \in \overline{\sigma}_n(w_\alpha)/\overline{\diamond}$. A similar condition must hold for the elements of $\phi/\overline{\diamond}$.

**Definition 76.** *Let $\Psi \subseteq \overline{\mathcal{F}}_n$ be a set of modal minterms. The set of* ***minterms possibly satisfiable together*** *(notation $\overline{sat}^*(\Psi)$) is defined as the stationary value of the sequence $(\Psi_n)_{n \ge 0}$, where $\Psi_0 = \Psi$ and, for $n \ge 0$, $\Psi_{n+1} = \overline{sat}(\Psi_n)$, with $\overline{sat}$ is given by:*

$$\begin{aligned} \overline{sat}(\Psi) = \{\phi \in \Psi \mid &\forall \alpha(\alpha \in \phi/\diamond \Rightarrow \exists \beta(\beta \in \Psi \textbf{ and } \rho(\beta) = \alpha \textbf{ and } \rho(\phi) \in \beta/\overline{\diamond})), \\ &\forall \alpha(\alpha \in \phi/\overline{\diamond} \Rightarrow \exists \beta(\beta \in \Psi \textbf{ and } \rho(\beta) = \alpha \textbf{ and } \rho(\phi) \in \beta/\diamond))\} \end{aligned} \tag{4.80}$$

*Furthermore, we define $\overline{sat}^*_{-1}(\Psi) = \{\rho(\psi) \mid \psi \in \overline{sat}^*(\Psi)\}$, the set of modal minterms of degree $n-1$ that are possibly satisfiable with $\bigvee \Psi$.*

We are now ready to prove the counterpart of Lemma 5 for $ML(\diamond, \overline{\diamond}, \mathbf{E})$.

**Lemma 6.** *Let $\Psi \subseteq \overline{\mathcal{F}}_n$ be a set of modal minterms of degree $n$ and let $\phi \in \overline{\mathcal{F}}_{n-1}$. Then $\phi \wedge \mathbf{A}(\bigvee \Psi)$ is satisfiable iff $\phi \in \overline{sat}^*_{-1}(\Psi)$.*

*Proof.* The proof of the direct implication is virtually identical to the one provided for Lemma 5. We will focus on the converse. We construct the model $\mathfrak{M} = (W, R, V)$ given by:

$$\begin{aligned} &W = \{w_\phi \mid \phi \in \overline{sat}^*(\Psi)\} \\ &R = \{(w_\phi, w_\psi) \in W \times W \mid \overline{\rho}(\psi) \in \phi/\diamond \text{ and } \overline{\rho}(\phi) \in \psi/\overline{\diamond}\} \\ &V(p) = \{w_\phi \in W \mid p \in \phi/pr\} \end{aligned} \tag{4.81}$$

We can now prove by induction on $i$ that, for every $\phi \in \overline{sat}^*(\Psi)$, $\overline{\sigma}_i(w_\phi) = \overline{\tau}_i(\phi)$ and, thus, $\mathfrak{M}, w_\phi \Vdash \overline{\rho}(\phi)$ (for $i = n-1$). For the base case, $i = 0$, we need to show that $\overline{\sigma}_0(w_\phi) = \phi/pr$, which follows from the choice of the valuation function $V$.

For the induction step, let us consider an arbitrary $\phi \in \overline{sat}^*(\Psi)$. Then, for every $\alpha \in \phi/\diamondcdot$ the definition of $\overline{sat}^*$ ensures that there exists $\psi \in \overline{sat}^*(\Psi)$ such that $\overline{\rho}(\psi) = \alpha$ and $\overline{\rho}(\phi) \in \psi/\overline{\diamondcdot}$. But then it means that we also have $(w_\phi, w_\psi) \in R$. A similar result holds for $\alpha \in \phi/\overline{\diamondcdot}$. Based on the defintion of $R$ and on that of $\overline{\tau}$, we can deduce that $\overline{\sigma}_{i+1}(w_\phi) = \overline{\sigma}_0(w_\phi) \wedge \diamondcdot\{\overline{\sigma}_i(w_\psi) \mid (w_\phi, w_\psi) \in R\} \wedge \overline{\diamondcdot}\{\overline{\sigma}_i(w_\psi) \mid (w_\psi, w_\phi) \in R\} = \phi/pr \wedge \diamondcdot\{\overline{\tau}_i(\psi) \mid \overline{\rho}(\psi) \in \phi/\diamondcdot\} \wedge \overline{\diamondcdot}\{\overline{\tau}_i(\psi) \mid \overline{\rho}(\psi) \in \phi/\overline{\diamondcdot}\} = \tau_{i+1}(\phi_\alpha)$, which is the desired result. This concludes our proof. □

Note that in practice it may be possible to construct a model that satisfies the requirements of the proof using just a subset of $\Psi$.

**Theorem 4.** *Let $\phi \in ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ be a global modal formula. The satisfiability of $\phi$ can be verified sintactically and, if $\phi$ is satisfiable, a model that satisfies it can be constructed recursively.*

*Proof.* The algorithm is the same as the one presented in the proof of Theorem 3, but using Lemma 6 instead of Lemma 5 and by adapting lemmas 1, 2 and 3 to work with the converse modal language. □

We will see examples for the use of theorems 3 and 4 in the following chapter, where we link modal logic to principles of argumentation semantics.

As a final remark, note that the models that we construct are not always trees and not always acyclic, as the dependencies between minterms may require cycles. Indeed, this is also in agreement with the fact that some satisfiable global modal formulas can have no acyclic model. Consider $\phi = \mathbf{A}(p \wedge \Diamond p)$. There is a very simple finite model that satisfies it: $\mathfrak{M} = (\{w\}, \{(w,w)\}, V)$, with $V(p) = \{w\}$. But the model is not a tree and, in fact, no finite tree model would satisfy $\Diamond p$ at its leaves.

## 4.6 Chapter Summary

In this chapter we discussed normal forms for several modal languages combining basic, converse and global modalities. We were also interested in the satisfaction of these normal forms and the actual construction of models that satisfy given formulas.

In Section 4.1 we introduced the modal languages and the notion of modal satisfaction. We have also discussed bisimulations, not used in this chapter but relevant for Chapter 5. Furthermore, we introduced several notations that helped us write modal formulas in a more compact form.

In Section 4.2 we discussed the negation normal form, which is quite often used in the modal logic literature. This form has helped us write the translations corresponding to the other normal forms more easily.

In sections 4.3 and 4.4 we presented the disjunctive normal forms for $ML(\Diamond)$ and $Ml(\Diamond, \overline{\Diamond})$. The approach is based on (Fine, 1975), but an equivalent form was used for the minterms, so as to allow a more intuitive and compact representation. Special emphasis was put on the satisfiability of the minterms. In the case of $ML(\Diamond, \overline{\Diamond})$ we had to provide constraints for the normal form so that the obtained minterms are indeed satisfiable.

In Section 4.5 we discussed the satisfiability of $ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ formulas. The most important result of this chapter is the algorithm for deciding the satisfiability of such formulas based on a syntactical analysis that can also lead to the construction of a model that satisfies the given formula. This algorithm will be a key part in the proof of important results that link modal logic and the evaluation principles of argumentation semantics, results that are presented in Chapter 5.

*Chapter 5*

# Argumentation in Modal Logic

Modal logic can easily capture some of the most important argumentation concepts, such as attack and defense. It has been used in (Grossi, 2010) for describing $\mathcal{CF}$, $\mathcal{AS}$, $\mathcal{CO}$ and $\mathcal{ST}$. This link established between argumentation and modal logic allows the transfer of techniques and results between the two domains. Thus, it is interesting to investigate how far this connection goes. In this chapter we show that the modal definability of argumentation semantics is strongly related to some of their properties and that many semantics fall beyond the expressiveness of simpler modal logics. In order to reach these conclusions, we will use results from Chapter 4.

We start by discussing the use of modal logic as a meta-language for talking about argumentation concepts, in Section 5.1. We then discuss the normal form of formulas that can describe argumentation semantics in Section 5.2. The main results are presented in Section 5.3 and Section 5.4, where we analyze the use of $ML(\Diamond, \mathbf{E})$, respectively $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$, for describing argumentation semantics. The chapter ends with a summary of the results in Section 5.5.

## 5.1 Modal Logic and Argumentation

This section focuses on the use of modal logic for argumentation, based on (Grossi, 2010). We start with the remark that, at least from a superficial point of view, argumentation frameworks and Kripke models are quite similar, in the sense that they both rely on a directed graph and give some interpretation to it. For Kripke models, we have sets of worlds that satisfy propositional symbols, while for argumentation frameworks we have extensions satisfying certain properties.

What is distinctive for argumentation semantics is that their properties generally refer to the attackers of an argument so, in a sense, the converse of the attack relation is used. This observation is implicitly used in (Grossi, 2010) for defining the modal satisfiability of $\Diamond$ and $\Box$. While the results are essentially the same, our approach is slightly different, in the sense that we do not alter modal satisfiability to fit the intrinsic nature of argumentation semantics, but we evaluate the extensions in a distinct framework based on the model.

**Definition 77.** *Let $\mathcal{S}$ be an argumentation semantics and let $ML$ be some modal language. We define the argumentation-based operator $\mathcal{S}$, such that, for any model $\mathfrak{M} = (W, R, V)$, any world $w \in W$ and any formula $\phi \in ML$, we have:*

$$\mathfrak{M}, w \Vdash \mathcal{S}\phi \Leftrightarrow V^*(\phi) \in \mathcal{E}_{\mathcal{S}}(W, R^-) \tag{5.1}$$

*Where $V^*$ is the extended valuation function and $R^-$ stands for the converse of $R$: $R^- \triangleq \{(y,x) \mid (x,y) \in R\}$.*

In words, the modal operator based on the argumentation semantics $\mathcal{S}$ holds when the valuation of $\phi$ is an extension of the argumentation framework that uses the converse of the accessibility relation as an attack relation. The challenge is to see whether this operator can be equivalently formulated within a given modal language. Proposition 25 lists the main results from (Grossi, 2010).

**Proposition 25.** *The following argumentation-based modalities can be captured by global modal formulas:*

$$\begin{aligned} &\Vdash \mathcal{CF}\phi \leftrightarrow \mathbf{A}(\phi \to \neg\Diamond\phi) \\ &\Vdash \mathcal{AS}\phi \leftrightarrow \mathbf{A}(\phi \to \neg\Diamond\phi \wedge \Box\Diamond\phi) \\ &\Vdash \mathcal{CO}\phi \leftrightarrow \mathbf{A}((\phi \to \neg\Diamond\phi) \wedge (\phi \leftrightarrow \Box\Diamond\phi)) \\ &\Vdash \mathcal{ST}\phi \leftrightarrow \mathbf{A}(\phi \leftrightarrow \neg\Diamond\phi) \end{aligned} \tag{5.2}$$

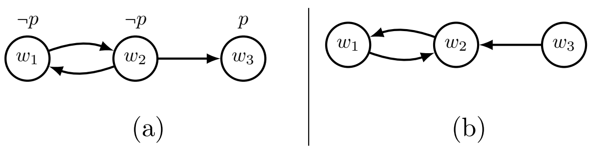



**Figure 5.1:** *Example for the satisfaction of argumentation-based modalities: (a) Kripke model $\mathfrak{M} = (W, R, V)$; (b) corresponding argumentation framework $F = (W, R^-)$.*

Instead of providing a proof for the equivalences from Proposition 25, we show that reading a modal formula from an argumentation perspective can lead directly to the definition of the corresponding semantics. The satisfaction of a formula $\phi$ at $w$ means that $w$ is an element of $V^*(\phi)$. Thus, the formula $\mathbf{A}(\phi \to \neg\Diamond\phi)$ can be read as: for any argument $a$, if $a$ is in $V^*(\phi)$, then there exists no attacker of $a$ that is in $V^*(\phi)$, which is a reformulation of the definition of conflict-free sets. The other formulas can be read similarly.

Let us see how this works for the example from Figure 5.1. We will show that $\mathfrak{M}, w_1 \Vdash \mathcal{CF}(p \vee \Diamond\Diamond p)$. Indeed, we have $V^*(p \vee \Diamond\Diamond p) = \{w_1, w_3\}$ and $\{w_1, w_3\} \in \mathcal{E}_{\mathcal{CF}}(F)$. Also, let us see that the equivalence from Proposition 25 does indeed hold. For this, we should prove that $\mathfrak{M}, w_1 \Vdash \mathbf{A}(p \vee \Diamond\Diamond p \to \neg\Diamond(p \vee \Diamond\Diamond p))$. Equivalently, we must show that $\mathfrak{M}, w \Vdash p \vee \Diamond\Diamond p \to \neg\Diamond(p \vee \Diamond\Diamond p)$, for all $w \in W$.

First, let us see that $w_1$ and $w_2$ satisfy the desired formula because they do not satisfy $p$ or $\Diamond\Diamond p$. On the other hand, $w_3$ satisfies $p \vee \Diamond\Diamond p$, so we must check that it satisfies $\neg\Diamond(p \vee \Diamond\Diamond p)$ as well, so that we obtain the satisfaction of the implication.

Since $w_3$ has no successors, it satisfies $\neg\Diamond(p \vee \Diamond\Diamond p)$ trivially, thus we can conclude that the equivalent formulation of $\mathcal{CF}$ in $ML(\Diamond, \mathbf{E})$ is indeed satisfied.

A characterization of the grounded semantics within $ML(\Diamond, \mathbf{E})$ is not provided in (Grossi, 2010). Instead, a $\mu$-calculus (Bradfield and Stirling, 2007) formula is given. Using our notations and convention for argumentation-based modalities, this formula can be written as:

$$\Vdash \mathcal{GR}\phi \leftrightarrow \mathbf{A}(\phi \leftrightarrow \mu Z.\Box\Diamond Z) \tag{5.3}$$

where $\mu$ is the minimal fixpoint operator, which brings second-order expressiveness to modal logic. We do not go into more details about $\mu$-calculus, as this logic is not the main focus or our approach.

As far as the preferred semantics is concerned, a proof based on total bisimulation is provided in (Grossi, 2010) for showing that $ML(\Diamond, \mathbf{E})$ is not expressive enough to capture $\mathcal{PR}$. We extend this result in (Gratie et al., 2012a) by showing that full hybrid $\mu$-calculus (Sattler and Vardi, 2001) is also unable to describe the preferred semantics. Since full hybrid $\mu$-calculus combines hybrid logics (Areces and ten Cate, 2007), $\mu$-calculus, the global modalities and the converse modalities, we can easily deduce that $\mathcal{PR}$ cannot be described within $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ either.

$ML(\Diamond, \mathbf{E})$ is extended in (Grossi, 2011) with monadic second order quantification. As stated in the paper, the purpose is not to find the appropriate expressive power for describing each semantics in isolation, but to accommodate them all at once. The resulting logic is shown to be the binary fragment of Monadic Second Order Logic (MSOL), which is more expressive than $\mu$-calculus. The advantage of using this logic is the fact that its model checking game can be applied to argumentation. In his paper, Grossi provides descriptions for $\mathcal{GR}$, $\mathcal{PR}$ and $\mathcal{SST}$. The use of MSOL for argumentation is also discussed in (Dvorak et al., 2012), where descriptions of $\mathcal{STA}$, $\mathcal{CF}2$ and $\mathcal{GR}^*$ are provided.

Our goal, on the other hand, is to find the right expressive power that is required for capturing each argumentation semantics. Our detailed analysis is limited to $ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$, but the methods and results that we provide can be reused for other logics as well. Furthermore, we strengthen the link between argumentation and modal logic by relating modal formulas to evaluation principles satisfied by argumentation semantics.

While the works presented so far are most closely related to our approach, there are several other ideas worth mentioning. For example the use of modal logic with argument labelings is discussed in (Caminada and Gabbay, 2009). The modal language used there has a fixed set of three propositional symbols, corresponding to the possible labels of an argument. Furthermore, an approach based on provability logic is discussed in (Gabbay, 2009a).

## 5.2 Formulas That Describe Semantics

In this section we aim to characterize modal formulas that can capture the operators that are based on argumentation semantics. More precisely, we show that the additivity principle, which is satisfied by almost all semantics, has a modal counterpart which can be related to syntactical properties of modal formulas.

We recall that the modal operator corresponding to an argumentation semantics is given by:

$$\mathfrak{M}, w \Vdash \mathcal{S}\phi \Leftrightarrow V^*(\phi) \in \mathcal{E}_{\mathcal{S}}(W, R^-) \tag{5.4}$$

If the converse modalities are available in the language, we can either use 5.4 as well, or adjust the definition to use $R$ directly, instead of its converse:

$$\mathfrak{M}, w \Vdash \overline{\mathcal{S}}\phi \Leftrightarrow V^*(\phi) \in \mathcal{E}_{\mathcal{S}}(W, R) \tag{5.5}$$

It is not very important which of the two is used when the converse is also available since, whenever the argumentation-based operator $\overline{\mathcal{S}}$ can be captured by a $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ formula, one can interchange all occurrences of $\Diamond$ and $\Box$ with $\overline{\Diamond}$ and $\overline{\Box}$ respectively to get the formula for $\mathcal{S}$. For the rest of this subsection we assume that the first convention is used.

In both 5.4 and 5.5 it is clear that the actual world $w$ where the satisfaction is tested is not relevant and that the satisfaction will either hold for all the worlds in the model or for none of them. Thus, we can write:

$$\mathfrak{M}, w \Vdash \mathcal{S}\phi \Leftrightarrow \forall w'(w' \in W \Rightarrow \mathfrak{M}, w' \Vdash \mathcal{S}\phi) \tag{5.6}$$

In modal logic, the right hand side amounts to saying that $\mathcal{S}\phi$ is globally true in $\mathfrak{M}$, which is usually denoted as $\mathfrak{M} \Vdash \phi$ (Blackburn et al., 2001). On the other hand, 5.7 also leads to the validity $\mathcal{S}\phi \leftrightarrow \mathbf{A}\mathcal{S}\phi$. We can exploit this observation and characterize the normal form of $\mathcal{S}\phi$ when this formula can be captured by either $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$.

**Theorem 5.** *Any global formula $\mathbf{A}\phi$ from $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ can be translated to the following normal form:*

$$tr'_{GNF}(\mathbf{A}\phi) = \bigvee_{i=1}^{n} \left( \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi'_i \right) \tag{5.7}$$

*which is equivalent to the original formula, i.e. $\mathbf{A}\phi \leftrightarrow tr'_{GNF}(\mathbf{A}\phi)$ is a validity. The formulas $\phi_{ij}$ and $\phi'_i$ are from $ML(\Diamond)$, respectively $ML(\Diamond, \overline{\Diamond})$.*

*Proof.* We rely on the translation to global normal form, as summarized in Proposition 24. Recall that the translation is given by $tr_{GNF} = tr_{E2G} \circ tr_{ENF} \circ tr_{NNF}$. For a formula $\mathbf{A}\phi$, the negation normal form will be $\mathbf{A}tr_{NNF}(\phi)$. Furthermore, the translation to extracted normal form using 4.57 cannot give rise to any subformula not within the scope of a global modality. This means that, in fact, the translation $tr_{ENF}(tr_{NNF}(\mathbf{A}\phi))$ gives rise to a formula defined by the following BNF:

$$\phi ::= \mathbf{A}\psi \mid \mathbf{E}\psi \mid \phi \wedge \phi \mid \phi \vee \phi \tag{5.8}$$

where $\psi$ is in either $ML(\Diamond)$ or $ML(\Diamond, \overline{\Diamond})$. This allows us to alter the definition of $tr_{E2G}$ and write $tr'_{E2G}$ as:

$$\begin{aligned} tr'_{E2G}(\mathbf{A}\psi) &= \mathbf{A}\psi \\ tr'_{E2G}(\mathbf{E}\psi) &= \mathbf{E}\psi \wedge \mathbf{A}\top \end{aligned} \tag{5.9}$$

The definitions for the disjunction and conjunction are the same as the ones for $tr_{E2G}$ presented in (4.65) and (4.66). With this, we can define $tr'_{GNF} = tr'_{E2G} \circ tr_{ENF} \circ tr_{NNF}$. The equivalence between $\mathbf{A}\phi$ and $tr'_{GNF}(\mathbf{A}\phi)$ holds based on Proposition 24 and the fact that the changes we applied to $tr_{E2G}$ preserve the modal satisfaction of formulas. □

**Corollary 1.** *Let $\mathcal{S}$ be an argumentation-based operator such that $\mathcal{S}\phi$ can be captured by $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. Then $\mathcal{S}\phi$ is equivalent to a global normal form without non-global terms:*

$$\mathcal{S}\phi \leftrightarrow \bigvee_{i=1}^{n} \left( \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi'_i \right) \tag{5.10}$$

*Proof.* Follows from Theorem 5 and the observation that $\mathcal{S} \leftrightarrow \mathbf{A}\mathcal{S}$ is valid. □

Next, let us notice that all the global formulas presented from (Grossi, 2010) are of the form $\mathbf{A}\phi$, where $\phi$ is in the basic modal language. With respect to the normal form 5.10, this is a quite simplified formula, as we have $n = 1$ and $n_1 = 0$. Furthermore, the meaning of such a formula is that some local constraint is required from all the worlds in the model.

Furthermore, it is easy to see that whenever two disjoint models satisfy such a formula, their disjoint union also satisfies it, corresponding in some sense to the additivity principle that we have introduced for argumentation semantics. In fact, we shall show that modal formulas (from $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$) that describe argumentation semantics satisfying additivity can be written in this simplified form.

**Definition 78.** *We say that a modal formula $\phi$ is* ***additive*** *(or that it satisfies* ***additivity****) iff, for any two disjoint models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ we have:*

$$\mathfrak{M}_1 \uplus \mathfrak{M}_2 \Vdash \phi \Leftrightarrow \mathfrak{M}_1 \Vdash \phi \textit{ and } \mathfrak{M}_2 \Vdash \phi \tag{5.11}$$

**Proposition 26.** *For any argumentation-based operator $\mathcal{S}$ that corresponds to an argumentation semantics that satisfies additivity, the formula $\mathcal{S}\phi$ is additive.*

This important observation lets us translate the argumentation problem of satisfying the additivity principle to a problem of modal logic. We are now ready to prove the main result of this section.

**Theorem 6.** *Let $\mathbf{A}\phi$ be a formula from $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. Then $\mathbf{A}\phi$ is additive iff it is equivalent to a formula $\mathbf{A}\psi$, where $\psi \in ML(\Diamond)$, respectively $\psi \in ML(\Diamond, \overline{\Diamond})$.*

*Proof.* The $\Leftarrow$ part of the proof is obvious, so we focus on the $\Rightarrow$ part. We start by converting $\mathbf{A}\phi$ to its normal form using Theorem 5:

$$tr'_{GNF}(\mathbf{A}\phi) = \bigvee_{i=1}^{n} \left( \bigwedge_{j=1}^{n_i} \mathbf{E}\phi_{ij} \wedge \mathbf{A}\phi'_i \right) \triangleq \bigvee_{i=1}^{n} \psi_i \tag{5.12}$$

where we have used $\psi_i$ to denote each term of the disjunction. If there exists any $\psi_k$ that is not satisfiable, we can simply remove it from the disjunction and obtain an equivalent shorter formula. Without loss of generality, we assume that we have already done so.

If there exists a $k$ such that $\neg\phi \wedge \mathbf{A}\phi'_k$ is satisfiable, then let $\mathfrak{M}_1$ be a model that satisfies it. Also, let $\mathfrak{M}_2$ be a model that satisfies $\psi_k$ (such a model exists because, according to the above assumption, $\psi_k$ is satisfiable for any $k$). Let $\mathfrak{M} = \mathfrak{M}_1 \uplus \mathfrak{M}_2$ be the disjoint union of the two models. We have that $\mathfrak{M}$ satisfies $\bigwedge_{j=1}^{n_k} \mathbf{E}\phi_{kj}$ because $\mathfrak{M}_2 \Vdash \psi_k$ and the satisfaction of existential formulas is preserved. Furthermore, $\mathfrak{M}$ satisfies $\mathbf{A}\phi'_k$ because both $\mathfrak{M}_1$ and $\mathfrak{M}_2$ satisfy it. Hence $\mathfrak{M}$ satisfies $\phi$. However, $\mathfrak{M}_1$ does not satisfy $\phi$, which contradicts the additivity of $\phi$.

This means that $\neg\phi \wedge \mathbf{A}\phi'_k$ is not satisfiable for any $k$, which leads to the validity $\mathbf{A}\phi'_k \to \phi$, for all $k$. Now, if there exists a $k$ such that $\phi \wedge \neg\mathbf{A}\phi'_k$ is not satisfiable, we get $\phi \leftrightarrow \mathbf{A}\psi_k$, which is the desired result.

If no such $k$ exists, on the other hand, we have that for all $k$ there is a model $\mathfrak{M}_k$ such that $\mathfrak{M}_k \Vdash \phi \wedge \neg\mathbf{A}\phi'_k$. The disjoint union $\biguplus_{k=1}^{n} \mathfrak{M}_k$ does not satisfy $\phi$, however, because for every $k$ there is at least one model, namely $\mathfrak{M}_k$, that does not satisfy the corresponding $\mathbf{A}\psi_k$. This contradicts the additivity of $\phi$. □

**Corollary 2.** *Let $\mathcal{S}$ be an argumentation semantics that satisfies additivity. If $\mathcal{S}\phi$ can be described by a formula from $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$, then it can also be described by a formula $\mathbf{A}\psi$, with $\psi \in ML(\Diamond)$ (respectively $\psi \in ML(\Diamond, \overline{\Diamond})$).*

The meaning of Corollary 2 is that whenever global modal logic can describe an argumentation semantics that is additive, it can also describe it with a simpler formula that imposes only local constraints that must hold for all arguments. This result is quite general because it just relies on additivity, a property satisfied by virtually all argumentation semantics ($\mathcal{EPS}$ is the only one that fails to do so, but it is also one of the semantics that received limited attention in the literature).

## 5.3 Doing Argumentation in $ML(\Diamond, \mathbf{E})$

In this section we discuss the use of $ML(\Diamond, \mathbf{E})$ for describing argumentation semantics. Based on the result from Corollary 2, we will investigate only formulas $\mathbf{A}\phi$, with $\phi \in ML(\Diamond)$. We start with a thorough analysis of first degree formulas in Subsection 5.3.1, then follow a more argumentation-oriented analysis based on attack and defense in Subsection 5.3.2. In Subsection 5.3.3 we provide a general result that shows the limitations of $ML(\Diamond, \mathbf{E})$ with respect to satisfying principles of argumentation semantics. We discuss the modal definability of the semantics that are not covered by the general result in Subsection 5.3.4, where we provide several bisimulation-based proofs. We end the section with a discussion of our results and a general graphical overview of the argumentation semantics that are captured by $ML(\Diamond, \mathbf{E})$, in Subsection 5.3.5.

### 5.3.1 Simple formulas

Based on the results of the previous section, the analysis of modal formulas that describe argumentation semantics can be performed directly on $\mathbf{A}\phi$ formulas, with

$\phi \in ML(\Diamond)$. Recall the definition of argumentation-based operators:

$$\mathfrak{M}, w \Vdash \mathcal{S}\phi \Leftrightarrow V^*(\phi) \in \mathcal{E}_{\mathcal{S}}(W, R^-) \tag{5.13}$$

Let us see that we do not really need to deal with general formulas $\mathcal{S}\phi$, but it is enough to discuss $\mathcal{S}p$, where $p$ is a propositional symbol. Extending the previous relation from propositional symbols to arbitrary formulas follows directly from the definition of modal satisfaction. Thus, for the rest of this chapter we will assume that the modal language is based on $\mathcal{P}rop = \{p\}$.

We make use of the disjunctive normal form. For a start, let us see how many distinct formulas we can write and how many of them are meaningful for argumentation. Recall that the modal minterms of degree 1 are:

$$\begin{aligned}\mathcal{F}_1 = \{&p \wedge \diamond\varnothing, p \wedge \diamond\{p\}, p \wedge \diamond\{\neg p\}, p \wedge \diamond\{p, \neg p\},\\ &\neg p \wedge \diamond\varnothing, \neg p \wedge \diamond\{p\}, \neg p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\{p, \neg p\}\}\end{aligned} \tag{5.14}$$

So there are a total of 8 minterms, leading to $2^8 = 256$ combinations for a formula $\mathbf{A}(\bigvee \Phi)$, with $\Phi \subseteq \mathcal{F}_1$. We will distinguish the minterms that have $p$ in positive form from those that have it negated, as follows:

$$\begin{aligned}\mathcal{F}_1^+ &= \{p \wedge \diamond\varnothing, p \wedge \diamond\{p\}, p \wedge \diamond\{\neg p\}, p \wedge \diamond\{p, \neg p\}\}\\ \mathcal{F}_1^- &= \{\neg p \wedge \diamond\varnothing, \neg p \wedge \diamond\{p\}, \neg p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\{p, \neg p\}\}\end{aligned} \tag{5.15}$$

Now, if $\Phi \subseteq \mathcal{F}_1^+$, then $\mathbf{A}(\bigvee \Phi)$ can only be satisfied for $V(p) = W$, while if $\Phi \subseteq \mathcal{F}_1^-$ the formula can be satisfied only for $V(p) = \varnothing$. Thus, such choices for $\Phi$ cannot actually be used for distinguishing the extensions of an argumentation framework. Thus, we will only examine sets of minterms $\Phi$ such that $\Phi \cap \mathcal{F}_1^+ \neq \varnothing$ and $\Phi \cap \mathcal{F}_1^- \neq \varnothing$, leading to $15 \times 15 = 225$ choices.

Conflict-freeness is one of the most important evaluation principles for argumentation semantics and is satisfied by all semantics proposed so far in the literature. Before discussing its impact on the possible choices of $\Phi$, let us also note that there are some sets of arguments that play an important role in argumentation without being conflict free. These are the isolated and the unattacked sets, which are important for stating the additivity, non-interference and directionality principles. These sets can be described using the following formulas

$$\begin{aligned}&\mathcal{IS}p \leftrightarrow \mathbf{A}(p \wedge \Box p \vee \neg p \wedge \Box\neg p) \leftrightarrow \mathbf{A}(p \wedge \diamond\varnothing \vee p \wedge \diamond\{p\} \vee \neg p \wedge \diamond\varnothing \vee \neg p \wedge \diamond\{\neg p\})\\ &\mathcal{US}p \leftrightarrow \mathbf{A}(p \rightarrow \Box p) \leftrightarrow \mathbf{A}(p \wedge \diamond\varnothing \vee p \wedge \diamond\{p\} \vee \bigvee \mathcal{F}_1^-)\end{aligned} \tag{5.16}$$

where we have also provided the equivalent disjunctive normal forms. In words, the formula for isolated sets states that if an argument $a$ is in an isolated set $S$, all its attackers are in $S$ as well and if $a$ is not in $S$, nor are any of its attackers. For unattacked sets, the formula says that the attackers of the arguments that are included in an unattacked set $S$ should be included in $S$ as well.

Aside from these sets, when we actually talk about semantics, we must ensure that conflict-freeness holds. This means that we do not allow accepted arguments to be attacked by other accepted arguments. In our setting, this means that the modal minterms $p \wedge \diamonddot\{p\}$ and $p \wedge \diamonddot\{p, \neg p\}$ should not be part of $\Phi$.

Furthermore, let us discuss unattacked arguments. If neither $p \wedge \diamonddot\varnothing$, nor $\neg p \wedge \diamonddot\varnothing$ is in $\Phi$, then $\mathbf{A}(\bigvee \Phi)$ cannot hold for any framework that contains unattacked arguments. On the other hand, if we are to include one of the two minterms, the one that makes most sense is the former. Indeed, it makes little sense in argumentation to specifically reject unattacked arguments. Thus, we only need to discuss the following two cases: $\Phi \cap \mathcal{F}_1^+ = \{p \wedge \diamonddot\varnothing\}$ and $\Phi \cap \mathcal{F}_1^+ = \{p \wedge \diamonddot\varnothing, p \wedge \diamonddot\{\neg p\}\}$.

We start with the first one. Let us first see that, if $\Phi \cap \mathcal{F}_1^+ = \{p \wedge \diamonddot\varnothing\}$, then $\mathbf{A}(\bigvee \Phi)$ can hold only if $p$ is true only for unattacked arguments. We can write this as $V(p) \subseteq \mathcal{UA}(W, R^-)$, where we have used $\mathcal{UA}(F)$ to refer to the set of unattacked arguments of the argumentation framework $F$. Based on $\Phi \cap \mathcal{F}_1^-$, the formula will only hold for some sets of unattacked arguments, depending on the properties of the rejected arguments. We distinguish two meaningful such properties for rejected arguments. The first one is to require that all unattacked arguments are accepted, in which case we obtain the description of $\mathcal{UA}$ as:

$$\begin{aligned} \mathcal{UA}p &\leftrightarrow \mathbf{A}(p \wedge \diamonddot\varnothing \vee \neg p \wedge \diamonddot\{p\} \vee \neg p \wedge \diamonddot\{\neg p\} \vee \neg p \wedge \diamonddot\{p, \neg p\}) \\ &\leftrightarrow \mathbf{A}(p \leftrightarrow \Box\bot) \end{aligned} \tag{5.17}$$

The other meaningful distinction that we can make is to require that rejected arguments are not attacked by accepted arguments, leading to the acceptance of sets of isolated arguments only. We will use $\mathcal{IA}(F)$ for the set of isolated arguments from $F$ and $\mathcal{PIA}(F)$ for all the sets consisting of isolated arguments. With these notations, we can write:

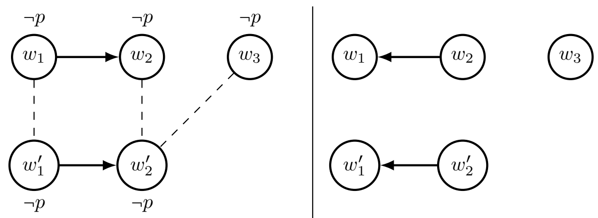
$$\begin{aligned} \mathcal{PIA}p &\leftrightarrow \mathbf{A}(p \wedge \diamonddot\varnothing \vee \neg p \wedge \diamonddot\varnothing \vee \neg p \wedge \diamonddot\{\neg p\}) \\ &\leftrightarrow \mathbf{A}((p \rightarrow \Box\bot) \wedge (\neg p \rightarrow \Box\neg p)) \end{aligned} \tag{5.18}$$

Furthermore, we can leave $\Phi \cap \mathcal{F}_1^- = \mathcal{F}_1^-$ and describe all sets consisting of unattacked arguments, denoted in what follows by $\mathcal{PUA}$:

$$\begin{aligned} \mathcal{PUA}p &\leftrightarrow \mathbf{A}(p \wedge \diamonddot\varnothing \vee \bigvee \mathcal{F}_1^-) \leftrightarrow \mathbf{A}(p \wedge \Box\bot \vee \neg p) \\ &\leftrightarrow \mathbf{A}(p \rightarrow \Box\bot) \end{aligned} \tag{5.19}$$

On the other hand, we will show that $ML(\diamond, \mathbf{E})$ is not expressive enough to capture the actual set of all isolated arguments. Since this is the first bisimulation-based justification from this chapter, let us first discuss the layout and meaning of Figure 5.2, as we shall use the same convention throughout the chapter. On the left hand side there are two models, one at the top and one at the bottom, denoted by $\mathfrak{M} = (W, R, V)$ respectively $\mathfrak{M}' = (W', R', V')$. The two models are total bisimilar and one total bisimulation between them is represented with dashed lines connecting pairs of bisimilar worlds. On the right hand side we show the argumentation framework corresponding to the two models. The frameworks are obtained by replacing the accessibility relation with its converse. The valuation of $p$ in a model will correspond to a set of arguments in the associated framework. The frameworks will be denoted with $F = (W, R^-)$ and $F' = (W', R'^-)$.

When using such bisimulation examples to prove the limitations of the modal language what we look for is a valuation such that the corresponding sets of arguments behave differently with respect to the argumentation semantics considered (one is an extension and the other is not). For the models from Figure 5.2 the valuation of $p$ is the empty set. However, the set of isolated arguments is only empty in $F'$, while in $F$ it is $\{w_3\}$.



(a) models (b) frameworks

**Figure 5.2:** $ML(\Diamond, \mathbf{E})$ *cannot describe the set of all isolated arguments.*

We now move on to the case $\Phi \cap \mathcal{F}_1^+ = \{p \wedge \Diamond\varnothing, p \wedge \Diamond\{\neg p\}\}$. With no constraints on rejected arguments, i.e. $\Phi \cap \mathcal{F}_1^- = \mathcal{F}_1^-$, we get the characterization of conflict-free sets:

$$\begin{aligned} \mathcal{CF}p &\leftrightarrow \mathbf{A}(p \wedge \Diamond\varnothing \vee p \wedge \Diamond\{\neg p\} \vee \bigvee \mathcal{F}_1^-) \\ &\leftrightarrow \mathbf{A}(p \to \Box\neg p) \leftrightarrow \mathbf{A}(p \to \neg\Diamond p) \end{aligned} \tag{5.20}$$

We can think of the four minterms from $\mathcal{F}_1^-$ as defining classes of rejected arguments, as presented in Table 5.1. Removing any of them from $\Phi$ means that we discard the conflict-free sets for which the corresponding class of arguments is not empty. From an argumentation perspective, it makes no sense to forbid rejected arguments that are attacked by accepted arguments, so we prefer sets $\Phi$ that contain $\neg p \wedge \Diamond\{p\}$ and $\neg p \wedge \Diamond\{p, \neg p\}$.

| Minterm | Meaning |
|---|---|
| $\neg p \wedge \Diamond\varnothing$ | rejected arguments that are unattacked |
| $\neg p \wedge \Diamond\{p\}$ | rejected arguments attacked by accepted arguments |
| $\neg p \wedge \Diamond\{\neg p\}$ | rejected arguments attacked by rejected arguments |
| $\neg p \wedge \Diamond\{p, \neg p\}$ | rejected arguments attacked by both attacked and rejected arguments |

**Table 5.1:** *Intuitive meaning of the negative minterms of modal degree* 1

On the other hand we can remove one or both of the remaining minterms. If we remove just $\neg p \wedge \Diamond\varnothing$, we get conflict-free sets that contain all the unattacked arguments. If we remove $\neg p \wedge \Diamond\{\neg p\}$, we get conflict free sets such that all rejected arguments are either unattacked or they are attacked by an accepted argument (they cannot be attached just by other rejected arguments). Such sets are not particularly interesting for argumentation. If we remove both minterms on the other hand, we get the characterization of the stable extensions:

$$\begin{aligned} \mathcal{ST}p &\leftrightarrow \mathbf{A}(p \wedge \diamondplus \varnothing \vee p \wedge \diamondplus\{p\} \vee \neg p \wedge \diamondplus\{p\} \vee \neg p \wedge \diamondplus\{p, \neg p\}) \\ &\leftrightarrow \mathbf{A}(p \leftrightarrow \Box\neg p) \leftrightarrow \mathbf{A}(p \leftrightarrow \neg\Diamond p) \end{aligned} \tag{5.21}$$

To sum up, we have found several simple formulas that describe relevant sets of arguments that are important in argumentation, whether as the extensions of an argumentation semantics or as sets that help describe desirable properties of semantics. We have summarized our findings in Table 5.2. The formulas for conflict-free sets and for stable extensions are already known from (Grossi, 2010), but here they are presented as part of a systematic analysis of simple modal formulas from an argumentation perspective.

| Notation | Formula | Description |
|---|---|---|
| $\mathcal{ST}$ | $\mathbf{A}(p \leftrightarrow \neg\Diamond p)$ | stable extensions |
| $\mathcal{CF}$ | $\mathbf{A}(p \rightarrow \neg\Diamond p)$ | conflict-free sets |
| $\mathcal{IS}$ | $\mathbf{A}(p \wedge \Box p \vee \neg p \wedge \Box\neg p)$ | isolated sets |
| $\mathcal{US}$ | $\mathbf{A}(p \rightarrow \Box p)$ | unattacked sets |
| $\mathcal{UA}$ | $\mathbf{A}(p \leftrightarrow \Box\bot)$ | the set of all unattacked arguments |
| $\mathcal{PUA}$ | $\mathbf{A}(p \rightarrow \Box\bot)$ | sets of unattacked arguments |
| $\mathcal{PIA}$ | $\mathbf{A}((p \rightarrow \Box\bot) \wedge (\neg p \rightarrow \Box\neg p))$ | sets of isolated arguments |

**Table 5.2:** *Argumentation concepts that can be captured by $ML(\Diamond, \mathbf{E})$ formulas*

### 5.3.2 Attack and defense

Instead of performing a thorough analysis for $\mathbf{A}\phi$ formulas based on second degree formulas $\phi$, we will refer to the approach presented in (Gratie and Florea, 2010), where we explore formulas in terms of attack and defense. What this means is that we can define the following notations:

$$\begin{aligned} \mathbf{c} &\triangleq p \\ \mathbf{a} &\triangleq \Diamond p \\ \mathbf{d} &\triangleq \Box\Diamond p \end{aligned} \tag{5.22}$$

We will also use $\overline{\mathbf{c}}$, $\overline{\mathbf{a}}$ and $\overline{\mathbf{d}}$ for the negation of these formulas, so that we are closer to the approach in (Gratie and Florea, 2010). At the end of the subsection, we will translate the results back to the modal language $ML(\Diamond, \mathbf{E})$.

This approach is based on the intuition that the simple argumentation semantics (the ones that do not involve maximality, intersections or more complex approaches) are generally based only on attack and defense. As we shall see later, what we get from this approach is all that the global modal language can do for describing argumentation semantics (for remaining semantics we will show that the corresponding argumentation-based modality cannot be captured by $ML(\Diamond, \mathbf{E})$ formulas).

We will use the abbreviations from (5.22) as building blocks for minterms and then see the meaningful formulas that we can build with them. The set of minterms is:

$$\begin{aligned}\mathcal{M} = \{&\mathbf{c}\wedge\mathbf{a}\wedge\mathbf{d}, \mathbf{c}\wedge\mathbf{a}\wedge\overline{\mathbf{d}}, \mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}, \mathbf{c}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}},\\ &\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\mathbf{d}, \overline{\mathbf{c}}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}}, \overline{\mathbf{c}}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}, \overline{\mathbf{a}}\wedge\overline{\mathbf{c}}\wedge\overline{\mathbf{d}}\}\end{aligned} \tag{5.23}$$

For any set $\Phi \subseteq \mathcal{M}$ we consider the formula $\mathbf{A}(\bigvee\Phi)$. We are interested in formulas that are meaningful for argumentation. Just as in the previous subsection, we can split the minterms in two classes, those that refer to accepted arguments (they contain $\mathbf{c}$) and those that refer to rejected arguments (they contain $\overline{\mathbf{c}}$).

$$\begin{aligned}\mathcal{M}^+ &= \{\mathbf{c}\wedge\mathbf{a}\wedge\mathbf{d}, \mathbf{c}\wedge\mathbf{a}\wedge\overline{\mathbf{d}}, \mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}, \mathbf{c}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}}\}\\ \mathcal{M}^- &= \{\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\mathbf{d}, \overline{\mathbf{c}}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}}, \overline{\mathbf{c}}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}, \overline{\mathbf{a}}\wedge\overline{\mathbf{c}}\wedge\overline{\mathbf{d}}\}\end{aligned} \tag{5.24}$$

For meaningful results we must have $\Phi\cap\mathcal{M}^+ \neq \varnothing$ and $\Phi\cap\mathcal{M}^- \neq \varnothing$. Again, we can consider that each minterm refers to a class of arguments and see whether we are interested in keeping or discarding sets of arguments for which the corresponding class is not empty.

For example, it makes little sense from an argumentation perspective to forbid accepted arguments that are not attacked by the chosen set and are also defended by it. Furthermore, based on the conflict-free principle, we cannot accept arguments that are contained in the set and also attacked by it. Thus, for $\Phi \cap \mathcal{M}^+$ we can choose between $\{\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}\}$ and $\{\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}, \mathbf{c}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}}\}$.

If no constraints are imposed on $\Phi \cap \mathcal{M}^-$, then we get the formulas that characterize admissible and conflict-free sets, respectively:

$$\begin{aligned}\mathcal{AS}p &\leftrightarrow \mathbf{A}(\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}\vee\bigvee\mathcal{M}^-) \leftrightarrow \mathbf{A}(\mathbf{c}\to\overline{\mathbf{a}}\wedge\mathbf{d})\\ \mathcal{CF}p &\leftrightarrow \mathbf{A}(\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}\vee\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}}\vee\bigvee\mathcal{M}^-) \leftrightarrow \mathbf{A}(\mathbf{c}\to\overline{\mathbf{a}})\end{aligned} \tag{5.25}$$

Let us now see that in both formulas the minterm $\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\mathbf{d}$ is not satisfiable because $\mathbf{c}\wedge\mathbf{a}$ is not allowed. Indeed, an argument that is both attacked and defended by a set, must be defended through an attack against an argument from the set, which makes the set non-conflict-free. Being unsatisfiable, this minterm can be either kept in the formula or removed, for example based on how the formula can be simplified to a more understandable form. In other words, if we remove this minterm from either disjunction, we obtain equivalent formulas.

Furthermore, arguments that are not contained and are attacked should not be forbidden, so we will keep $\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\overline{\mathbf{d}}$ in $\Phi$. If we focus on admissible sets and look for reducing the corresponding set of minterms to get new semantics, we have three choices:

$$\begin{aligned}\mathcal{CO}p &\leftrightarrow \mathbf{A}(\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}\vee\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\overline{\mathbf{d}}\vee\overline{\mathbf{c}}\wedge\overline{\mathbf{a}}\wedge\overline{\mathbf{d}}) \leftrightarrow \mathbf{A}((\mathbf{c}\to\overline{\mathbf{a}}\wedge\mathbf{d})\wedge(\overline{\mathbf{c}}\to\overline{\mathbf{d}}))\\ \mathcal{ST}p &\leftrightarrow \mathbf{A}(\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}\vee\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\overline{\mathbf{d}}) \leftrightarrow \mathbf{A}(\mathbf{c}\leftrightarrow\overline{\mathbf{a}}\wedge\mathbf{d})\\ \mathcal{AS}^{da}p &\triangleq \mathbf{A}(\mathbf{c}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}\vee\overline{\mathbf{c}}\wedge\mathbf{a}\wedge\overline{\mathbf{d}}\vee\overline{\mathbf{c}}\wedge\overline{\mathbf{a}}\wedge\mathbf{d}) \leftrightarrow \mathbf{A}((\mathbf{c}\to\overline{\mathbf{a}}\wedge\mathbf{d})\wedge(\overline{\mathbf{c}}\to\mathbf{a}\vee\mathbf{d})\end{aligned} \tag{5.26}$$

We obtained formulas that characterize the complete and the stable semantics. Furthermore, it is interesting to note that we have obtained an equivalent, but more descriptive characterization of the stable semantics. Furthermore, we have also found a formula that describes a novel argumentation semantics. $\mathcal{AS}^{da}$ extensions are admissible sets that either defend or attack the arguments that they do not contain. A possible reason for desiring such a behavior is that arguments that are neither attacked nor defended, may be regarded as too undecided.

Next, we focus on conflict-free sets and discuss the choices for $\Phi \cap \mathcal{M}^-$. We have the same three choices. Let us see that in order for $\mathbf{c} \wedge \overline{\mathbf{a}} \wedge \overline{\mathbf{d}}$ to hold, there must exist an argument that satisfies $\overline{\mathbf{c}} \wedge \overline{\mathbf{a}}$. Thus, if we remove both $\overline{\mathbf{c}} \wedge \overline{\mathbf{a}} \wedge \mathbf{d}$ and $\overline{\mathbf{c}} \wedge \overline{\mathbf{a}} \wedge \overline{\mathbf{d}}$ we end up again with the description of the stable semantics. The other two formulas are:

$$\begin{aligned}
\mathcal{CF}^{sr}p &\leftrightarrow \mathbf{A}(\mathbf{c} \wedge \overline{\mathbf{a}} \wedge \mathbf{d} \vee \mathbf{c} \wedge \overline{\mathbf{a}} \wedge \overline{\mathbf{d}} \vee \overline{\mathbf{c}} \wedge \mathbf{a} \wedge \overline{\mathbf{d}} \vee \overline{\mathbf{c}} \wedge \overline{\mathbf{a}} \wedge \overline{\mathbf{d}}) \\
&\leftrightarrow \mathbf{A}((\mathbf{c} \rightarrow \overline{\mathbf{a}}) \wedge (\overline{\mathbf{c}} \rightarrow \overline{\mathbf{d}})) \\
\mathcal{CF}^{da}p &\triangleq \mathbf{A}(\mathbf{c} \wedge \overline{\mathbf{a}} \wedge \mathbf{d} \vee \mathbf{c} \wedge \overline{\mathbf{a}} \wedge \overline{\mathbf{d}} \vee \overline{\mathbf{c}} \wedge \mathbf{a} \wedge \overline{\mathbf{d}} \vee \overline{\mathbf{c}} \wedge \overline{\mathbf{a}} \wedge \mathbf{d}) \\
&\leftrightarrow \mathbf{A}((\mathbf{c} \rightarrow \overline{\mathbf{a}}) \wedge (\overline{\mathbf{c}} \rightarrow \mathbf{a} \vee \mathbf{d}))
\end{aligned} \tag{5.27}$$

The second formula is another new semantics, the conflict-free counterpart of $\mathcal{AS}^{da}$, i.e. conflict-free sets that either attack or defend outer arguments. Let us now see that $\mathcal{CF}^{da}$ and $\mathcal{AS}^{da}$ are indeed distinct semantics. An inspection of the map of argumentation semantics from Figure 3.18 (page 58) shows that we need to compare $\mathcal{AS}^{da}$ with $\mathcal{AS}^{wr}$, $\mathcal{AS}^{P}$ and with admissible sets. For $\mathcal{CF}^{da}$ we need to perform comparisons with $\mathcal{CF}^{cr}$, $\mathcal{CF}^{wr}$, $\mathcal{AS}$, $\mathcal{CF}^{P}$ and the skeptical semantics.

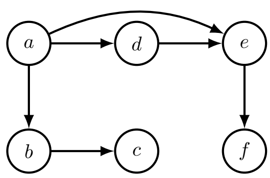



**Figure 5.3:** *Argumentation framework for exemplifying $\mathcal{AS}^{da}$ and $\mathcal{CF}^{da}$*

To keep things simple, the framework from Figure 5.3 does not consider the distinctions with respect to skeptical semantics or ideal sets. Let us see that the set $\{a, f\}$ attacks $b$, $d$ and $e$ and defends $c$, thus it is a $\mathcal{AS}^{da}$ extension. On the other hand it is not a $\mathcal{AS}^{wr}$ extension because the grounded extension of $F$ is $\{a, c, f\}$ and it is also not a prudent admissible set because $a$ indirectly attacks $f$. Furthermore, the same set is also a $\mathcal{CF}^{da}$ extension but is not a prudent conflict-free set and also not in $\mathcal{CF}^{wr}$ or $\mathcal{CF}^{cr}$. For a set that is in $\mathcal{CF}^{da}$ and is not admissible, consider $\{b, d, f\}$.

To sum up, we have described several semantics using this approach. Most of them are the ones already covered in (Grossi, 2010), but presented in the context of a systematic search for semantics that can be described in terms of attack and defense. We have also described the novel $\mathcal{CF}^{sr}$ semantics that we have introduced for exploiting the reinstatement principle. Furthermore, the analysis has also uncovered two new argumentation semantics, $\mathcal{AS}^{da}$ and $\mathcal{CF}^{da}$. We provide an overview of this subsection's results in Table 5.3. As promised, we also translate the formulas based on attack and defense to $ML(\Diamond, \mathbf{E})$ formulas.

| Notation | $(\mathbf{c}, \mathbf{a}, \mathbf{d})$ and $ML(\Diamond, \mathbf{E})$ formula | Description |
|---|---|---|
| $\mathcal{CF}$ | $\mathbf{A}(\mathbf{c} \to \overline{\mathbf{a}})$<br>$\mathbf{A}(p \to \neg\Diamond p)$ | conflict-free sets |
| $\mathcal{AS}$ | $\mathbf{A}(\mathbf{c} \to \overline{\mathbf{a}} \wedge \mathbf{d})$<br>$\mathbf{A}(p \to \neg\Diamond p \wedge \Box\Diamond p)$ | admissible sets |
| $\mathcal{CO}$ | $\mathbf{A}((\mathbf{c} \to \overline{\mathbf{a}} \wedge \mathbf{d}) \wedge (\overline{\mathbf{c}} \to \overline{\mathbf{d}}))$<br>$\mathbf{A}((p \to \neg\Diamond p \wedge \Box\Diamond p) \wedge (\neg p \to \neg\Box\Diamond p))$ | complete extensions |
| $\mathcal{ST}$ | $\mathbf{A}(\mathbf{c} \leftrightarrow \overline{\mathbf{a}} \wedge \mathbf{d})$<br>$\mathbf{A}(p \leftrightarrow \neg\Diamond p \wedge \Box\Diamond p)$ | stable extensions |
| $\mathcal{CF}^{sr}$ | $\mathbf{A}((\mathbf{c} \to \overline{\mathbf{a}}) \wedge (\overline{\mathbf{c}} \to \overline{\mathbf{d}}))$<br>$\mathbf{A}((p \to \neg\Diamond p) \wedge (\neg p \to \neg\Box\Diamond p))$ | conflict-free sets + strong reinstatement |
| $\mathcal{CF}^{da}$ | $\mathbf{A}((\mathbf{c} \to \overline{\mathbf{a}}) \wedge (\overline{\mathbf{c}} \to \mathbf{a} \vee \mathbf{d}))$<br>$\mathbf{A}((p \to \neg\Diamond p) \wedge (\neg p \to \Diamond p \vee \Box\Diamond p))$ | conflict-free sets + attack or defend |
| $\mathcal{AS}^{da}$ | $\mathbf{A}((\mathbf{c} \to \overline{\mathbf{a}} \wedge \mathbf{d}) \wedge (\overline{\mathbf{c}} \to \mathbf{a} \vee \mathbf{d}))$<br>$\mathbf{A}((p \to \neg\Diamond p \wedge \Box\Diamond p) \wedge (\neg p \to \Diamond p \vee \Box\Diamond p))$ | admissible sets + attack or defend |

**Table 5.3:** *Argumentation semantics that can be described using attack and defense.*

This approach is carried out further in (Gratie and Florea, 2010) to also use maximality with attack and defense. This leads to descriptions for $\mathcal{PR}$, $\mathcal{GR}$ and $\mathcal{SST}$, but such descriptions can can no longer be translated to $ML(\Diamond, \mathbf{E})$, as maximality is not expressible in this logic. In fact, maximality in general is not even captured by $\mu$-calculus.

### 5.3.3 $ML(\Diamond, \mathbf{E})$ vs argumentation principles

In this subsection we provide a negative result that shows the limitations of modal logic with respect to describing argumentation semantics and also how these limitations are connected to important evaluation principles of argumentation semantics. More precisely, we will show that universally-defined subsemantics of $\mathcal{CO}$ that also satisfy additivity cannot be captured by global modal formulas. We will prove this using satisfiability results from the previous section.

For a start, we recall the formula characterizing the complete semantics:

$$\mathcal{CO}p \leftrightarrow \mathbf{A}((p \to \neg\Diamond p) \wedge (p \leftrightarrow \Box\Diamond p)) \triangleq \mathbf{A}\phi_{\mathcal{CO}} \tag{5.28}$$

The first thing we need to do is convert $\phi_{\mathcal{CO}}$ to disjunctive normal form. We start by converting it to negation normal form:

$$\begin{aligned} \phi_{\mathcal{CO}} &= (p \to \neg\Diamond p) \wedge (p \leftrightarrow \Box\Diamond p) \\ &\leftrightarrow (\neg p \vee \neg\Diamond p) \wedge (p \wedge \Box\Diamond p \vee \neg p \wedge \neg\Box\Diamond p) \\ &\leftrightarrow (\neg p \vee \Box\neg p) \wedge (p \wedge \Box\Diamond p \vee \neg p \wedge \Diamond\Box\neg p) \\ &\leftrightarrow p \wedge \Box\neg p \wedge \Box\Diamond p \vee \neg p \wedge \Diamond\Box\neg p \vee \neg p \wedge \Box\neg p \wedge \Diamond\Box\neg p \\ &\leftrightarrow p \wedge \Box\neg p \wedge \Box\Diamond p \vee \neg p \wedge \Diamond\Box\neg p \end{aligned} \tag{5.29}$$

Next, we take thee two disjuncts separately and convert them to second degree disjunctive normal form.

$$\begin{aligned} p \wedge \Box\neg p \wedge \Box\Diamond p &\leftrightarrow p \wedge \Box(\neg p \wedge \Diamond p) \\ &\leftrightarrow p \wedge \Box(\neg p \wedge \diamond\{p\} \vee \neg p \wedge \diamond\{p, \neg p\}) \\ &\leftrightarrow \bigvee\{p \wedge \diamond\Phi \mid \Phi \subseteq \{\neg p \wedge \diamond\{p\}, \neg p \wedge \diamond\{p, \neg p\}\}\} \\ &\triangleq \bigvee \Phi^{+}_{\mathcal{CO}} \end{aligned} \tag{5.30}$$

Note that there are a total of 4 minterms in $\Phi^{+}_{\mathcal{CO}}$. For the second disjunct of $\phi_{\mathcal{CO}}$ we have:

$$\begin{aligned} \neg p \wedge \Diamond\Box\neg p &\leftrightarrow \neg p \wedge \Diamond(p \wedge \diamond\varnothing \vee p \wedge \diamond\{\neg p\} \vee \neg p \wedge \diamond\varnothing \vee \neg p \wedge \diamond\{\neg p\}) \\ &\leftrightarrow \bigvee\{\neg p \wedge \diamond\Phi \mid \Phi \subseteq \mathcal{F}_1, \\ &\qquad\qquad \Phi \cap \{p \wedge \diamond\varnothing, p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\varnothing, \neg p \wedge \diamond\{\neg p\}\} \neq \varnothing\} \\ &\triangleq \bigvee \Phi^{-}_{\mathcal{CO}} \end{aligned} \tag{5.31}$$

Now let us see that for this part of the formula we have quite an impressive number of minterms. Indeed, the intersection condition tells us that we must choose at least one element from that set of first degree minterms, for a total of $2^4 - 1 = 15$ combination, while from the remaining 4 minterms we can choose unrestricted, so we have a total of $15 \times 16 = 240$ minterms. We denote $\Phi_{\mathcal{CO}} = \Phi^{+}_{\mathcal{CO}} \cup \Phi^{-}_{\mathcal{CO}}$ and we get:

$$\phi_{\mathcal{CO}} \leftrightarrow \bigvee \Phi_{\mathcal{CO}} \tag{5.32}$$

where $\Phi_{\mathcal{CO}}$ has a total number of 244 minterms. Now, recall that in fact not all these minterms are satisfiable together and, moreover, that we can build models for $\phi \wedge \mathbf{A}\phi_{\mathcal{CO}}$ based on the reduction of satisfiable minterms. In our case, we shall see that this simplifies the problem a lot.

We are now going to compute $sat^{*}(\Phi_{\mathcal{CO}})$ and $sat^{*}_{-1}(\Phi_{\mathcal{CO}})$. For this, let us first compute the set of reductions for the minterms in $\Phi_{\mathcal{CO}}$, $\rho(\Phi_{\mathcal{CO}})$:

$$\begin{aligned} \rho(\Phi_{\mathcal{CO}}) &= \{\rho(\phi) \mid \phi \in \Phi_{\mathcal{CO}}\} \\ &= \{p \wedge \diamond\varnothing, p \wedge \diamond\{\neg p\}\} \cup \{\neg p \wedge \diamond\Phi \mid \Phi \subseteq \mathcal{F}_0, \Phi \cap \{p, \neg p\} \neq \varnothing\} \\ &= \{p \wedge \diamond\varnothing, p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\{p\}, \neg p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\{p, \neg p\}\} \end{aligned} \tag{5.33}$$

Thus, we have only 5 reduced minterms. We use them to compute $sat(\Phi_{\mathcal{CO}})$:

$$\begin{aligned}
\Phi'_{\mathcal{CO}} \triangleq sat(\Phi_{\mathcal{CO}}) &= \{\phi \in \Phi_{\mathcal{CO}} \mid \phi/\diamond \subseteq \rho(\Phi_{\mathcal{CO}})\} \\
&= \{p \wedge \diamond\Phi \mid \Phi \subseteq \{\neg p \wedge \diamond\{p\}, \neg p \wedge \diamond\{p, \neg p\} \cap \rho(\Phi_{\mathcal{CO}})\}\} \\
&\cup \{p \wedge \diamond\Phi \mid \Phi \subseteq \rho(\Phi_{\mathcal{CO}}), \Phi \cap \{p \wedge \diamond\varnothing, p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\varnothing, \neg p \wedge \diamond\{\neg p\}\} \neq \varnothing\} \\
&= \{p \wedge \diamond\varnothing, p \wedge \diamond\{\neg p \wedge \diamond\{p\}\}, p \wedge \diamond\{\neg p \wedge \diamond\{p, \neg p\}\}, \\
&\qquad p \wedge \diamond\{\neg p \wedge \diamond\{p\}, \neg p \wedge \diamond\{p, \neg p\}\}\} \\
&\cup \{p \wedge \diamond\Phi \mid \Phi \subseteq \rho(\Phi_{\mathcal{CO}}), \Phi \cap \{p \wedge \diamond\varnothing, p \wedge \diamond\{\neg p\}, \neg p \wedge \diamond\{\neg p\}\} \neq \varnothing\}
\end{aligned} \tag{5.34}$$

Let us see that the total number of minterms has become $4 + (2^3 - 1) \times 2^2 = 32$, which is significantly less than the initial 240. Furthermore, let us see that $\rho(\Phi'_{\mathcal{CO}}) = \rho(\Phi_{\mathcal{CO}})$, which leads to $sat(\Phi'_{\mathcal{CO}}) = \Phi'_{\mathcal{CO}}$. This, in turn, means that $sat^*(\Phi_{\mathcal{CO}}) = \Phi'_{\mathcal{CO}}$ and, furhtermore, $sat^*_{-1}(\Phi_{\mathcal{CO}}) = \rho(\Phi_{\mathcal{CO}})$. We introduce the following notations:

$$\begin{aligned}
\alpha_1 &= p \wedge \diamond\varnothing \\
\alpha_2 &= p \wedge \diamond\{\neg p\} \\
\alpha_3 &= \neg p \wedge \diamond\{p\} \\
\alpha_4 &= \neg p \wedge \diamond\{\neg p\} \\
\alpha_5 &= \neg p \wedge \diamond\{p, \neg p\}
\end{aligned} \tag{5.35}$$

In order to build the model that satisfies each of the above formulas, we need to choose a 2-nd degree modal minterm for each of them, just as suggested in the proof of Lemma 5. We make the following choices:

$$\begin{aligned}
\beta_1 &= p \wedge \diamond\varnothing \\
\beta_2 &= p \wedge \diamond\{\neg p \wedge \diamond\{p\}\} = p \wedge \diamond\{\alpha_3\} \\
\beta_3 &= \neg p \wedge \diamond\{p \wedge \diamond\varnothing\} = \neg p \wedge \diamond\{\alpha_1\} \\
\beta_4 &= \neg p \wedge \diamond\{\neg p \wedge \diamond\{\neg p\}\} = \neg p \wedge \diamond\{\alpha_4\} \\
\beta_5 &= \neg p \wedge \diamond\{p \wedge \diamond\varnothing, \neg p \wedge \diamond\{p\}\} = \neg p \wedge \diamond\{\alpha_1, \alpha_3\}
\end{aligned} \tag{5.36}$$

Now, in our model $\mathfrak{M}_{\mathcal{CO}} = (W_{\mathcal{CO}}, R_{\mathcal{CO}}, V_{\mathcal{CO}})$ we have a world $w_i$ for each $\beta_i$ and the accessibility relation satisfies:

$$(w_i, w_j) \in R_{\mathcal{CO}} \Leftrightarrow \alpha_j \in \beta_i/\diamond \tag{5.37}$$

Furthermore, the valuation function $V_{\mathcal{CO}}$ is defined so as to satisfy $\mathfrak{M}_{\mathcal{CO}}, w_i \Vdash \alpha_i/pr$. Based on the above considerations, which follow the construction from the proof of Lemma 5 we get the model depicted in Figure 5.4.

The most important property of the corresponding argumentation framework $F_{\mathcal{CO}} \triangleq (W_{\mathcal{CO}}, R^-_{\mathcal{CO}})$ is the fact that is has a single complete extension. Indeed, its grounded extension $\{w_1, w_2\}$ attacks $w_3$ and $w_4$, while $w_5$ is self-attacking, so there can be no other complete extension.

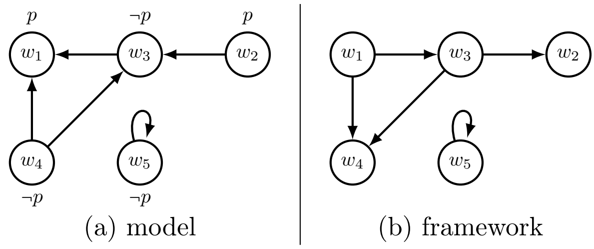



**Figure 5.4:** *Model that satisfies all first degree minterms from* $sat^*_{-1}(\Phi_{\mathcal{CO}})$

**Lemma 7.** *Any satisfiable formula* $\phi \wedge \mathbf{A}\phi_{\mathcal{CO}}$ *can be satisfied by a model* $\mathfrak{M} = (W, R, V)$ *such that the argumentation framework* $F = (W, R^-)$ *has s single complete extension.*

*Proof.* Based on Lemma 3 it is enough to consider the case when $\phi$ is a modal minterm. We will prove the claim by induction on the modal degree of $\phi$, following the constructive proof provided for Lemma 4.

If $md(\phi) \leq 1$, then we look for the modal minterm $\alpha_i \in sat^*_{-1}(\Phi_{\mathcal{CO}})$ such that $\tau_{md(\phi)}(\alpha_i) = \phi$ and we can choose the pointed model $(\mathfrak{M}_{\mathcal{CO}}, w_i)$ to satisfy $\phi$. We have seen already that this model has a single complete extension.

For $md(\phi) \geq 2$ we proceed by induction. We can consider the cases $md(\phi) = 0$ and $md(\phi) = 1$ as base cases and just prove the induction step. We employ the same recursive construction from Lemma 4. In other words, we assume that we have a pointed model $(\mathfrak{M}_\psi, w_\psi)$ that satisfies $\psi \wedge \mathbf{A}\phi_{\mathcal{CO}}$ for each $\psi \in \phi/\diamond$, where $\mathfrak{M}_\psi = (W_\psi, R_\psi, V_\psi)$. We also consider the pointed model $(\mathfrak{M}_0, w_0)$ that satisfies $\phi/pr$. We construct the model $\mathfrak{M} = (W, R, V)$ that satisfies $\phi \wedge \mathbf{A}\phi_{\mathcal{CO}}$ as follows:

$$\begin{aligned} W &= W_0 \cup \bigcup_{\psi \in \phi/\diamond} W_\psi \\ R &= R_0 \cup \bigcup_{\psi \in \phi/\diamond} (R_\psi \cup \{(w_0, w_\psi)\}) \\ V(p) &= V_0(p) \cup \bigcup_{\psi \in \phi/\diamond} V_\psi(p) \end{aligned} \tag{5.38}$$

We have already seen in the proof of Lemma 4 that the model constructed like this satisfies the desired formula. What we need to show here is that whenever the models $\mathfrak{M}_\psi$ are such that the argumentation framework $F_\psi \triangleq (W_\psi, R^-_\psi)$ has a single complete extension, then $F \triangleq (W, R^-)$ has the same property.

Let us see that the sets $W_\psi$ are unattacked in $F$, so we can use the directionality principle, satisfied by the complete semantics, to get that the intersection of any complete extension of $F$ with $W_\psi$ is a complete extension of $F_\psi$. Thus, we can deduce that the intersection of any complete extension of $F$ with $W \setminus \{w_0\}$ is unique, call this intersection $E$. The only way to have multiple complete extensions for $F$ is to have both $E$ and $E \cup \{w_0\}$ as complete extensions.

However, if $E \cup \{w_0\}$ is complete, then $E$ defends $w_0$ (because $w_0$ does not defend itself against any of its attackers), which means that $E$ is not complete. Thus, there is a single complete extension for $F$. □

**Theorem 7.** *$\mathcal{CO}$ is the only universally defined argumentation semantics that can be described with a $ML(\Diamond, \mathbf{E})$ formula and also satisfies admissibility, reinstatement and additivity.*

*Proof.* First of all, note that the complete semantics does satisfy all the properties mentioned in the theorem. What we need to show is that there is no other argumentation semantics that satisfies them as well.

Suppose that there exists such an argumentation semantics. Since it satisfies both admissibility and reinstatement, it means that it gives complete extensions, i.e. $\mathcal{S} \subseteq \mathcal{CO}$. Also we have that $\mathcal{S}$ can be described by a formula from $ML(\Diamond, \mathbf{E})$ and is additive. Based on Theorem 6, this means that there exists a formula $\mathbf{A}\phi_{\mathcal{S}}$ such that $Sp \leftrightarrow \mathbf{A}\phi_{\mathcal{S}}$ is valid and $\phi_{\mathcal{S}} \in ML(\Diamond)$.

We want $\mathcal{S}$ to be different from $\mathcal{CO}$, so there must exist a framework $F = (\mathcal{A}, \mathcal{R})$ and a set of arguments $E \subseteq \mathcal{A}$ such that $E \in \mathcal{E}_{\mathcal{CO}}(F) \setminus \mathcal{E}_{\mathcal{S}}(F)$. We construct the model $\mathfrak{M} = (\mathcal{A}, \mathcal{R}^{-}, V)$, where $V(p) = E$. Then it must be that $\mathfrak{M} \nVdash \mathbf{A}\phi_{\mathcal{S}}$, so there exists a world $w$ such that $\mathcal{M}, w \Vdash \neg\phi_{\mathcal{S}}$. Since $V(p)$ was chosen as a complete extension, we also have $\mathfrak{M}, w \Vdash \neg\phi_{\mathcal{S}} \wedge \mathbf{A}\phi_{\mathcal{CO}}$.

But then, based on Lemma 7, there exists a model $\mathfrak{M}' = (W', R', V')$ that satisfies the same formula and whose corresponding argumentation framework $F' = (W', R'^{-})$ has a single complete extension. Since the only complete extension of $F$, given by $V(p)$, is not an extension of $\mathcal{S}$, we would have that $\mathcal{S}$ gives no extension for $F'$, which is a contradiction with the assumption that $\mathcal{S}$ is universally defined.

Thus, the only argumentation semantics that satisfies all the properties from the theorem is the complete semantics. □

The immediate implication of this result for argumentation is that none of the following semantics can be described within $ML(\Diamond, \mathbf{E})$: $\mathcal{PR}$, $\mathcal{GR}$, $\mathcal{SST}$, $\mathcal{ID}$, $\mathcal{EAG}$, $\mathcal{GR}^*$, $\mathcal{PR}^*$, $\mathcal{SST}^*$, $\mathcal{ID}^*$, $\mathcal{GR}^{*id}$, $\mathcal{PR}^{cids}$, $\mathcal{SST}^{cids}$, $\mathcal{GR}^{*cids}$, $\mathcal{AD}1$ and $\mathcal{AD}2$. It also means that it is impossible to define within $ML(\Diamond, \mathbf{E})$ a novel argumentation semantics that is additive, universally defined and that gives complete extensions.

### 5.3.4 Total bisimulation in use

We have seen so far in this section that the global modal language can describe several argumentation semantics and other important sets for argumentation, by adding to the results in (Grossi, 2010). We have also seen the limitations of this modal language, namely the fact that many argumentation semantics cannot be captured because of their properties. In this subsection we aim to cover the remaining semantics and show that they are beyond the expressive power of $ML(\Diamond, \mathbf{E})$. We will justify this with several bisimulation examples.

In order to prove that an argumentation semantics $\mathcal{S}$ cannot be captured by a formula from $ML(\Diamond, \mathbf{E})$ we are looking for two total bisimilar models such that the valuation of $p$ is an $\mathcal{S}$ extension for the argumentation framework corresponding to one of the models, but not for the other model.

The first example that we consider relies on the distinction that is generally made by argumentation semantics between odd and even length cycles. Such a distinction is not available in $ML(\Diamond, \mathbf{E})$. See for example the total bisimilar models $\mathfrak{M}$ (above) and $\mathfrak{M}'$ (below) from Figure 5.5 (a).

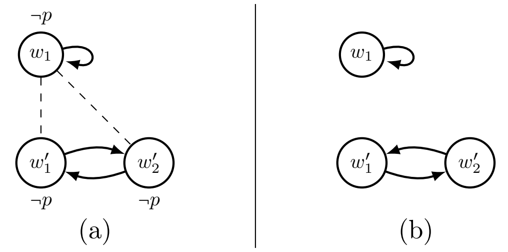



**Figure 5.5:** *Bisimulation example based on odd vs even length cycles.*

We will consider the following semantics: $\mathcal{MCF}$, $\mathcal{CF}1$, $\mathcal{CF}2$, $\mathcal{STA}$, $\mathcal{STA}2$, $\mathcal{EPS}$ and $\mathcal{PR}^P$. For all of them, the only extension that corresponds to the upper framework $F$ is the empty set. For the lower framework $F'$, all these semantics prescribe the extensions $\{w'_1\}$ and $\{w'_2\}$. Thus, the empty set, which is the valuation of $p$ in both models, is an extension for $F$ but not for $F'$. As a result, none of these semantics can be captured by a global modal formula.

Next, we focus on the skeptical semantics. The example from Figure 5.6 covers $\mathcal{PR}^S$, $\mathcal{SST}^S$ and $\mathcal{GR}^{*S}$. For the upper framework from the figure, $\{w_0, w_3\}$ and $\{w_1, w_3\}$ are the extensions of $\mathcal{PR}$, $\mathcal{SST}$ and $\mathcal{GR}^*$, so their skeptical versions yield $\{w_3\}$. For the lower framework on the other hand the empty set is the only extensions, so this will be the extension provided by the skeptical versions as well. Thus, the set encoded by the valuation of $p$, namely $\{w_3\}$ respectively $\{w'_3\}$, is an extension in one framework but not in the other. We conclude that $ML(\Diamond, \mathbf{E})$ cannot describe $\mathcal{PR}^S$, $\mathcal{SST}^S$ and $\mathcal{GR}^{*S}$.

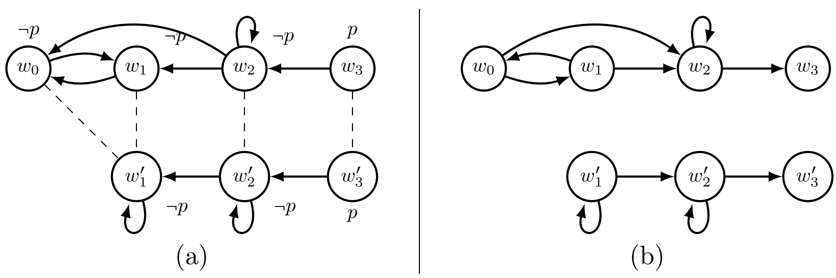



**Figure 5.6:** $ML(\Diamond, \mathbf{E})$ *cannot describe* $\mathcal{PR}^S$, $\mathcal{SST}^S$ *and* $\mathcal{GR}^{*S}$

Next, we move on to versions of $\mathcal{MCF}$ and $\mathcal{STA}$, based on the argumentation frameworks from Figure 5.7. The relevant extensions are given in Table 5.4. The valuation of $p$ in the upper model is $\{w_2\}$ and is only an extension for the base semantics that we considered. In the lower framework on the other hand, the valuation of $p$ is $\{w'_2\}$ and is an extension for all the semantics based on $\mathcal{MCF}$ and $\mathcal{STA}$. Thus, we conclude that $ML(\Diamond, \mathbf{E})$ cannot describe $\mathcal{MCF}^S$, $\mathcal{STA}^S$, $\mathcal{MCF}^{ids}$, $\mathcal{STA}^{ids}$, $\mathcal{MCF}^{id}$ and $\mathcal{STA}^{id}$. Thus, we have covered all skeptical semantics. Furthermore, let us see that, for $\mathcal{S}em \in \{\mathcal{PR}, \mathcal{SST}, \mathcal{GR}^*\}$, if there is a description in $ML(\Diamond, \mathbf{E})$ for $\mathcal{S}em^{ids}$, then we can obtain a description of $\mathcal{S}em^{cids}$, as $\mathcal{S}em^{cids} = \mathcal{S}em^{ids} \cap \mathcal{CO}$. However, we know that the complete ideal sets cannot be described, so it must be that $ML(\Diamond, \mathbf{E})$ cannot capture $\mathcal{IDS}$, $\mathcal{EAGS}$ and $\mathcal{GR}^{*ids}$.

| $\mathcal{S}em$ | $\mathcal{E}_{\mathcal{S}em}(F)$ | $\mathcal{E}_{\mathcal{S}em}(F')$ |
|---|---|---|
| $\mathcal{MCF}$, $\mathcal{STA}$ | $\{\{w_0\},\{w_1\},\{w_2\}\}$ | $\{\{w_2'\}\}$ |
| $\mathcal{MCF}^S$, $\mathcal{STA}^S$ | $\{\varnothing\}$ | $\{\{w_2'\}\}$ |
| $\mathcal{MCF}^{ids}$, $\mathcal{STA}^{ids}$ | $\{\varnothing\}$ | $\{\varnothing,\{w_2'\}\}$ |
| $\mathcal{MCF}^{id}$, $\mathcal{STA}^{id}$ | $\{\varnothing\}$ | $\{\{w_2'\}\}$ |

**Table 5.4:** *Extensions of semantics based on $\mathcal{MCF}$ and $\mathcal{STA}$ for the frameworks from Figure 5.7*

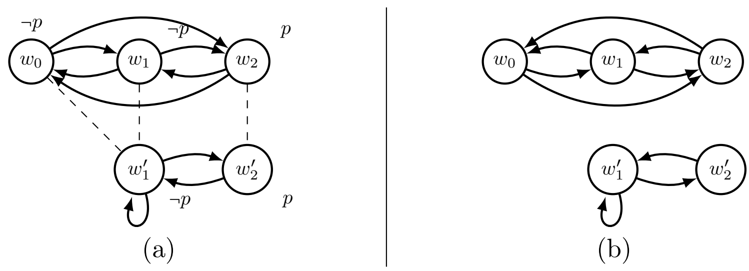



**Figure 5.7:** *Bisimulation example for semantics based on $\mathcal{MCF}$ and $\mathcal{STA}$*

The last bisimulation example that we consider is presented in Figure 5.8. For the upper framework we have $\mathcal{E}_{\mathcal{CF}^{cr}}(F) = \mathcal{E}_{\mathcal{CF}^{cw}}(F) = \mathcal{E}_{\mathcal{CO}^P}(F = \{\{\varnothing,\{w_1\},\{w_2,w_4\},\{w_3\},\{w_4\}\}$. Thus, the valuation of $p$, $V(p) = \{w_2\}$ is not an extension. For the lower framework on the other hand, we have $\mathcal{E}_{\mathcal{CF}^{cr}}(F') = \mathcal{E}_{\mathcal{CF}^{cw}}(F') = \mathcal{E}_{\mathcal{CO}^P}(F) = \{\varnothing,\{w_1'\},\{w_2'\},\{w_3'\}\}$ so the valuation of $p$ in the corresponding model $V'(p) = \{w_2'\}$ is an extension. We have thus shown that $ML(\Diamond, \mathbf{E})$ is not expressive enough to capture $\mathcal{CF}^{cr}$, $\mathcal{CF}^{cw}$ and $\mathcal{CO}^P$.

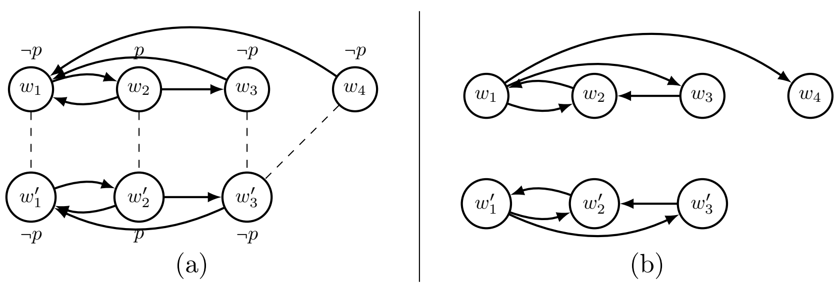



**Figure 5.8:** *Bisimulation example for $\mathcal{CF}^{cr}$, $\mathcal{CF}^{cw}$ and $\mathcal{CO}^P$*

### 5.3.5 Discussion

In this section we have explored the use of $ML(\Diamond, \mathbf{E})$ for describing argumentation semantics. We have expanded on the results from (Grossi, 2010) and we have

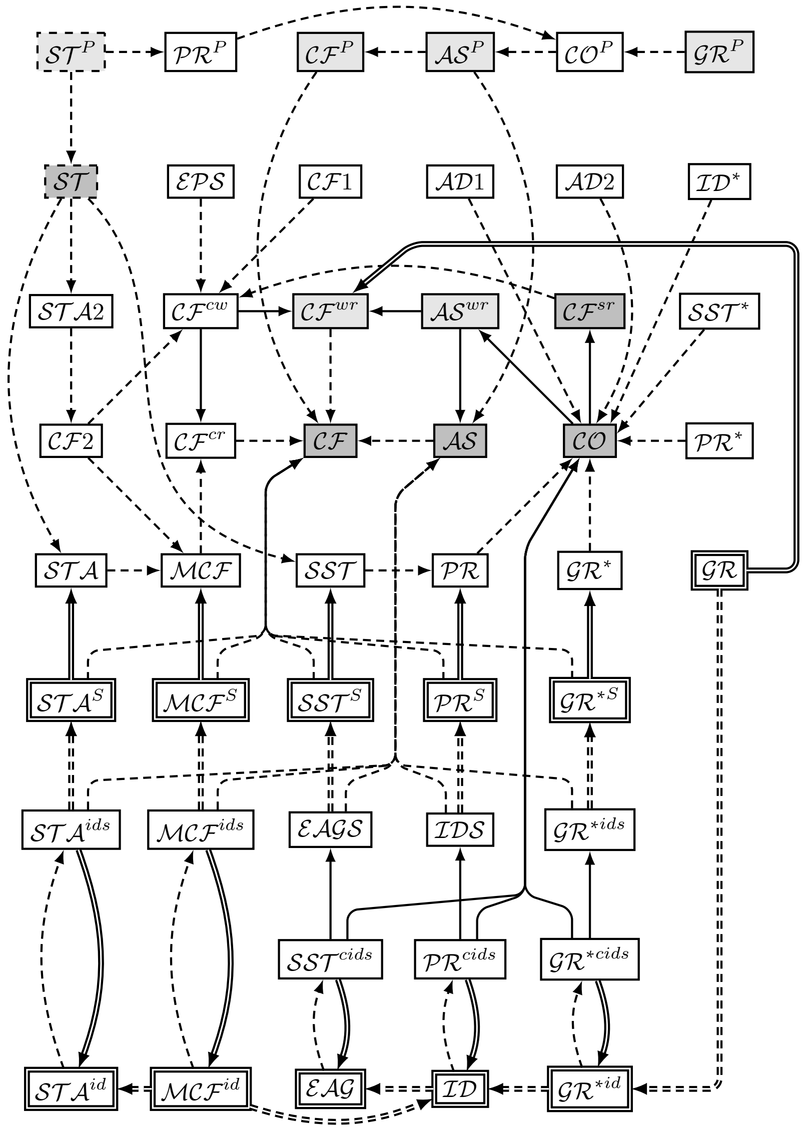



**Figure 5.9:** *Overview of the use of $ML(\Diamond, \mathbf{E})$ for describing argumentation semantics: dark gray background = can be described, white background = cannot be described, light gray background = undecided*

also provided proofs for the semantics that cannot be captured. An overview is presented in Figure 5.9. As can be seen in the figure, there are several semantics that are still undecided, in the sense that we do not have a description formula for them, but also we have no proof that such a formula does not exist. This is because all these semantics can be described within $\mu$-calculus with global modalities as follows:

$$
\begin{aligned}
&\mathcal{CF}^{wr} p \leftrightarrow \mathbf{A}((p \rightarrow \neg\Diamond p) \wedge (\mu X.\Box\Diamond X \rightarrow p)) \\
&\mathcal{AS}^{wr} p \leftrightarrow \mathbf{A}((p \rightarrow \neg\Diamond p \wedge \Box\Diamond p) \wedge (\mu X.\Box\Diamond X \rightarrow p)) \\
&\mathcal{CF}^{P} p \leftrightarrow \mathbf{A}(p \rightarrow \neg\mu X.(\Diamond p \vee \Diamond\Diamond X)) \\
&\mathcal{AS}^{P} p \leftrightarrow \mathbf{A}(p \rightarrow \neg\mu X.(\Diamond p \vee \Diamond\Diamond X) \wedge \Box\Diamond p) \\
&\mathcal{ST}^{P} p \leftrightarrow \mathbf{A}((p \rightarrow \neg\mu X.(\Diamond p \vee \Diamond\Diamond X)) \wedge (\neg p \rightarrow \Diamond p)) \\
&\mathcal{GR}^{P} p \leftrightarrow \mathbf{A}(p \leftrightarrow \mu X.\Box\mu Y.\Diamond X \vee \Diamond\Diamond Y)
\end{aligned}
\tag{5.39}
$$

Since $\mu$-calculus with the global modality is invariant under total bisimulation, a proof based on total bisimulations cannot be provided for showing that these semantics cannot be captured by $ML(\Diamond, \mathbf{E})$.

In light of this observation, the general result presented in Theorem 7 becomes more important, as it covers the grounded semantics, which is also total bisimulation invariant. In addition, that result is also important because it links modal logic to actual evaluation principles for argumentation semantics.

## 5.4 Doing Argumentation in $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$

In this section we explore the capabilities of the converse modalities for argumentation. We shall follow a similar path as that of the previous section, by first providing simple formulas that are meaningful for argumentation. We will, however, focus on the added expressive power rather then exploring all possible formulas. We will also prove a result similar to Theorem 7, but applicable to $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$.

Recall that so far we have used the following convention for defining a modal operator based on the argumentation semantics $\mathcal{S}$:

$$\mathfrak{M}, w \Vdash \mathcal{S}p \Leftrightarrow V^*(p) \in \mathcal{E}_{\mathcal{S}}(W, R^-) \tag{5.40}$$

In other words, we had to compute the extensions based on the converse of the accessibility relation from the model. This was due to the fact that the acceptability of arguments mostly depends on the acceptability of their attackers and thus leads to an implicit use of the converse relation. On the other hand, we now have the converse modalities in the language, so we can use the accessibility relation directly. We will put a bar above argumentation-based operators defined this way, in order to have a clear distinction between the two types:

$$\mathfrak{M}, w \Vdash \overline{\mathcal{S}}p \Leftrightarrow V^*(p) \in \mathcal{E}_{\mathcal{S}}(W, R) \tag{5.41}$$

Furthermore, let us see that whenever $\mathcal{S}p$ can be translated to a $ML(\Diamond, \mathbf{E})$ formula, we can obtain a formula for $\overline{\mathcal{S}}$ simply by replacing all occurrences of $\Diamond$

and $\Box$ with $\overline{\Diamond}$ and $\overline{\Box}$ respectively.

$$\begin{aligned}
&\overline{\mathcal{CF}}p \leftrightarrow \mathbf{A}(p \to \neg\overline{\Diamond}p)\\
&\overline{\mathcal{AS}}p \leftrightarrow \mathbf{A}(p \to \neg\overline{\Diamond}p \wedge \overline{\Box}\overline{\Diamond}p)\\
&\overline{\mathcal{CO}}p \leftrightarrow \mathbf{A}((p \to \neg\overline{\Diamond}p \wedge \overline{\Box}\overline{\Diamond}p) \wedge (\neg p \to \neg\overline{\Box}\overline{\Diamond}p))\\
&\overline{\mathcal{ST}}p \leftrightarrow \mathbf{A}(p \leftrightarrow \neg\overline{\Diamond}p)\\
&\overline{\mathcal{CF}^{sr}}p \leftrightarrow \mathbf{A}((p \to \neg\overline{\Diamond}p) \wedge (\neg p \to \neg\overline{\Diamond}p \vee \overline{\Box}\overline{\Diamond}p))
\end{aligned} \tag{5.42}$$

Although being able to use the accessibility relation directly as an attack relation for defining argumentation-based concepts is a significant advantage in itself, we are also interested in the additional expressive power that the converse modalities bring. Although there is a definite gain in expressiveness, we shall also see that most of the negative results from the previous section hold for $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ as well.

### 5.4.1 Additional expressive power

In this subsection we aim to identify the added value of using the converse modalities in argumentation, aside from the fact that we no longer need to work with the converse of the accessibility relation as an attack relation, but can use the model directly as an argumentation framework.

We have seen in Subsection 5.3.1 that $ML(\Diamond, \mathbf{E})$ cannot describe the set of all isolated arguments. On the other hand if we use the converse modalities, we can use the following formula:

$$\overline{\mathcal{IA}}p \leftrightarrow \mathbf{A}(p \leftrightarrow \Box\bot \wedge \overline{\Box}\bot) \tag{5.43}$$

Furthermore, let us consider the following formula:

$$\mathbf{A}(\Diamond p \wedge \overline{\Diamond}p \to p) \tag{5.44}$$

What this formula says about the set of arguments corresponding to the valuation of $p$ is that if an argument $a$ is attacked by the set but also attacks the set, then $a$ is included in the set. This idea is related to SCC-recursiveness, because such sets are in fact unions of strongly connected components. They can be seen as a generalization of SCC's and used for defining SCC-recursiveness as a directionality-like principle (Gratie et al., 2012b).

With the above examples, we already see that the additional expressive power given by the converse modalities is relevant for argumentation. Nevertheless, our main concern is related to describing argumentation semantics. So let us see that now we can describe $\mathcal{CF}$-reinstatement and the corresponding semantics, $\mathcal{CF}^{cr}$:

$$\mathcal{CF}^{cr}p \leftrightarrow \mathbf{A}((p \to \neg\overline{\Diamond}p) \wedge (\overline{\Box}\overline{\Diamond}p \wedge \neg\overline{\Diamond}p \wedge \neg\Diamond p \to p)) \tag{5.45}$$

Recall that $\mathcal{CF}$-reinstatement requires that for any extension $E$ and any argument $a$ such that $E$ defends $a$ and $E \cup \{a\}$ is conflict-free it holds that $a \in E$. Let us now consider the second conjunct of the formula describing $\mathcal{CF}^{cr}$, the one that corresponds to $\mathcal{CF}$-reinstatement: $\overline{\Box}\overline{\Diamond}p \wedge \neg\overline{\Diamond}p \wedge \neg\Diamond p \rightarrow p$. Clearly we have that $\overline{\Box}\overline{\Diamond}p$ corresponds to the fact that the extension $E$ defends the argument $a$ (at which the evaluation of the formula is done). On the other hand, let us note that $\neg\overline{\Diamond}p \wedge \neg\Diamond p$ does not really correspond to $E \cup \{a\}$ being conflict-free, but only to $E \not\rightarrow a$ and $a \not\rightarrow E$, which is a weaker condition, because $a$ may be self-attacking. For describing $\mathcal{CF}$-reinstatement this is not an issue, because of the defense part: if $a$ is self-attacking and $E$ defends $a$, it must hold that $E \rightarrow a$, so the condition for $\mathcal{CF}$-reinstatement cannot hold. On the other hand, the observation that we cannot capture $E \cup \{a\}$ being conflict-free with a modal formula means that we cannot describe, for example, the naive semantics.

We do not employ a systematic analysis of $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ as we did in the previous section for $ML(\Diamond, \mathbf{E})$ because there are too many distinct minterms in this case, but even more importantly because we shall see further on that the advantages of the converse modalities are limited to what we have presented already, at least as far as describing argumentation semantics is concerned.

### 5.4.2 $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ vs argumentation principles

In this subsection we show that the connection that we have established between $ML(\Diamond, \mathbf{E})$ and several evaluation principles of argumentation semantics (as well as the negative implications of this connection) also holds for $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. We recall the negation normal form for the description of the complete semantics presented in (5.29):

$$\mathcal{CO}p \leftrightarrow \mathbf{A}(p \wedge \Box(\neg p \wedge \Diamond p) \vee \neg p \wedge \Diamond\Box\neg p) \tag{5.46}$$

and we translate this formula in the new context of using the accessibility relation directly as an attack relation:

$$\overline{\mathcal{CO}}p \leftrightarrow \mathbf{A}(p \wedge \overline{\Box}(\neg p \wedge \overline{\Diamond}p) \vee \neg p \wedge \overline{\Diamond}\overline{\Box}\neg p) \triangleq \mathbf{A}\phi_{\mathcal{CO}} \tag{5.47}$$

Our goal in what follows is to convert $\phi_{\mathcal{CO}}$ to its equivalent disjunctive normal form in $ML(\Diamond, \overline{\Diamond})$. The challenge in this case consists in the number of minterms of this translation which, as we shall see shortly, is significantly greater than it was in the case of $ML(\Diamond)$. We start by translating the first degree subformulas. We start with $\neg p \wedge \overline{\Diamond}p$:

$$\begin{aligned}\neg p \wedge \overline{\Diamond}p &\leftrightarrow \bigvee\{\neg p \wedge \diamonddot\Psi_1 \wedge \overline{\diamonddot}\Psi_2 \mid \Psi_1 \subseteq \overline{\mathcal{F}}_0, \Psi_2 \subseteq \overline{\mathcal{F}}_0, p \in \Psi_2\} \\ &\leftrightarrow \bigvee\{\neg p \wedge \diamonddot\Psi_1 \wedge \overline{\diamonddot}\{p\}, \neg p \wedge \diamonddot\Psi_1 \wedge \overline{\diamonddot}\{p, \neg p\} \mid \Psi_1 \subseteq \overline{\mathcal{F}}_0\} \\ &\triangleq \bigvee \Theta_1\end{aligned} \tag{5.48}$$

Note that there are a total of 8 minterms in $\Theta_1$, as opposed to only two in the translation of the corresponding subformula in $ML(\Diamond)$. We now convert the first degree subformula $\overline{\Box}\neg p$:

$$
\begin{aligned}
\overline{\Box}\neg p &\leftrightarrow \bigvee\{p \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2, \neg p \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \mid \Phi_1 \subseteq \overline{\mathcal{F}}_0, \Phi_2 \subseteq \{\neg p\}\} \\
&\leftrightarrow \bigvee\{p \wedge \Diamonddot\Psi_1 \wedge \overline{\Diamonddot}\varnothing, p \wedge \Diamonddot\Psi_1 \wedge \overline{\Diamonddot}\{\neg p\}, \\
&\qquad \neg p \wedge \Diamonddot\Psi_1 \wedge \overline{\Diamonddot}\varnothing, \neg p \wedge \Diamonddot\Psi_1 \wedge \overline{\Diamonddot}\{\neg p\} \mid \Psi_1 \subseteq \overline{\mathcal{F}}_0\} \\
&\triangleq \bigvee \Theta_2
\end{aligned}
\tag{5.49}
$$

The number of minterms in $\Theta_2$ is 16. We can now compute the second degree minterms, separately for each disjunct, just as we did in Subsection 5.3.3. Note that in this case we must also account for the satisfiability constraint applicable to minterms from $ML(\Diamond, \overline{\Diamond})$. For the disjunct corresponding to accepted arguments, we have:

$$
\begin{aligned}
p \wedge \overline{\Box}(\neg p \wedge \Diamond p) &\leftrightarrow \bigvee\{\phi = p \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \mid \Phi_1 \subseteq \overline{\mathcal{F}}_1, \Phi_2 \subseteq \Theta_1, \\
&\qquad \forall\psi(\psi \in \Phi_1 \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\overline{\Diamonddot}), \forall\psi(\psi \in \Phi_2 \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\Diamonddot)\} \\
&\leftrightarrow \bigvee\{p \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \mid \Phi_1 \subseteq \{\psi \in \overline{\mathcal{F}}_1 \mid p \in \psi/\overline{\Diamonddot}\}, \\
&\qquad \Phi_2 \subseteq \{\neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p\}, \neg p \wedge \Diamonddot\{p\} \wedge \overline{\Diamonddot}\{p, \neg p\}, \\
&\qquad\qquad \neg p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\{p\}, \neg p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\{p, \neg p\}\}\} \\
&\triangleq \bigvee \Phi^+_{\mathcal{CO}}
\end{aligned}
\tag{5.50}
$$

There are a total of $2^{16} \times 2^4 = 2^{20}$ minterms. Simlarly, we compute the disjunctive normal form for the part that corresponds to rejected arguments:

$$
\begin{aligned}
\neg p \wedge \overline{\Diamond}\overline{\Box}\neg p &\leftrightarrow \bigvee\{\phi = \neg p \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \mid \Phi_1 \subseteq \overline{\mathcal{F}}_1, \Phi_2 \subseteq \overline{\mathcal{F}}_1, \Phi_2 \cap \Theta_2 \neq \varnothing, \\
&\qquad \forall\psi(\psi \in \Phi_1 \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\overline{\Diamonddot}), \forall\psi(\psi \in \Phi_2 \Rightarrow \overline{\rho}(\overline{\rho}(\phi)) \in \psi/\Diamonddot)\} \\
&\leftrightarrow \bigvee\{\neg p \wedge \Diamonddot\Phi_1 \wedge \overline{\Diamonddot}\Phi_2 \mid \Phi_1 \subseteq \{\psi \in \overline{\mathcal{F}}_1 \mid \neg p \in \psi/\overline{\Diamonddot}\}, \\
&\qquad \Phi_2 \subseteq \{\psi \in \overline{\mathcal{F}}_1 \mid \neg p \in \psi/\Diamonddot\}, \Phi_2 \cap \Theta_2' \neq \varnothing\} \\
&\triangleq \bigvee \Phi^-_{\mathcal{CO}}
\end{aligned}
\tag{5.51}
$$

where

$$
\begin{aligned}
\Theta_2' &\triangleq \Theta_2 \cap \{\psi \in \overline{\mathcal{F}}_1 \mid \neg p \in \psi/\Diamonddot\} \\
&= \{p \wedge \Diamonddot\{\neg p\} \wedge \overline{\Diamonddot}\varnothing, p \wedge \Diamonddot\{\neg p\} \wedge \overline{\Diamonddot}\{p\}, p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\varnothing, \\
&\qquad p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\{p\}, \neg p \wedge \Diamonddot\{\neg p\} \wedge \overline{\Diamonddot}\varnothing, \neg p \wedge \Diamonddot\{\neg p\} \wedge \overline{\Diamonddot}\{p\}, \\
&\qquad \neg p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\varnothing, \neg p \wedge \Diamonddot\{p, \neg p\} \wedge \overline{\Diamonddot}\{p\}\}
\end{aligned}
\tag{5.52}
$$

There are a total of $2^{16} \times 2^8 \times (2^8 - 1) = 255 \times 2^{24} = 4080 \times 2^{20}$ minterms.

We denote $\Phi_{\mathcal{CO}} \triangleq \Phi^+_{\mathcal{CO}} \cup \Phi^-_{\mathcal{CO}}$. The complete translation of $\phi_{\mathcal{CO}}$ to disjunctive normal form is given by $\phi_{\mathcal{CO}} \leftrightarrow \bigvee \Phi_{\mathcal{CO}}$. Note that $\Phi_{\mathcal{CO}}$ contains a total of $4081 \times 2^{20}$ minterms, which is roughly 4 billion, as compared to the 240 that we had for the $ML(\Diamond, \mathbf{E})$ disjunctive normal form.

Just as we did before, let us compute $sat(\Phi_{\mathcal{CO}})$ and hope for a significant reduction in the number of minterms. First, we need $\overline{\rho}(\Phi_{\mathcal{CO}})$:

$$
\begin{aligned}
\overline{\rho}(\Phi_{\mathcal{CO}}) = \{ & p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\varnothing, p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\varnothing, p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\varnothing, \\
& p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\varnothing, p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\neg p\}, p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{\neg p\}, \\
& p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\{\neg p\}, p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\{\neg p\}\} \\
\cup \{ & \neg p \wedge \diamonddot\Phi_1 \wedge \overline{\diamonddot}\Phi_2 \mid \Phi_1 \subseteq \overline{\mathcal{F}}_0, \Phi_2 \subseteq \overline{\mathcal{F}}_0, \Phi_2 \neq \varnothing\}
\end{aligned}
\tag{5.53}
$$

There are a total of 20 first degree minterms. We do not provide all the details, but after computing $sat(\Phi_{\mathcal{CO}})$, it follows that 6 of these 20 first degree minterms are not satisfiable. Indeed, let us see for example that $p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\varnothing$ is not satisfiable because it needs a formula $p \wedge \diamonddot\Psi_1 \wedge \overline{\diamonddot}\Psi_2$ such that $p \in \Psi_2$, but none of the 20 minterms satisfies this.

The 6 minterms that are not satisfiable are:

$$
\begin{aligned}
& p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\varnothing, \\
& p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{\neg p\}, \\
& p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\varnothing, \\
& p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\{\neg p\}, \\
& \neg p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{\neg p\}, \\
& \neg p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\{\neg p\}
\end{aligned}
\tag{5.54}
$$

The remaining 14 minterms are satisfiable, as we shall see by building the model that satisfies them. Note how the analysis of satisfiability has taken us from 4 billion second degree minterms to just 14 first degree minterms. In what follows, we give a name to each of the 14 satisfiable minterms and choose a second degree minterm to use for the model according to the approach presented in Lemma 6.

$$
\begin{array}{ll}
\alpha_1 = p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\varnothing & \beta_1 = p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\varnothing \\
\alpha_2 = p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\varnothing & \beta_2 = p \wedge \diamonddot\{\alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_{14}\} \wedge \overline{\diamonddot}\varnothing \\
\alpha_3 = p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\neg p\} & \beta_3 = p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\alpha_6, \alpha_8, \alpha_{12}, \alpha_{14}\} \\
\alpha_4 = p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\{\neg p\} & \beta_4 = p \wedge \diamonddot\{\alpha_6, \alpha_{11}, \alpha_{12}, \alpha_{13}\} \wedge \overline{\diamonddot}\{\alpha_6\} \\
\alpha_5 = \neg p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{p\} & \beta_5 = \neg p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\alpha_2\} \\
\alpha_6 = \neg p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{p\} & \beta_6 = \neg p \wedge \diamonddot\{\alpha_3, \alpha_4\} \wedge \overline{\diamonddot}\{\alpha_2, \alpha_4\} \\
\alpha_7 = \neg p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\{p\} & \beta_7 = \neg p \wedge \diamonddot\{\alpha_{10}\} \wedge \overline{\diamonddot}\{\alpha_2\} \\
\alpha_8 = \neg p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\{p\} & \beta_8 = \neg p \wedge \diamonddot\{\alpha_3, \alpha_{10}\} \wedge \overline{\diamonddot}\{\alpha_2\} \\
\alpha_9 = \neg p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\neg p\} & \beta_9 = \neg p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\alpha_{10}\} \\
\alpha_{10} = \neg p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\{\neg p\} & \beta_{10} = \neg p \wedge \diamonddot\{\alpha_{10}, \alpha_9, \alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14}\} \\
 & \qquad \wedge \overline{\diamonddot}\{\alpha_{10}, \alpha_{13}, \alpha_{14}, \alpha_7, \alpha_8\} \\
\alpha_{11} = \neg p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{p, \neg p\} & \beta_{11} = \neg p \wedge \diamonddot\varnothing \wedge \overline{\diamonddot}\{\alpha_4, \alpha_{10}\} \\
\alpha_{12} = \neg p \wedge \diamonddot\{p\} \wedge \overline{\diamonddot}\{p, \neg p\} & \beta_{12} = \neg p \wedge \diamonddot\{\alpha_3\} \wedge \overline{\diamonddot}\{\alpha_4, \alpha_{10}\} \\
\alpha_{13} = \neg p \wedge \diamonddot\{\neg p\} \wedge \overline{\diamonddot}\{p, \neg p\} & \beta_{13} = \neg p \wedge \diamonddot\{\alpha_{10}\} \wedge \overline{\diamonddot}\{\alpha_4, \alpha_{10}\} \\
\alpha_{14} = \neg p \wedge \diamonddot\{p, \neg p\} \wedge \overline{\diamonddot}\{p, \neg p\} & \beta_{14} = \neg p \wedge \diamonddot\{\alpha_3, \alpha_{10}\} \wedge \overline{\diamonddot}\{\alpha_2, \alpha_{10}\}
\end{array}
\tag{5.55}
$$

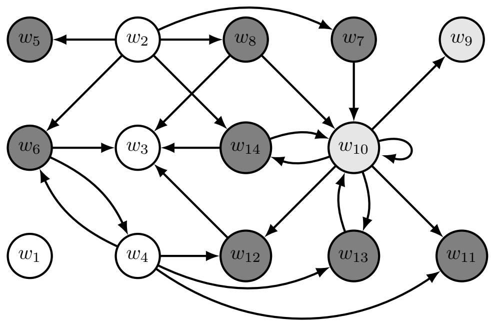



**Figure 5.10:** *Argumentation framework that satisfies the modal minterms from* $\overline{sat}^*_{-1}(\Phi_{\mathcal{CO}})$ *and has a single complete extension*

Let us see the model from Figure 5.10. White background stands for worlds that satisfy $p$ and gray background for those that satisfy $\neg p$. It is not difficult to see that each modal minterm $\beta_i$ is satisfied in this model at world $w_i$. Furthermore, let us see that this model, viewed as an argumentation framework, has a single complete extension. The white background arguments are the elements of the grounded extension, which should be a part of any complete extension. The dark gray arguments are attacked by the grounded extension so they cannot possibly be accepted in order to give rise to additional complete extensions. The two light gray arguments, although not attacked by the grounded extension, cannot be accepted either. Indeed, $w_{10}$ is self-attacking and $w_9$ cannot be defended against the attack from $w_{10}$.

**Lemma 8.** *Any satisfiable formula* $\phi \wedge \mathbf{A}\phi_{\mathcal{CO}} \in ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ *can be satisfied by a model* $\mathfrak{M} = (W, R, V)$ *such that the argumentation framework* $F = (W, R)$ *has a single complete extension.*

*Proof.* The proof is very similar to the one given for Lemma 7 and relies for the base case on the model from Figure 5.10. □

**Theorem 8.** $\mathcal{CO}$ *is the only universally defined argumentation semantics that can be described with a* $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ *formula and also satisfies admissibility, reinstatement and additivity.*

*Proof.* The proof is virtually the same as that of Theorem 7, but using Lemma 8 instead of Lemma 7. □

The immediate implication of this result for argumentation is that none of the following semantics can be described within $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ either: $\mathcal{PR}$, $\mathcal{GR}$, $\mathcal{SST}$, $\mathcal{ID}$, $\mathcal{EAG}$, $\mathcal{GR}^*$, $\mathcal{PR}^*$, $\mathcal{SST}^*$, $\mathcal{ID}^*$, $\mathcal{GR}^{*id}$, $\mathcal{PR}^{cids}$, $\mathcal{SST}^{cids}$, $\mathcal{GR}^{*cids}$, $\mathcal{AD}1$ and $\mathcal{AD}2$.

### 5.4.3 Total converse bisimulation in use

Most of the total bisimulation examples from Subsection 5.3.4 are in fact total converse bisimulations, so they can be reused here for showing that $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ is not expressive enough to describe the corresponding semantics. In order to explain more formally how the results are to be reused, we consider the total bisimulation example from Figure 5.6. The figure depicts the two models $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$ but also the corresponding argumentation frameworks $F = (W, R^-)$ and $F' = (W', R'^-)$. It is not difficult to see that the total bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$ is in fact a total converse bisimulation, so we have $\mathfrak{M} \leftrightarroweq_{tc} \mathfrak{M}'$.

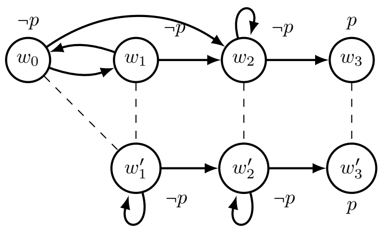



**Figure 5.11:** *$ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ cannot describe $\mathcal{PR}^S$, $\mathcal{SST}^S$ and $\mathcal{GR}^{*S}$*

Let us compare the way we use argumentation-based operators in this section to the approach that was used in the previous section:

$$\mathfrak{M}, w \Vdash \mathcal{S}p \leftrightarrow V^*(p) \in \mathcal{E}_{\mathcal{S}}(W, R^-) \leftrightarrow (W, R^-, V), w \Vdash \overline{\mathcal{S}}p \tag{5.56}$$

What this means is that we can use the frameworks from the figure as models for the converse version of argumentation-based operators. Furthermore, it is quite immediate that:

$$(W, R, V) \leftrightarroweq_{tc} (W', R', V') \Leftrightarrow (W, R^-, V) \leftrightarroweq_{tc} (W', R'^-, V') \tag{5.57}$$

Thus, we can use the models from Figure 5.11 to deduce that $\mathcal{PR}^S$, $\mathcal{SST}^S$ and $\mathcal{GR}^{*S}$ cannot be described within $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. In a similar way we can reuse all the examples from Subsection 5.3.4, except for the last one, whose models are reproduced in Figure 5.12. Let us see that the bisimulation from the figure is not a total converse bisimulation.

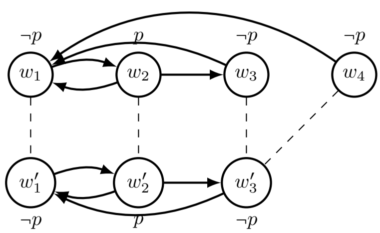



**Figure 5.12:** *Total bisimulation that is not a total converse bisimulation.*

Indeed, we have that $(w_2', w_3') \in R'$ and $(w_4, w_3') \in Z$, where $Z$ stands for the bisimulation depicted in the figure. Thus, based on the zag condition for the converse modality, there should exist $w \in W$ such that $(w, w_2') \in Z$ and $(w, w_4) \in R$. However, in our example this is not the case. On the other hand note that this example was used for the argumentation semantics $\mathcal{CF}^{cr}$, $\mathcal{CF}^{cw}$ and $\mathcal{CO}^{P}$. Of these, the first one actually has a description in $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. The other two fall into the undecided category because they can be described in $\mu$-calculus extended with the global and converse modalities, so they are both invariant with respect to total converse bisimulations.

### 5.4.4 Discussion

While, as shown in the beginning of this section, the converse modality does bring useful additional expressive power for doing argumentation in modal logic, its use is somewhat limited. We did manage to describe an additional semantics and a few sets of useful arguments, but most semantics can still not be captured. However, the use of the converse modality also allows us to use the Kripke models directly as argumentation frameworks, providing a more intuitive setting for the use of modal logic in argumentation.

On the other hand we managed to put the normal form satisfiability introduced in Section 4.5.4 to good use and extended the result from the previous section to $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. The important challenge in this case has been the large number of minterms. Furthermore, the argumentation framework that we generated for the satisfiability of first order modal minterms proved to be significantly larger and more complex than the example from the previous section.

We have also seen that most total bisimulation examples from the previous section were in fact total converse bisimulations, thus leading to the impossibility of describing the corresponding semantics in $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ as well. We provide a complete picture of the definability of argumentation semantics using the global converse modal language in Figure 5.13.

Note that in this case we have two additional semantics that cannot be decided with bisimulations. Both of them can be captured by $\mu$-calculus with converse and global modalities, as follows:

$$\begin{aligned}
\mathcal{CO}^{P}p \leftrightarrow \mathbf{A}((p \to \neg\mu X.(\overline{\Diamond}p \vee \overline{\Diamond\Diamond}X) \wedge \Box\Diamond p) \\
\wedge (\overline{\Box}\overline{\Diamond}p \wedge \neg\mu X.(\overline{\Diamond}p \vee \overline{\Diamond\Diamond}X) \wedge \neg\mu X.(\Diamond p \vee \Diamond\Diamond X) \to p)) \\
\mathcal{CF}^{cw}p \leftrightarrow \mathbf{A}((p \to \neg\mu X.(\overline{\Diamond}p \vee \overline{\Diamond\Diamond}X)) \\
\wedge (\overline{\Box}\overline{\Diamond}p \wedge \neg\mu X.(\overline{\Diamond}p \vee \overline{\Diamond\Diamond}X) \wedge \neg\mu X.(\Diamond p \vee \Diamond\Diamond X) \to p))
\end{aligned} \tag{5.58}$$

## 5.5 Chapter Summary

This chapter has focused on the link between argumentation and modal logic, as initially proposed in (Grossi, 2010). We proved a very general result that links modal logic and the evaluation principles of argumentation semantics, leading to a limitative result concerning the actual use of $ML(\Diamond, \mathbf{E})$ for argumentation. Furthermore, we have extended this result to cover a more expressive language,

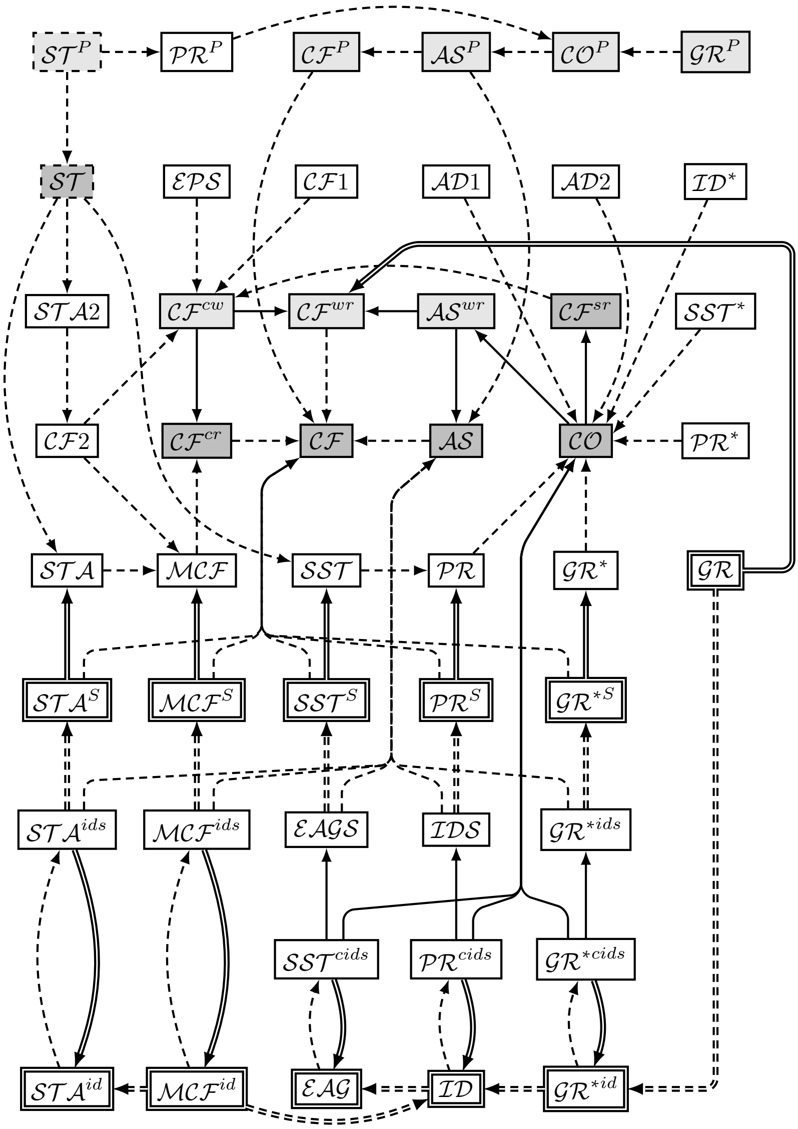



**Figure 5.13:** *Overview of the use of $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ for describing argumentation semantics: dark gray background = can be described, white background = cannot be described, light gray background = undecided*

$ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. We also provided bisimulation-based proofs to show that other argumentation semantics cannot be described by these languages.

In Section 5.1 we provided an overview of the use of modal logic as a meta-language for talking about argumentation semantics, as discussed in (Grossi, 2010).

In Section 5.2 we showed that the modal formulas that describe argumentation semantics can be converted to simpler normal forms with respect to arbitary formulas, especially when the corresponding semantics are additive.

We discussed the use of the global modal language for describing argumentation semantics in Section 5.3. We have extended the results from (Grossi, 2010) with a systematic analysis of first degree formulas and another one based on attack and defense. The latter also gave rise to new argumentation semantics. Furthermore, we provided a general result that links the modal description of argumentation semantics to admissibility, reinstatement, additivity and universality, principles that most argumentation semantics follow. The result is limitative and implies that further use of modal logic as a meta-language for argumentation must either give up one or more of the four principles, or increase the expressiveness of the language. We concluded the section by providing bisimulation examples to show that most argumentation semantics not covered by the general result are still beyond the expressive power of the global modal language. We also pointed out that such bisimulation examples cannot be provided for the argumentation semantics that can be described using a combination of $\mu$-calculus and the global modalities, since this language is invariant with respect to total bisimulations.

We discussed the most straightforward candidate for extending the modal language in Section 5.4, where we focused on $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. We showed that the increase in expressive power is useful for argumentation, by allowing the description of a few additional concepts. However, the general limitative result still holds, as well as most of the bisimulation examples. The $\mu$-calculus formulas presented in sections 5.3 and 5.4 suggest that $\mu$-calculus is able to bring more expressive power to argumentation semantics. On the other hand, all bisimulation examples are still applicable.

*Chapter 6*

# Constrained Argumentation Semantics

In this chapter we focus on the use of logic together with abstract argumentation for applying constraints to argumentation semantics. We discuss the works that have inspired our approach in Section 6.1. In Section 6.2 we introduce 16 novel labeling-based semantics and we prove their most important properties. In Section 6.3 we describe a negotiation scenario modeled using arguments and show that it can benefit from using the semantics we have introduced. We formalize the scenario into an argumentation-based negotiation game in Section 6.4. We conclude the chapter with a summary in Section 6.5.

## 6.1 Preliminaries

In this section we discuss approaches from the argumentation literature that our work is based on or strongly related to. We start with argument labelings in Subsection 6.1.1. Constrained argumentation frameworks are presented in Subsection 6.1.2. We end this section with conditional labelings in Subsection 6.1.3. A comparison of our approach with all these and also with the enhanced preferred semantics will be provided at the end of Section 6.2.

### 6.1.1 Argument labelings

Labelings of argumentation frameworks were introduced in (Caminada, 2006a), but they can also be traced back to earlier works such as (Prakken, 2004) and (Jakobovits and Vermeir, 1999). Whereas the extension-based semantics prescribe sets of arguments that are accepted (extensions), with the implicit assumption that the arguments that are not in the extension are rejected, labelings assign to each argument one of the following labels: `in`, `out`, `undec`.

**Definition 79.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework. A* ***labeling*** *is a* *total function* $\mathcal{L}: \mathcal{A} \rightarrow \{\texttt{in}, \texttt{out}, \texttt{undec}\}$. *We will use the following notations:*

$$\begin{aligned} \texttt{in}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{in}\} \\ \texttt{out}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{out}\} \\ \texttt{undec}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{undec}\} \end{aligned} \tag{6.1}$$

*Alternatively, a labeling* $\mathcal{L}$ *of an argumentation framework* $F = (\mathcal{A}, \mathcal{R})$ *can be seen as a partition of the set* $\mathcal{A}$ *into three sets* $(\texttt{in}(\mathcal{L}), \texttt{out}(\mathcal{L}), \texttt{undec}(\mathcal{L}))$.

An argument labeling can be read as follows:
- the `in` label is assigned to arguments that are accepted
- the `out` label is assigned to arguments that are rejected
- the `undec` label is assigned to arguments for which no decision is taken (their status is undecided)

Just as the extension-based semantics provide restrictions for the sets of arguments that can be accepted, an argument labeling is meaningful if the assigned labels satisfy certain properties, which leads to several kinds of labelings (Caminada and Gabbay, 2009). Here we only focus on complete labelings.

**Definition 80.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework and let* $\mathcal{L}$ *be a labeling of* $F$. $\mathcal{L}$ *is a* ***complete labeling*** *iff, for every argument* $a \in \mathcal{A}$ *it holds that:*

*(a)* $a$ *is labeled* `in` *iff all the attackers of* $a$ *are labeled* `out`

$$\mathcal{L}(a) = \texttt{in} \Leftrightarrow \forall b(b \in \mathcal{A} \textbf{ and } b \rightarrow a \Rightarrow \mathcal{L}(b) = \texttt{out}) \tag{6.2}$$

*(b)* $a$ *is labeled* `out` *iff* $a$ *has an attacker that is labeled* `in`

$$\mathcal{L}(a) = \texttt{out} \Leftrightarrow \exists b(b \in \mathcal{A} \textbf{ and } b \rightarrow a \textbf{ and } \mathcal{L}(b) = \texttt{in}) \tag{6.3}$$

*The set of all complete labelings of* $F$ *is denoted by* $\mathcal{L}_{\mathcal{CO}}(F)$.

Note that the complete labelings were first introduced in (Caminada, 2006a) as reintstatement labelings. We rely on the more recent terminology used in (Caminada and Gabbay, 2009) and (Baroni et al., 2011a).

Let us see that Definition 80 stipulates no explicit constraints for the `undec` label. On the other hand, as mentioned for example in (Caminada and Gabbay, 2009), the conditions from the definition give rise to the following possible cases for a given argument $a$:
- $a$ is labeled `in` and all its attackers are labeled `out`
- $a$ is labeled `out` and one of its attackers is labeled `in` (the other attackers may have any label)
- $a$ is labeled `undec`, none of its attackers are labeled `in` and at least one attacker is labeled `undec`

Thus, we do have some constraints that are implicit for the assignment of the `undec` label.

As an example, consider the argumentation framework from Figure 6.1. The complete labelings of $F$ are:

$$\begin{aligned} &\mathcal{L}_1: \quad \mathcal{L}_1(a) = \texttt{in}, \quad \mathcal{L}_1(b) = \texttt{out}, \quad \mathcal{L}_1(c) = \texttt{in} \\ &\mathcal{L}_2: \quad \mathcal{L}_2(a) = \texttt{out}, \quad \mathcal{L}_2(b) = \texttt{in}, \quad \mathcal{L}_2(c) = \texttt{out} \\ &\mathcal{L}_3: \quad \mathcal{L}_3(a) = \texttt{undec}, \quad \mathcal{L}_3(b) = \texttt{undec}, \quad \mathcal{L}_3(c) = \texttt{undec} \end{aligned} \tag{6.4}$$

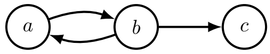



**Figure 6.1:** *Example argumentation framework for labelings.*

Let us see that $\mathcal{L}_1$ is indeed a complete labeling. Argument $a$ is labeled `in` and its only attacker, $b$, is labeled `out`, as required by Definition 80. Furthermore, $b$ is labeled out and one of its attackers, namely $a$ is labeled `in`. Finally, $c$ is labeled `in` and its only attacker, $b$, is labeled `out`. Thus, the conditions from Definition 80 are satisfied and we can conclude that $\mathcal{L}_1$ is indeed a complete labeling. A similar reasoning can be used for showing that $\mathcal{L}_2$ and $\mathcal{L}_3$ are also complete labelings.

On the other hand let us see that the labeling $\mathcal{L}'$ given by

$$\begin{aligned} \mathcal{L}'(a) &= \texttt{out} \\ \mathcal{L}'(b) &= \texttt{undec} \\ \mathcal{L}'(c) &= \texttt{in} \end{aligned} \tag{6.5}$$

violates the completeness conditions imposed in Definition 80. First of all, $a$ is labeled `out` but it has no attacker labeled `in`. Furthermore, $b$ has all its attackers labeled `out`, but is labeled `undec` instead of `in`. Finally, $c$ is labeled `in` but not all its attackers are labeled `out`, as $b$ is labeled `undec`. Thus, $\mathcal{L}'$ is not a complete labeling.

As pointed out in (Caminada, 2006a), argument labelings are strongly related to extension-based semantics. For a start, let us consider the sets of arguments that are labeled `in` by the complete labelings from (6.4): $\texttt{in}(\mathcal{L}_1) = \{a, c\}$, $\texttt{in}(\mathcal{L}_2) = \{b\}$, $\texttt{in}(\mathcal{L}_3) = \varnothing$. These sets are all the extensions prescribed by the complete semantics, i. e. $\mathcal{E}_{\mathcal{CO}}(F) = \{\varnothing, \{a, c\}, \{b\}\}$. This suggests that for going from labelings to extensions it may be enough to consider the set of arguments that are labeled `in`. The less trivial translation is the one from extensions to labelings.

**Definition 81.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. For any conflict-free set $S$ we consider the corresponding labeling* **Ext2Lab**$(S) = \mathcal{L}$*, given by:*

$$\begin{aligned} \texttt{in}(\mathcal{L}) &= S \\ \texttt{out}(\mathcal{L}) &= \{a \in \mathcal{A} \mid S \rightarrow a\} \\ \texttt{undec}(\mathcal{L}) &= \{a \in \mathcal{A} \setminus S \mid S \not\rightarrow a\} \end{aligned} \tag{6.6}$$

For example. if we take $S = \{a, c\}$, which is a complete extension of the framework from Figure 6.1, we have **Ext2Lab**$(S) = \mathcal{L}_1$, the complete labeling that we have discussed already.

**Proposition 27.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{L}$ be a labeling of $F$. $\mathcal{L}$ is a complete labeling iff there exists a complete extension $S \in \mathcal{E}_{\mathcal{CO}}(F)$ such that $\mathcal{L} =$* **Ext2Lab**$(S)$*.*

In words, the result from Proposition 27 implies that every complete extension corresponds to a complete labeling and vice versa. This correspondence can further be used for identifying the complete labelings that correspond to other argumentation semantics that give complete extensions. As it turns out, minimizing or maximizing the set of arguments with some label leads to known extension-based semantics (Caminada, 2006a). The most important results are presented in Table 6.1.

| Constraints | Semantics |
|---|---|
| no constraint | $\mathcal{CO}$ |
| minimal `in` or minimal `out` or maximal `undec` | $\mathcal{GR}$ |
| maximal `in` or maximal `out` | $\mathcal{PR}$ |
| minimal `undec` | $\mathcal{SST}$ |
| empty `undec` | $\mathcal{ST}$ |

**Table 6.1:** *Extension-based semantics that can be obtained by imposing constraints on complete labelings.*

For example. semi-stable extensions correspond to complete labelings that have their `undec`-labeled part minimal with respect to set inclusion. In our case, these labelings are $\mathcal{L}_1$ and $\mathcal{L}_2$, so both $\{a, c\}$ and $\{b\|$ are semi-stable extensions of the argumentation framework from Figure 6.1. The grounded extension, on the other hand, corresponds to the labeling that has minimal `in`, minimal `out` and maximal `undec` labeled parts. Any of these conditions leads to the same labeling, namely $\mathcal{L}_3$, which does indeed correspond to the grounded extension of $F$, the empty set.

Another property of complete labelings that has inspired our approach is the fact that a complete labeling is uniquely defined by its `in`-labeled part, as well as by its `out`-labeled part (Caminada and Gabbay, 2009), in the sense of Proposition 28.

**Proposition 28.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{L}_1, \mathcal{L}_2 \in \mathcal{L}_{\mathcal{CO}}(F)$ be two complete labelings of $F$.*
*(a) if* $\mathtt{in}(\mathcal{L}_1) = \mathtt{in}(\mathcal{L}_2)$ *then* $\mathcal{L}_1 = \mathcal{L}_2$
*(b) if* $\mathtt{out}(\mathcal{L}_1) = \mathtt{out}(\mathcal{L}_2)$ *then* $\mathcal{L}_1 = \mathcal{L}_2$

Note that a similar property does not hold for the `undec`-labeled part, as we have for our example that $\mathtt{undec}(\mathcal{L}_1) = \mathtt{undec}(\mathcal{L}_2) = \varnothing$, yet the two labelings are distinct.

Other approaches have been proposed in the literature aside from the three-labels approach we have presented here. A notable example consists in using real values from the interval $[0, 1]$ and is part of the equational approach presented in (Gabbay, 2009b) in the larger context of fibring argumentation frames. We have also considered numerical labels in (Gratie and Florea, 2012b), where we provide an algorithm for approximating numerical complete labelings of an argumentation framework.

### 6.1.2 Constrained argumentation frameworks

In this subsection we discuss constrained argumentation frameworks, as proposed in (Coste-Marquis et al., 2006). We provide and discuss some of the formal definitions here, as constraints play an important role in our approach as well.

**Definition 82.** *We will use $PL(\mathcal{P}rop)$ to refer to the propositional language defined by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \tag{6.7}$$

*where $p$ ranges over the set of propositional symbols $\mathcal{P}rop$.*

Note that, in contrast with the definitions of modal languages from previous chapters, we have included the set of propositional symbols $\mathcal{P}rop$ to the notation. This is due to the fact that the actual set of propositional symbols will be the same as the set of arguments, so it will vary across argumentation frameworks.

**Definition 83.** *A **constrained argumentation framework** is defined as a tuple $F = (\mathcal{A}, \mathcal{R}, \mathcal{C})$, where $\mathcal{A}$ is a finite set of arguments, $\mathcal{R}$ is a binary attack relation on $\mathcal{A}$ and $\mathcal{C}$ is a propositional formula from $PL(\mathcal{A})$.*

In other words, a constrained argumentation framework is simply an argumentation framework $(\mathcal{A}, \mathcal{R})$ that is augmented with a propositional formula $\mathcal{C}$ that uses the arguments of the framework as propositional symbols.

Such formulas can be used for encoding constraints that can be imposed on the extensions of the argumentation framework. In (Coste-Marquis et al., 2006) the authors define the completion of a set of arguments $S$ as $\hat{S} = \{a \mid a \in S\} \cup \{\neg a \mid a \in \mathcal{A} \setminus S\}$ and require that $\hat{S} \Vdash \mathcal{C}$. We provide an equivalent formalization of satisfaction, which is closer to what we will use in the next section.

**Definition 84.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\mathcal{C} \in PL(\mathcal{A})$ be a formula that uses propositional symbols from $\mathcal{A}$, i.e. $(\mathcal{A}, \mathcal{R}, \mathcal{C})$ is a constrained argumentation framework. Let $S \subseteq A$ be a set of arguments. We recursively define the **satisfaction** of $\mathcal{C}$ by $S$ in $F$ (notation $S \Vdash_F \mathcal{C}$) as follows:*

$$\begin{aligned}
&S \Vdash_F \top \\
&S \nVdash_F \bot \\
&S \Vdash_F a \Leftrightarrow a \in S, \text{ for all } a \in \mathcal{A} \\
&S \Vdash_F \neg\phi \Leftrightarrow S \nVdash_F \phi, \text{ for all } \phi \in PL(\mathcal{A}) \\
&S \Vdash_F \phi \wedge \psi \Leftrightarrow S \Vdash_F \phi \textbf{ and } S \Vdash_F \psi, \text{ for all } \phi, \psi \in PL(\mathcal{A}) \\
&S \Vdash_F \phi \vee \psi \Leftrightarrow S \Vdash_F \phi \textbf{ or } S \Vdash_F \psi, \text{ for all } \phi, \psi \in PL(\mathcal{A})
\end{aligned} \tag{6.8}$$

*We will drop the subscript $F$ whenever the argumentation framework is implicit. Whenever $S \Vdash \phi$, we will also say that $S$ is a **model** of $\phi$.*

Note that there is a significant distinction between the use of logic formulas and satisfaction here and their use in previous chapters. Indeed, in the previous chapters we have used modal formulas as a meta-language for describing argumentation semantics, with sets of arguments corresponding to propositional symbols or to the valuation of formulas. Furthermore, satisfaction was defined with respect to a world in the model (in our case, an argument). Here, on the other hand, propositional symbols correspond to arguments (and we have exactly one propositional symbol for each argument), while satisfaction is defined with respect to a set of arguments.

We now use our formulation of satisfaction and follow the approach from (Coste-Marquis et al., 2006) to introduce $\mathcal{C}$-admissible sets.

**Definition 85.** *Let $F = (\mathcal{A}, \mathcal{R}, \mathcal{C})$ be a constrained argumentation framework. A set of arguments $S \subseteq \mathcal{A}$ is a $\mathcal{C}$-admissible set iff it is an admissible set of the argumentation framework $(\mathcal{A}, \mathcal{R})$ and, in addition, $S \Vdash_{(\mathcal{A},\mathcal{R})} \mathcal{C}$.*

In other words, $\mathcal{C}$-admissible sets are admissible sets for which we impose some constraints with respect to which arguments are to be included or excluded from the set. As an example, let us consider two constrained frameworks based on Figure 6.2.

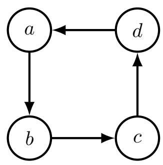



**Figure 6.2:** *Example argumentation framework for the satisfaction of propositional constraints.*

$$\begin{aligned} F_1 &= (\mathcal{A}, \mathcal{R}, \mathcal{C}_1), \quad \mathcal{C}_1 = a \wedge \neg b \\ F_2 &= (\mathcal{A}, \mathcal{R}, \mathcal{C}_2), \quad \mathcal{C}_2 = a \wedge \neg c \vee \neg a \wedge c \end{aligned} \tag{6.9}$$

The admissible sets of the original, non-constrained, framework $F = (\mathcal{A}, \mathcal{R})$ from Figure 6.2 are $\mathcal{E}_{\mathcal{AS}}(F) = \{\varnothing, \{a, c\}, \{b, d\}\}$. From this, it is easy to see that the only $\mathcal{C}_1$-admissible set of $F_1$ is $\{a, c\}$, while $F_2$ has no $\mathcal{C}_2$-admissible set.

**Definition 86.** *A constrained argumentation framework $F = (\mathcal{A}, \mathcal{R}, \mathcal{C})$ is **consistent** if it has at least one $\mathcal{C}$-admissible set.*

Thus, $F_1$ is consistent, while $F_2$ is not. As we can see, it is rather easy to get no $\mathcal{C}$-admissible set for a constrained argumentation framework. Since the other semantics introduced in (Coste-Marquis et al., 2006) for constrained argumentation frameworks correspond to the classical semantics, but are defined with respect to $\mathcal{C}$-admissible sets instead of admissible sets, it will be even more probable that the approach will produce no extension for many cases.

In our approach, we assume that we can alter the argumentation framework by adding arguments and attacks and we show what one can do in order to satisfy any reasonable constraint, including some of the constraints that are not satisfied by any admissible set, such as $\mathcal{C}_2$ above. Furthermore, we will separate the propositional formula from the framework and associate it to the semantics, leading to parameterized semantics for the same argumentation framework.

### 6.1.3 Conditional labelings

Another approach that is related to our work is that from (Boella et al., 2011), where logic and labelings are used together. The idea is to find sets of arguments that should be attacked in order for a specific argument to be labeled `in`, `out` or `undec`. The reference labeling is taken to be the grounded one and can be defined either as **Ext2Lab**$(E)$, where $E$ is the grounded extension, or as the complete labeling that has minimal `in` or minimal `out` or maximal `undec`-labeled parts. All these characterizations follow from Subsection 6.1.1.

The approach presented in (Boella et al., 2011) is based on the idea that each argument $a$ from an argumentation framework can be assigned three conditional labels with the following meaning:

- $a^+$ – which arguments should be attacked in order for $a$ to be `in`?
- $a^-$ – which arguments should be attacked in order for $a$ to be `out`?
- $a^?$ – which arguments should be attacked in order for $a$ to be `undec`?

Each of the three labels is in fact a propositional formula that encodes the answer to the corresponding question.

**Definition 87.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. The corresponding language for conditional labelings is given by the following BNF:*

$$\phi ::= \top \mid \bot \mid a^0 \mid a^+ \mid a^- \mid a^? \mid \phi \wedge \phi \mid \phi \vee \phi \tag{6.10}$$

*where $a$ ranges over the set of arguments $\mathcal{A}$.*

The meaning of the atomic formulas with respect to the satisfaction of the proposed goal (the label of some argument) is as follows:

- $\top$: the desired label is the same as the current label, nothing needs to be done
- $\bot$: nothing can be done in order to ensure the desired label – for example a self-attacking argument cannot be `in`
- $a^0$: $a$ should be attacked in order for the goal to be satisfied
- $a^+$: $a$ should be `in` in order for the goal to be satisfied
- $a^-$: $a$ should be `out` in order for the goal to be satisfied
- $a^?$: $a$ should be `undec` in order for the goal to be satisfied

Thus, the formula that corresponds to a conditional label can be computed starting from the properties of complete labelings and then using substitution to replace occurrences of other conditional labels. We explain the approach in more detail on the example from Figure 6.3.

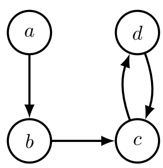



**Figure 6.3:** *Example argumentation framework for conditional labelings.*

For any argumentation framework and any argument $a$ we can assign the following conditional labelings, that are based on the completeness conditions from Definition 80 and their implications for the `undec` label, as discussed in Subsection 6.1.1:

$$\begin{aligned} a^+ :&\quad \bigwedge\{b^- \mid b \to a\} \\ a^- :&\quad a^0 \vee \bigvee\{b^+ \mid b \to a\} \\ a^? :&\quad \bigvee\{b^? \mid b \to a\} \wedge \bigwedge\{b^? \vee b^- \mid b \to a\} \end{aligned} \tag{6.11}$$

For example, in order to ensure that $a$ is `out`, $a$ must be attacked directly or the right arguments are to be attacked in order to ensure that one of $a$'s attackers is `in`. The relations from (6.11) are in fact a rewriting of similar relations presented in (Boella et al., 2011).

For our example framework from Figure 6.3, we get the following conditional labels:

$$\begin{array}{lll} a^+ : \top & a^- : a^0 & a^? : \bot \\ b^+ : a^- & b^- : b^0 \vee a^+ & b^? : a^? \\ c^+ : b^- \wedge d^- & c^- : c^0 \vee b^+ \vee d^+ & c^? : (b^? \vee d^?) \wedge (b^? \vee b^-) \wedge (d^? \vee d^-) \\ d^+ : c^- & d^- : d^0 \vee c^+ & d^? : c^? \end{array} \tag{6.12}$$

Note that we have also used some simplifications, based on equivalences such as $\phi \wedge \bot \leftrightarrow \phi$. The conditional labels that we obtained answer questions that are slightly different from those that we presented in the beginning of this subsection. Indeed, these labels tell us what should happen in the rest of the framework in order for the label of some argument to have a specified value, but they do not tell us exactly the arguments that should be attacked. This happens because the formulas also contain other conditional labels.

Thus, the real challenge consists in converting these formulas to a form defined by the following BNF:

$$\begin{aligned} \phi &::= \top \mid \bot \mid \psi \\ \psi &::= \xi \mid \psi \vee \psi \\ \xi &::= a^0 \mid \xi \wedge \xi \end{aligned} \tag{6.13}$$

where $a$ ranges over the set of arguments $\mathcal{A}$. This is in fact a disjunctive normal form that can be easily translated to sets of arguments that should be attacked for satisfying the corresponding goal of the conditional label.

In order to reach this form, we can use substitution to replace conditional labels that occur within formulas with their actual values. This approach can be applied repeatedly, together with simplifications, and leads to the desired form for acyclic frameworks. However, problems arise in the presence of cycles. The following rule is proposed in (Boella et al., 2011) for dealing with cycles: if $a^i$ appears in the body of $a^j$, where $i, j \in \{+, -, ?\}$, then:

- if $i = j = ?$ then $a^i$ can be substituted with $\top$
- otherwise, $a^i$ is substituted with $\bot$

We will make use of this rule, as our example does contain a cycle. First of all note that there is no substitution needed for $a$'s labels, as they are already in the desired form. The labels tell us that $a$ is already `in`, that it can be turned `out` only by directly attacking it and also that it cannot be turned `undec` by attacking arguments from the framework. For $b$'s labels we have, after the substitution:

$$\begin{aligned} & b^+ : a^0 \\ & b^- : b^0 \vee \top \leftrightarrow \top \\ & b^? : \bot \end{aligned} \tag{6.14}$$

The meaning of $b$'s labels is that $b$ can be accepted if $a$ is defeated, $b$ is already rejected and $b$ can not become undecided just by attacking arguments from the framework. Let us now substitute $b$'s labels in the formulas for $c$'s conditional labels:

$$\begin{aligned} &c^+ : \top \wedge d^- \leftrightarrow d^- \\ &c^- : c^0 \vee a^0 \vee d^+ \\ &c^? : (\bot \vee d^?) \wedge (\bot \vee \top) \wedge (d^? \vee d^-) \leftrightarrow d^? \end{aligned} \tag{6.15}$$

Now we can substitute $d$'s labels as well and get the following:

$$\begin{aligned} &c^+ : d^0 \vee c^+ \leftrightarrow d^0 \vee \bot \leftrightarrow d^0 \\ &c^- : c^0 \vee a^0 \vee c^- \leftrightarrow c^0 \vee a^0 \vee \bot \leftrightarrow c^0 \vee a^0 \\ &c^? : c^? \leftrightarrow \top \end{aligned} \tag{6.16}$$

This means that $c$ can be accepted if $d$ is attacked, $c$ can be rejected if it is directly attacked or if $a$ is attacked. It also means that $c$ is already undecided, so nothing needs to be done in this case. Last, we substitute $c$'s labels into $d$'s formulas:

$$\begin{aligned} &d^+ : c^0 \vee a^0 \\ &d^- : d^0 \vee d^0 \leftrightarrow d^0 \\ &d^? : \top \end{aligned} \tag{6.17}$$

This means that $d$ can be accepted if either $a$ or $c$ are defeated, $d$ can be rejected if it is directly attacked and $d$ is already undecided. With this we have completed the analysis of the example framework.

## 6.2 Constrained Argumentation Semantics

In this section we propose several labeling-based semantics that can deal with the challenges of using argumentation for multi-agent systems. Our approach combines ideas from the enhanced preferred semantics and constrained argumentation frameworks in order to provide labelings that show which arguments need to be attacked in order to satisfy certain constraints. The work presented here was published in (Gratie and Florea, 2012a). We first introduce the concept of open labelings in Subsection 6.2.1. We then define the constrained labelings in Subsection 6.2.2, where we also discuss some of their properties. A comparison with other approaches from the argumentation literature is provided in Subsection 6.2.3.

### 6.2.1 Open labelings

We have seen in Subsection 2.8 that the enhanced preferred semantics is based on ignoring a number of arguments so that admissible sets are available. We apply the same idea for labelings by enriching the usual set of labels $\{\texttt{in}, \texttt{out}, \texttt{undec}\}$ with an additional label, $\texttt{ign}$, which stands for arguments that are ignored.

On the other hand, in contrast with the enhanced preferred semantics approach, we feel that working with extensions that are computed for subframeworks and are not admissible with respect to the full framework is somewhat questionable. Thus, our semantics will only refer to labelings, i.e. the $\texttt{in}$-labeled part is not supposed to be used separately. Instead, we regard the set of ignored arguments as the set that should be attacked in order to obtain the desired labels for the other arguments.

In a multi-agent scenario, where an agent constantly receives new information about the world, from various sources and with various degrees of reliability, it may even be reasonable to actually ignore some of the arguments, for example in the case when a desired labeling cannot occur only because of arguments coming from the same source. We will revisit these aspects at the end of this section.

We start by introducing open labelings, which correspond to complete labelings of the subframeworks of an argumentation framework.

**Definition 88.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an arbitrary argumentation framework. An* ***open labeling*** *is a total function $\mathcal{L} : \mathcal{A} \to \{\texttt{in}, \texttt{out}, \texttt{undec}, \texttt{ign}\}$ such that, for any argument $a \in \mathcal{A}$:*

*(a) $a$ is labeled* $\texttt{in}$ *iff all the attackers of $a$ are labeled either* $\texttt{out}$ *or* $\texttt{ign}$

$$\mathcal{L}(a) = \texttt{in} \Leftrightarrow \forall b(b \in \mathcal{A} \textit{ and } b \to a \Rightarrow \mathcal{L}(b) = \texttt{out} \textit{ or } \mathcal{L}(b) = \texttt{ign}) \qquad (6.18)$$

*(b) $a$ is labeled* $\texttt{out}$ *iff $a$ has an attacker that is labeled* $\texttt{in}$

$$\mathcal{L}(a) = \texttt{out} \Leftrightarrow \exists b(b \in \mathcal{A} \textit{ and } b \to a \textit{ and } \mathcal{L}(b) = \texttt{in}) \qquad (6.19)$$

*The set of all open labelings of $F$ is denoted by $\mathcal{L}_{\mathcal{O}}(F)$. We will also use the following notations:*

$$\begin{aligned} \texttt{in}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{in}\} \\ \texttt{out}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{out}\} \\ \texttt{undec}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{undec}\} \\ \texttt{ign}(\mathcal{L}) &= \{a \in \mathcal{A} \mid \mathcal{L}(a) = \texttt{ign}\} \end{aligned} \qquad (6.20)$$

Note that an open labeling of an argumentation framework can also be seen as a partition of the set of arguments into four sets. Thus, we will often write a labeling $\mathcal{L}$ as $(\texttt{in}(\mathcal{L}), \texttt{out}(\mathcal{L}), \texttt{undec}(\mathcal{L}), \texttt{ign}(\mathcal{L}))$. As mentioned already, an open labeling can also be seen as a complete labeling of a certain subframework:

$$\mathcal{L} \in \mathcal{L}_{\mathcal{O}}(F) \Leftrightarrow (\texttt{in}(\mathcal{L}), \texttt{out}(\mathcal{L}), \texttt{undec}(\mathcal{L})) \in \mathcal{L}_{\mathcal{CO}}(F{\downarrow}_{\mathcal{A} \setminus \texttt{ign}(\mathcal{L})}) \qquad (6.21)$$

Let us consider the framework from Figure 6.4. The open labelings of this framework, grouped by the set of ignored arguments, is:

$$\begin{aligned}
&(\varnothing,\varnothing,\{a,b,c\},\varnothing),(\{a\},\{b,c\},\varnothing,\varnothing),(\{b\},\{a,c\},\varnothing,\varnothing),\\
&(\{b\},\{c\},\varnothing,\{a\}),\\
&(\{a\},\{c\},\varnothing,\{b\}),\\
&(\varnothing,\varnothing,\{a,b,\},\{c\}),(\{a\},\{b\},\varnothing,\{c\}),(\{b\},\{a\},\varnothing,\{c\}),\\
&(\{c\},\varnothing,\varnothing,\{a,b\}),\\
&(\{b\},\varnothing,\varnothing,\{a,c\}),\\
&(\{a\},\varnothing,\varnothing,\{b,c\}),\\
&(\varnothing,\varnothing,\varnothing,\{a,b,c\})
\end{aligned} \tag{6.22}$$

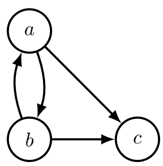



**Figure 6.4:** *Example argumentation framework for open labelings.*

Every complete labeling of an argumentation framework corresponds to an open labeling that ignores no argument. Thus, our approach subsumes complete labelings and generates significantly more labelings. This can also be seen from our example. Moreover, having a more general class of labelings, we expect to be able to enforce stronger properties for the arguments that are not ignored.

First, we focus on the set of undecided arguments. It is known that complete labelings that have no undecided argument (stable labelings) correspond to stable extensions. Since the stable semantics is not universally defined, it may happen that an argumentation framework has no stable labeling. On the other hand, we will see shortly that open labelings with no undecided argument always exist. In other words, we can require that all arguments that are not ignored are either `in` or `out` and still be sure to get at least one labeling.

**Definition 89.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework. An open labeling $\mathcal{L} \in \mathcal{L}_O(F)$ is said to be* ***decided*** *iff it has no* `undec`*-labeled argument. The set of all decided labelings of an argumentation framework $F$ is denoted by $\mathcal{L}_\mathcal{D}(F)$.*

$$\mathcal{L}_\mathcal{D}(F) = \{\mathcal{L} \in \mathcal{L}_\mathcal{O}(F) \mid \texttt{undec}(\mathcal{L}) = \varnothing\} \tag{6.23}$$

For the framework from Figure 6.4 only two of the open labelings are not decided, namely $(\varnothing,\varnothing,\{a,b,c\},\varnothing)$ and $(\varnothing,\varnothing,\{a,b\},\{c\})$. All the other open labelings are also decided. Let us see that decided labelings exist for any argumentation framework.

**Proposition 29.** *For any argumentation framework $F = (\mathcal{A}, \mathcal{R})$, the following open labelings are decided:*

*(a)* $\mathcal{L} = (S, \varnothing, \varnothing, \mathcal{A} \setminus \mathcal{S})$, *for all* $S \in \mathcal{E}_{\mathcal{CF}}(F)$.
*(b)* $\mathcal{L} = (\mathtt{in}(\mathcal{L}_c), \mathtt{out}(\mathcal{L}_c), \varnothing, \mathtt{undec}(\mathcal{L}_c))$, *for all* $\mathcal{L}_c \in \mathcal{L}_{\mathcal{CO}}(F)$.

*Proof.* (a) We only need to show that the labeling is open. This follows from the fact that the restriction of $F$ to any conflict-free set is an argumentation framework that only consists of isolated arguments and the only complete extension of such a framework is the set of all its arguments.

(b) In order to see that $\mathcal{L}$ is an open labeling in this case, compare Definition 88 with Definition 79 and note that taking a complete labeling and changing `undec` to `ign` does indeed lead to an open labeling. □

While the proof of Proposition 29 shows that we can easily get decided labelings from complete labelings simply by ignoring the undecided arguments, let us see that there are decided labelings whose `in` and `out`-labeled sets are not the same as those of any complete labeling of the same framework. We return to our example from Figure 6.4 and the open labelings from (6.22). The complete labelings of $F$ are $(\varnothing, \varnothing, \{a, b, c\})$, $(\{a\}, \{b, c\}, \varnothing)$ and $(\{b\}, \{a, c\}, \varnothing)$. On the other hand, note that $(\{c\}, \varnothing, \varnothing, \{a, b\})$ is a decided labeling where the set of `in`-labeled arguments is distinct from that of any complete labeling. Similarly, $(\{a\}, \{c\}, \varnothing, \{b\})$ is a decided labeling where the set of `out`-labeled arguments is not the same as that of any complete labeling.

We know from Proposition 28 that complete labelings are uniquely defined by their `in`-labeled part, as well as by their `out`-labeled part. On the other hand, let us see that this is not the case for open labelings. Indeed, for our example there are 4 labelings that have their `in`-labeled part equal to $\{a\}$: $(\{a\}, \{b, c\}, \varnothing, \varnothing)$, $(\{a\}, \{c\}, \varnothing, \{b\})$, $(\{a\}, \{b\}, \varnothing, \{c\})$ and $(\{a\}, \varnothing, \varnothing, \{b, c\})$. Furthermore, there are 2 open labelings that have their `out`-labeled part equal to $\{c\}$: $(\{a\}, \{c\}, \varnothing, \{b\})$ and $(\{b\}, \{c\}, \varnothing, \{a\})$. Also, note that all these open labelings are in fact decided as well.

Furthermore, for each set of `ign`-labeled arguments it is possible to have several open labelings (one for each complete labeling of the corresponding subframework). Also, we have seen that there may be several complete labelings that have the same `undec`-labeled part. Thus, none of the four sets that correspond to an open labeling does uniquely define it.

In what follows, we focus on open labelings that are uniquely defined by their `ign`-labeled part. We will call them unique labelings.

**Definition 90.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework. An open labeling* $\mathcal{L} \in \mathcal{L}_{\mathcal{O}}(F)$ *is said to be* ***unique*** *iff no other open labeling* $\mathcal{L}'$ *of* $F$ *has* $\mathtt{ign}(\mathcal{L}') = \mathtt{ign}(\mathcal{L})$. *The set of all unique labelings of* $F$ *is denoted by* $\mathcal{L}_{\mathcal{U}}(F)$.

$$\mathcal{L}_{\mathcal{U}}(F) = \{\mathcal{L} \in \mathcal{L}_{\mathcal{O}}(F) \mid \forall \mathcal{L}'(\mathcal{L}' \in \mathcal{L}_{\mathcal{O}}(F) \textit{ and } \mathtt{ign}(\mathcal{L}') = \mathtt{ign}(\mathcal{L}) \Rightarrow \mathcal{L}' = \mathcal{L})\} \quad (6.24)$$

We argue that unique labelings are useful in a multi-agent scenario. Indeed, an agent that uses open labelings has to attack all the ignored arguments in order to bring about the complete labeling of the corresponding subframework. If the chosen open labeling is unique, the agent is sure that upon attacking the ignored arguments the new framework will have a single complete extension. Thus, all the other agents from the system will draw the same conclusions from the framework (provided, of course, that they use some argumentation semantics that yields complete extensions).

Let us note that unique labelings exist for any argumentation framework, since the open labelings from Proposition 29 (a) are also unique. In the general case, on the other hand, uniqueness and decidedness are distinct notions.

**Proposition 30.** *Not all decided labelings are unique and not all unique labelings are decided.*

*Proof.* It suffices to find an example labeling for each of the two claims. First, let us see that $(\{a\},\{b,c\},\emptyset,\emptyset)$ is a decided labeling of the framework from Figure 6.4 and it is not unique, as $(\{b\},\{a,c\},\emptyset,\emptyset)$ is an open labeling of the same framework. On the other hand all the open labelings of $F$ that are not decided are not unique either.

For the second claim we consider the argumentation framework from Figure 6.5, call it $F'$. We have that $(\emptyset,\emptyset,\{a,b,c\},\emptyset)$ is a unique labeling of $F'$ and it is not decided. □

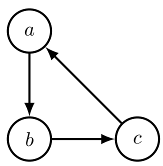



**Figure 6.5:** *Not all unique labelings are decided.*

In fact, let us see that the decided labelings correspond to stable labelings of subframeworks, whereas the unique labelings correspond to grounded labelings of subframeworks that have a single complete labeling.

**Definition 91.** *Let $F = (\mathcal{A},\mathcal{R})$ be an argumentation framework. An open labeling $\mathcal{L} \in \mathcal{L}_{\mathcal{O}}(F)$ is said to be **strict** iff $\mathcal{L}$ is both decided and unique. The set of all strict labelings of an argumentation framework $F$ is denoted by $\mathcal{L}_{\mathcal{S}}(F)$.*

$$\mathcal{L}_{\mathcal{S}}(F) = \mathcal{L}_{\mathcal{D}}(F) \cap \mathcal{L}_{\mathcal{U}}(F) \tag{6.25}$$

Again, we can rely on Proposition 29 (a) to see that strict labelings exist for any argumentation framework. Let us consider the argumentation framework from Figure 6.6 and compute the labelings for all the four semantics we have introduced so far. The results are presented in Table 6.2, the first four columns.

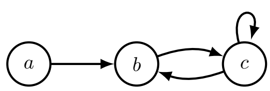



**Figure 6.6:** *Example argumentation framework for open, decided, unique and strict labelings.*

| $\mathcal{L}_{\mathcal{O}}$ | $\mathcal{L}_{\mathcal{D}}$ | $\mathcal{L}_{\mathcal{U}}$ | $\mathcal{L}_{\mathcal{S}}$ | $\mathcal{L}_{\mathcal{MO}}$ | $\mathcal{L}_{\mathcal{MD}}$ | $\mathcal{L}_{\mathcal{MU}}$ | $\mathcal{L}_{\mathcal{MS}}$ |
|---|---|---|---|---|---|---|---|
| $(\{a\},\{b\},\{c\},\varnothing)$ | | ✓ | | ✓ | | ✓ | |
| $(\{b\},\{c\},\varnothing,\{a\})$ | ✓ | | | | ✓ | | |
| $(\varnothing,\varnothing,\{b,c\},\{a\})$ | | | | | | | |
| $(\{a\},\varnothing,\{c\},\{b\})$ | | ✓ | | | | | |
| $(\{a\},\{b\},\varnothing,\{c\})$ | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| $(\varnothing,\varnothing,\{c\},\{a,b\})$ | | ✓ | | | | | |
| $(\{a\},\varnothing,\varnothing,\{b,c\})$ | ✓ | ✓ | ✓ | | | | ✓ |
| $(\{b\},\varnothing,\varnothing,\{a,c\})$ | ✓ | ✓ | ✓ | | | | |
| $(\varnothing,\varnothing,\varnothing,\{a,b,c\})$ | ✓ | ✓ | ✓ | | | | |

**Table 6.2:** *Labelings for the framework from Figure 6.6*

Note that the trivial open labeling that ignores all arguments, $(\varnothing,\varnothing,\varnothing,\mathcal{A})$, is both decided and unique. However, such a labeling has little practical use. In order to filter out this labeling and also other labelings that ignore too many arguments, we can refine our semantics by requiring that the set of ignored arguments is minimal in some sense. Note that we prefer minimization with respect to set inclusion, in constrast with the enhanced preferred semantics, where minimization was performed with respect to cardinality.

**Definition 92.** *Let $\mathcal{L}$ be a $\delta$ labeling of an argumentation framework $F$, where $\delta \in \{$open, decided, unique, strict$\}$. Then $\mathcal{L}$ is said to be a **minimal** $\delta$ labeling iff $\mathtt{ign}(\mathcal{L})$ is minimal (with respect to set inclusion) among $\mathtt{ign}(\mathcal{L}')$ of all $\delta$ labelings $\mathcal{L}'$ of $F$. The set of all minimal $\delta$ labelings of $F$ is denoted by $\mathcal{L}_{\mathcal{MO}}(F)$, $\mathcal{L}_{\mathcal{MD}}(F)$, $\mathcal{L}_{\mathcal{MU}}(F)$ and $\mathcal{L}_{\mathcal{MS}}(F)$ respectively.*

The last four columns of Table 6.2 correspond to the minimal versions of the open, decided, unique and strict labelings, respectively, computed for the argumentation framework from Figure 6.6. Note that, since any argumentation framework has at least one complete labeling, the minimal open labelings will be the ones that ignore no argument, i.e. will correspond to complete labelings.

We have seen that open labelings can help an agent identify the set of arguments it must defeat in order to bring about a certain labeling. For now, open labelings simply require a complete labeling of the arguments that are not ignored, while decided, unique and strict labelings provide additional restrictions. Going back to our agent, it is clear that whenever the same constraints can be satisfied by defeating a set of arguments or one of its subsets, the subset is preferable, as it requires less effort from the agent.

The most important property of the eight labeling-based semantics that we have introduced is that they are all universally defined, i.e. they provide at least one labeling for any argumentation framework.

### 6.2.2 Constrained labelings

In the previous subsection we have introduced several labeling-based semantics based on ignoring arguments in order to enforce other properties. We shall see now that we can also apply constraints on the labels of individual arguments, in a manner similar to that used in (Coste-Marquis et al., 2006) for constrained argumentation frameworks. Our approach is also related to conditional labelings (Boella et al., 2011). However, in contrast with both approaches, we consider that the constraints should specifically require that an argument is accepted or that it is rejected. In other words, we do not allow constraints that require an argument to be undecided.

This being said, the propositional language that we will use for constraints is slightly different.

**Definition 93.** *We will use $PL(\mathcal{P}rop)_{NNF}$ to refer to the propositional language defined by the following BNF:*

$$\phi ::= \top \mid \bot \mid p \mid \neg p \mid \phi \wedge \phi \mid \phi \vee \phi \tag{6.26}$$

*where $p$ ranges over the set of propositional symbols $\mathcal{P}rop$.*

Note that the propositional language from Definition 93 is in fact the negation normal form of the one used in (Coste-Marquis et al., 2006) and discussed in Subsection 6.1.2.

**Definition 94.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\phi$ be a propositional formula in negation normal form, using propositional symbols from $\mathcal{A}$, i.e. $\phi \in PL(\mathcal{A})_{NNF}$. Let $\mathcal{L} \in \mathcal{L}_{\mathcal{O}}(F)$ be an open labeling of $F$. We recursively define the* ***arg-satisfaction*** *of $\phi$ by $\mathcal{L}$ in $F$ (notation $\mathcal{L} \Vdash_F \phi$) as follows:*

$$\begin{aligned}
&\mathcal{L} \Vdash_F \top\\
&\mathcal{L} \nVdash_F \bot\\
&\mathcal{L} \Vdash_F a \Leftrightarrow a \in \mathtt{in}(\mathcal{L}),\ \textit{for all } a \in \mathcal{A}\\
&\mathcal{L} \Vdash_F \neg a \Leftrightarrow a \in \mathtt{out}(\mathcal{L}) \textit{ or } a \in \mathtt{ign}(\mathcal{L}),\ \textit{for all } a \in \mathcal{A}\\
&\mathcal{L} \Vdash_F \phi \wedge \psi \Leftrightarrow \mathcal{L} \Vdash_F \phi \textit{ and } \mathcal{L} \Vdash_F \psi,\ \textit{for all } \phi, \psi \in PL(\mathcal{A})_{NNF}\\
&\mathcal{L} \Vdash_F \phi \vee \psi \Leftrightarrow \mathcal{L} \Vdash_F \phi \textit{ or } \mathcal{L} \Vdash_F \psi,\ \textit{for all } \phi, \psi \in PL(\mathcal{A})_{NNF}
\end{aligned} \tag{6.27}$$

*We will drop the subscript $F$ whenever the argumentation framework is implicit. Whenever $\mathcal{L} \Vdash_\phi$ we will also say that $\mathcal{L}$* ***arg-satisfies*** *$\phi$ or that $\mathcal{L}$ is a* ***model*** *of $\phi$.*

From Definition 94 it is clear that our notion of satisfaction is rather distinct from the one presented in (Coste-Marquis et al., 2006), in the sense that the negation of a propositional argument means that we want the corresponding argument to be `out` or `ign`, which is not the same as requiring that the argument is not `in`, as we also exclude the possibility of it being `undec`. Note that we have used the term arg-satisfaction in Definition 94, as we will need to distinguish it from the usual notion of satisfaction from propositional logic.

Let us now look at the framework from Figure 6.7 and see that the following relations hold:

$$\begin{aligned}
&(\{a, c\}, \{b, d\}, \varnothing, \varnothing) \Vdash a \wedge \neg b\\
&(\{a\}, \{b\}, \varnothing, \{c, d\}) \Vdash a \wedge \neg c \vee \neg a \wedge c
\end{aligned} \tag{6.28}$$

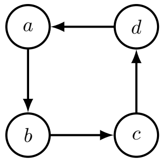



**Figure 6.7:** *Example argumentation framework for the satisfaction of constraints by open labelings.*

Note that these are in fact the same constraints that we have used for constrained argumentation frameworks in Subsection 6.1.2. We have seen there that the second constraint could not be satisfied by any admissible set. However, nothing is said in (Coste-Marquis et al., 2006) about what kind of formulas are not satisfiable for a particular framework. Here we aim to discuss this issue.

Let us first see that also in our case there exist formulas that are not arg-satisfiable by any open labeling. Indeed, consider the formula $\phi = a \wedge b$. In order to satisfy it, we would need to accept both $a$ and $b$, but this is not possible because $a \rightarrow b$. Furthermore, formulas that are not satisfiable in the propositional logic sense cannot be arg-satisfied either.

**Definition 95.** *Let $F = (\mathcal{A}, \mathcal{R})$ be an argumentation framework and let $\phi$ be a satisfiable propositional formula from $PL(\mathcal{A})_{NNF}$. We say that $\phi$ is **arg-consistent** in $F$ iff the set of formulas $\{\phi\} \cup \{\neg a \vee \neg b \mid (a, b) \in \mathcal{R}\}$ is satisfiable.*

The arg-consistency notion introduced in Definition 95 is in fact related only to conflict-freeness and can be seen as a minimal requirement for arg-satisfaction, in the sense that formulas that violate it cannot be arg-satisfied. In what follows we will show that the converse is also true, i.e. any arg-consistent formula is arg-satisfiable.

We first rely on the disjunctive normal form for providing an alternative characterization of arg-consistent formulas.

**Definition 96.** *We say that a propositional formula $\phi$ from $PL(\mathcal{P}rop)$ (or, alternatively, from $PL(\mathcal{P}rop)_{NNF}$) is in **disjunctive normal form** (DNF) iff:*

$$\phi = \bigvee \{pr(\Phi) \mid \Phi \in \Gamma\} \tag{6.29}$$

*where $\Gamma \subseteq \mathcal{P}(\mathcal{P}rop)$ is a set of sets of propositional symbols from $\mathcal{P}rop$ and*

$$pr(\Phi) = \bigwedge \{p \mid p \in \Phi\} \wedge \bigwedge \{\neg p \mid p \in \mathcal{P}rop \setminus \Phi\} \tag{6.30}$$

*is the same notation we have used for modal languages. We will use $PL(\mathcal{P}rop)_{DNF}$ to refer to the language consisting of all disjunctive normal forms.*

It is well known that propositional formulas can be converted to an equivalent disjunctive normal form, so we will not enter into the details of this. Furthermore, aside from $\bot$, all other disjunctive normal forms are satisfiable. We are now ready for an equivalent characterization of arg-consistent formulas.

**Lemma 9.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework and* $\phi \in PL(\mathcal{A})_{DNF}$ *a propositional formula in disjunctive normal form, i.e.:*

$$\phi = \bigvee\{pr(\Phi) \mid \Phi \in \Gamma\} \tag{6.31}$$

*where* $\Gamma \subseteq \mathcal{P}(\mathcal{A})$. *Then* $\phi$ *is arg-consistent in* $F$ *iff* $\Gamma \cap \mathcal{E}_{\mathcal{CO}}(F) \neq \varnothing$.

*Proof.* We start with the direct implication. We assume that $\phi$ is arg-consistent, so the set $S = \{\phi\} \cup \{\neg a \vee \neg b \mid (a, b) \in \mathcal{R}\}$ is satisfiable. Then there exists a truth assignment that satisfies all formulas from $S$. This truth assignment must then satisfy exactly one of the minterms of $\phi$, call it $pr(\Phi_0)$. Let us see that $\Phi_0 \in \mathcal{E}_{\mathcal{CF}}(F)$. Suppose that there exist two arguments $a, b \in \Phi_0$ such that $a \to b$. But then we would have $(a, b) \in \mathcal{R}$, so $\neg a \vee \neg b \in S$ is not satisfied, which is a contradiction. Thus, $\Phi_0$ is a conflict-free set in $F$, which is the desired result.

For the converse, we know that $\Gamma \cap \mathcal{E}_{\mathcal{CF}}(F) \neq \varnothing$ so we can choose an arbitrary $\Phi_0 \in \Gamma \cap \mathcal{E}_{\mathcal{CO}}(F)$. We assign $\top$ to all elements of $\Phi_0$ and $\bot$ to the arguments that are not in $\Phi_0$. We show that this truth assignment satisfies $S = \{\phi\} \cup \{\neg a \vee \neg b \mid (a, b) \in \mathcal{R}\}$. Indeed, the satisfaction of $\phi$ follows directly from the choice of the truth assignment to match $pr(\Phi_0)$. Furthermore, suppose that there exist arguments $a$ and $b$ such that $(a, b) \in \mathcal{R}$ and $\neg a \vee \neg b$ is not satisfied. Then both $a$ and $b$ must be in $\Phi_0$, which is a contradiction, as $\Phi_0$ was chosen conflict-free. This concludes our proof. □

**Theorem 9.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework. Any propositional formula* $\phi \in PL(\mathcal{A})_{NNF}$ *that is arg-consistent in* $F$ *is satisfied by at least one strict labeling of* $F$.

*Proof.* We start by converting $\phi$ to its equivalent disjunctive normal form:

$$\phi \leftrightarrow \bigvee\{pr(\Phi) \mid \Phi \in \Gamma\} \tag{6.32}$$

where $\Gamma \subseteq \mathcal{P}(\mathcal{A})$. We know from Lemma 9 that we can choose $\Phi_0 \in \Gamma \cap \mathcal{E}_{\mathcal{CF}}(F)$. Based on the definition of arg-satisfaction it is enough to prove that $pr(\Phi_0)$ is arg-satisfied by a strict labeling of $F$.

We show that the labeling $\mathcal{L} = (\Phi_0, \varnothing, \varnothing, \mathcal{A} \setminus \Phi_0)$ is strict and arg-satisfies $pr(\Phi_0)$. The restricted argumentation framework $F{\downarrow_{\Phi_0}}$ only conssits of isolated arguments (because $\Phi_0$ is conflict-free) so its only complete labeling is $(\Phi_0, \varnothing, \varnothing)$, which proves that $\mathcal{L}$ is a unique labeling. Since $\mathtt{undec}(\mathcal{L}) = \varnothing$, we have that $\mathcal{L}$ is also decided and, hence, strict. The fact that $\mathcal{L} \Vdash pr(\Phi_0)$ follows directly from the definition of arg-satisfaction. □

Since a strict labeling is also decided, unique and open, we can formulate the following corollary:

**Corollary 3.** *Let* $F = (\mathcal{A}, \mathcal{R})$ *be an argumentation framework. Any propositional formula from* $PL(\mathcal{A})_{NNF}$ *that is arg-consistent in* $F$ *is satisfied by at least one* $\delta$ *labeling, where* $\delta \in \{open, decided, unique, strict\}$.

The result is quite strong, as it shows that any reasonable constraint (one that does not require conflicting arguments to be labeled `in` at the same time) can be arg-satisfied. This means that imposing such constraints on open, decided, unique or strict labelings leads to semantics that are still able to provide at least one labeling for any argumentation framework (they are universally defined).

**Definition 97.** *Let $\mathcal{L}$ be a $\delta$ labeling of an argumentation framework $F = (\mathcal{A}, \mathcal{R})$, where $\delta \in \{open, decided, unique, strict\}$, and let $\phi \in PL(\mathcal{A})_{NNF}$ be a formula that is arg-consistent in $F$. We say that $\mathcal{L}$ is a **constrained** $\delta$ labeling (or, alternatively, a $\delta$ **labeling constrained by** $\phi$) iff $\mathcal{L}$ arg-satisfies $\phi$. The set of all constrained $\delta$ labelings of $F$ is denoted by $\mathcal{L}_{\mathcal{KO}}(F,\phi)$, $\mathcal{L}_{\mathcal{KD}}(F,\phi)$, $\mathcal{L}_{\mathcal{KU}}(F,\phi)$ and $\mathcal{L}_{\mathcal{KS}}(F,\phi)$ respectively.*

Note that $\mathcal{K}$ was chosen for "constrained" instead of $\mathcal{C}$ in order to avoid confusion, as we have already used $\mathcal{L}_{\mathcal{CO}}(F)$ to stand for the complete labelings of $F$.

Since all the semantics introduced in Definition 97 provide at least one labeling for any argumentation framework, it makes sense to minimize the set of ignored arguments, just as we did in the previous subsection. This generates another four labeling-based semantics.

**Definition 98.** *Let $\mathcal{L}$ be a $\delta$ labeling constrained by $\phi$ of an argumentation framework $F$, where $\delta \in \{open, decided, unique, strict\}$. Then $\mathcal{L}$ is said to be a **minimal constrained** $\delta$ labeling iff* $\texttt{ign}(\mathcal{L})$ *is minimal (with respect to set inclusion) among* $\texttt{ign}(\mathcal{L}')$ *of all constrained $\delta$ labelings $\mathcal{L}'$ of $F$. The set of all minimal constrained $\delta$ labelings of $F$ is denoted by $\mathcal{L}_{\mathcal{MKO}}(F,\phi)$, $\mathcal{L}_{\mathcal{MKD}}(F,\phi)$, $\mathcal{L}_{\mathcal{MKU}}(F,\phi)$ and $\mathcal{L}_{\mathcal{MKS}}(F,\phi)$ respectively.*

We end this subsection with an example. We consider the framework from Figure 6.6 and the formula $\phi = b \wedge \neg c \vee \neg a$. The constrained versions of the open, decided, unique and strict labelings are shown in Table 6.3.

| $\mathcal{L}_{\mathcal{O}}(F)$ | $\mathcal{L}_{\mathcal{KO}}(F,\phi)$ | $\mathcal{L}_{\mathcal{KD}}(F,\phi)$ | $\mathcal{L}_{\mathcal{KU}}(F,\phi)$ | $\mathcal{L}_{\mathcal{KS}}(F,\phi)$ |
|---|---|---|---|---|
| $(\{a\},\{b\},\{c\},\varnothing)$ | | | | |
| $(\{b\},\{c\},\varnothing,\{a\})$ | ✓ | ✓ | | |
| $(\varnothing,\varnothing,\{b,c\},\{a\})$ | ✓ | | | |
| $(\{a\},\varnothing,\{c\},\{b\})$ | | | | |
| $(\{a\},\{b\},\varnothing,\{c\})$ | | | | |
| $(\varnothing,\varnothing,\{c\},\{a,b\})$ | ✓ | | ✓ | |
| $(\{a\},\varnothing,\varnothing,\{b,c\})$ | | | | |
| $(\{b\},\varnothing,\varnothing,\{a,c\})$ | ✓ | ✓ | ✓ | ✓ |
| $(\varnothing,\varnothing,\varnothing,\{a,b,c\})$ | ✓ | ✓ | ✓ | ✓ |

**Table 6.3:** *Labeling-based semantics constrained by $\phi = b \wedge \neg c \vee \neg a$ for the argumentation framework from Figure 6.6*

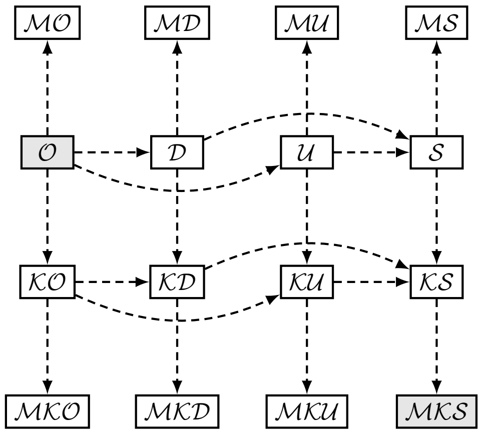



**Figure 6.8:** *Novel labeling-based argumentation semantics and relations between them.*

These results also lead to the following labelings for the minimal constrained versions of the semantics:

$$\begin{aligned}
\mathcal{L}_{\mathcal{MKO}}(F,\phi) &= \{(\{b\},\{c\},\varnothing,\{a\}),(\varnothing,\varnothing,\{b,c\},\{a\})\}\\
\mathcal{L}_{\mathcal{MKD}}(F,\phi) &= \{(\{b\},\{c\},\varnothing,\{a\})\}\\
\mathcal{L}_{\mathcal{MKU}}(F,\phi) &= \{(\varnothing,\varnothing,\{c\},\{a,b\}),(\{b\},\varnothing,\varnothing,\{a,c\})\}\\
\mathcal{L}_{\mathcal{MKS}}(F,\phi) &= \{(\{b\},\varnothing,\varnothing,\{a,c\})\}
\end{aligned} \tag{6.33}$$

This example also shows that the proposed semantics are distinct. To summarize, we have introduced 16 novel labeling-based semantics. We provide a map of the relations between these semantics in Figure 6.8. We have grayed out the most general (open) semantics and the most restrictive one (minimal constrained strict). Note that a comparison with the semantics from Chapter 2 is not possible, as we have not introduced extension-based versions for our semantics, because the set of `in`-labeled arguments only makes sense if we also know which of the arguments are ignored.

### 6.2.3 Discussion

In this subsection we aim to compare our approach based on open labelings with the works that inspired it and to point out the distinguishing features of our work.

We start with the enhanced preferred semantics (Zhang and Lin, 2010), as the idea of ignoring arguments comes from their work. The most important distinction between our labeling-based semantics and the enhanced preferred one is the fact that we consider that the set of `in`-labeled arguments is only meaningful if the set of ignored arguments is also known, so we only work with the whole labeling. This is the reason why we provided no extension-based version for our semantics.

Furthermore, the minimal versions of our semantics use set inclusion instead of cardinality, as we feel that cardinality alone may not be that relevant. In a multi-agent system scenario, for example. the source of the ignored arguments may be more relevant than their actual number: ten arguments coming from unreliable sources might be more easily defeated than a single argument from a trusted peer.

Next, let us compare our approach with the now standard complete labelings (Caminada, 2006a). First of all, it is easy to see and we have mentioned already that if no argument is ignored, our approach produces precisely the complete labelings. We may say that our approach subsumes the complete labelings, although in fact the intended use of our labelings is a bit distinct: we expect that the ignored arguments are to be defeated by adding new arguments to the framework.

Furthermore, note that while requiring complete labelings to have no undecided argument (stable labelings) can lead to no possible labeling, in our case we are sure to have at least one labeling. In fact, we can even enforce other constraints as well and still obtain universally defined labelings. Of course, this comes with the drawback that our labelings cannot be used in the same way as complete labelings but, as we shall see in the following section, they are rather useful in dynamic scenarios where one is allowed to add new arguments to the framework.

We will now discuss constrained argumentation frameworks, which are also strongly related to our approach. In contrast with (Coste-Marquis et al., 2006), we use labelings instead of extensions in our approach. Furthermore, the language that we use for constraints is restricted to negation normal form. While for constrained argumentation frameworks it is not very clear which formulas can be satisfied and which ones cannot, we provide a precise characterization of formulas that cannot be arg-satisfied by open labelings. In fact, as we have seen, any formula that does not require conflicting arguments to be accepted at the same time can be arg-satisfied by a strict labeling. This result is quite strong and contrasts with the case of constrained argumentation frameworks, where it is often the case that no admissible set satisfies the desired constraint.

In addition, our approach also uses more general constraints, such as an empty set of ignored arguments or uniqueness with respect to the set of ignored arguments. While the former can also be expressed using a propositional formula, we preferred to treat it separately due to the fact that it resembles the constraints that defines stable labelings.

Last, but not least, our proposal is related to conditional labelings (Boella et al., 2011). We have seen in Subsection 6.1.3 that conditional labelings tell us what arguments need to be defeated in order for a certain argument to be labeled with a particular label by the grounded labeling. Our work generalizes this idea as follows:

- our goal is not the label of a single argument, but a more complex constraint that may involve several arguments
- our labelings provide sets of arguments that need to be defeated in order for a complete labeling to satisfy the goal (we are, thus, not limited to the grounded labeling)

On the other hand, our approach assumes that undecided arguments are not desirable and provides no means for writing constraints that require an argument to be undecided. Our formulas can be seen as a combination of those used for constrained argumentation semantics and those used for conditional labelings, as for us the propositional symbol $a$ has the same meaning as $a^+$, while $\neg a$ is the same as $a^-$. We shall see an actual example of how our labelings can be used in the next section.

## 6.3 An Argumentation-based MAS Scenario

In this section we discuss the argumentation-based multi-agent scenario that we have proposed in (Gratie and Florea, 2012a) as an application of open labelings. While the setting and the storyline of the example are the same, there are several changes with respect to the original proposal. The scenario points out how abstract argumentation frameworks can be used for the modeling of a multi-agent scenario and also how open labelings appear as a natural generalization of complete labelings, helping agents decide their actions based on the desired outcome.

### 6.3.1 Auction scenario

We will consider a very simple negotiation scenario for the sale of two items. To keep things simple. we will use an unspecified currency (just a positive integer). We will consider only three persons in our scenario: Anthony, Brian and Carol.

We assume that Anthony wants to sell two old pieces of furniture: a chair and a table. Since Anthony has been collecting old items for quite some time, he knows that the value of the chair is 200, while the value of the table is 300. Anthony wishes to obtain at least these amounts for the items. Otherwise, he will not sell them. Two of Anthony's friends, Brian and Carol, are interested in the two items and meet him to negotiate the sale.

In order to have a well organized negotiation, Anthony proposes a set of rules leading to a protocol that can be seen as a special kind of auction. In a turn-based fashion, each participant is allowed to place bids (or, in Anthony's case, to ask for a minimum price) or to argue about the price or about who should receive each item. The negotiation ends when none of the participants has anything more to say. Then the outcome is decided based on the bids and all the other arguments that have been provided, using rules that are known to all participants from the beginning of the negotiation. We will provide more details about this mechanism in the following subsections.

We assume that Brian is rather rich, but knows very little about antiques, so he is willing to pay even twice the value of each item, i.e. 400 for the chair and 600 for the table. On the other hand, he wants to either get both items or none of them. Carol has more limited resources, so she cannot afford both items, but would really like to buy one of them. She is willing to pay 300 for the chair and 500 for the table.

We are going to model this informal scenario as a turn-based game featuring agents that work with abstract argumentation frameworks and constraints. In the following subsections we will discuss relevant parts of the scenario.

### 6.3.2 Outcome arguments

We start by discussing what we referred to as the environment in (Gratie and Florea, 2012a). Here we emphasize the fact that the relevant states of the environment are in fact the possible outcomes of the negotiation. In our case, these refer to the owners of each of the two items. We will use a first-order predicate language for writing the corresponding arguments, which will be referred to as outcome arguments.

We have two items, *chair* and *table*, and three possible owners: *Anthony*, *Brian* and *Carol*. We can assume that we use a first-order formula for each possible case:

$$\begin{aligned} s_1 &= has(Anthony, chair) \wedge has(Anthony, table) \\ s_2 &= has(Anthony, chair) \wedge has(Brian, table) \\ s_3 &= has(Anthony, chair) \wedge has(Carol, table) \end{aligned} \tag{6.34}$$

and so on, for a total of $3^2 = 9$ states. We can read these states as arguments, deduce that no two can occur at the same time and add attacks between all pairs of arguments. The attacks ensure that no two arguments will be `in` at the same time. We can also add the constraint $\phi = s_1 \vee s_2 \vee \ldots \vee s_9$ in order to keep only the complete labelings that do label one of the 9 states as `in`. While this representation is semantically correct, the number of arguments is exponential in the number of items.

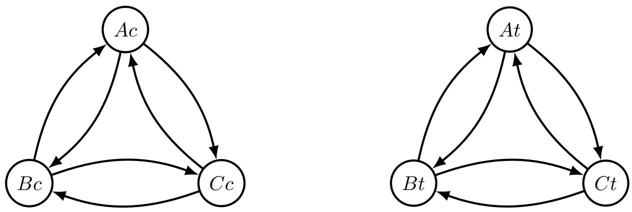



**Figure 6.9:** *State arguments for the auction example.*

As an alternative, let us see that the owner of the chair does not depend on the owner of the table and the actual restriction is that each item has exactly one owner. Thus, we can use the following outcome arguments:

$$\begin{aligned} Ac &= has(Anthony, chair) \\ Bc &= has(Brian, chair) \\ Cc &= has(Carol, chair) \\ At &= has(Anthony, table) \\ Bt &= has(Brian, table) \\ Ct &= has(Carol, table) \end{aligned} \tag{6.35}$$

A suitable and intuitive argumentation framework for this case is the one depicted in Figure 6.9, in conjunction with the following constraint: $\varphi_o = (Ac \vee Bc \vee Cc) \wedge (At \vee Bt \vee Ct)$. The constraint ensures that each item will have at least one owner, while the attacks between arguments ensure that there will not be more than one owner for each item. Note that this representation is linear in the number of items. We will further refer to $\varphi_o$ as the outcome consistency constraint.

In contrast with (Gratie and Florea, 2012a) we will assume here that the representation of the possible negotiation outcomes is fixed and cannot be altered during the game. It is possible, on the other hand, to add new arguments that attack or are attacked by outcome arguments.

### 6.3.3 Action and reaction

We will use first-order predicates to describe actions as well. We will refer to the corresponding arguments as practical or action arguments.

In our scenario we have two types of actions. Brian and Carol can bid for each of the items, while Anthony can require a minimum price for each item. For example, $bids(table, 350)$ encodes a bid of 350 for the table. We will annotate actions with the name of the agent performing them, as in $Carol : bids(table, 350)$. The implicit attacks between such arguments come from the fact that a higher bid for the same item is preferred to a lower one. We will assume that Anthony states his minimum prices by placing bids himself, with the desired amounts: if his bids are the highest, he will keep the items.

Furthermore, actions may have an impact on the outcome of the negotiation. For example, the bid $Carol : bids(table, 350)$, if highest, should imply that Carol becomes the owner of the table. For this, we should have the following attacks: $Carol : bids(table, 350) \rightarrow At$ and $Carol : bids(table, 350) \rightarrow Bt$. We assume that such implicit attacks are also common knowledge for all auction participants.

We assume that the first round of negotiation consisted in the following arguments:

$$\begin{aligned} Anthony & \begin{cases} bids(chair, 200) & \triangleq bAc \\ bids(table, 300) & \triangleq bAt \end{cases} \\ Brian & \begin{cases} bids(chair, 400) & \triangleq bBc \\ bids(table, 450) & \triangleq bBt \end{cases} \\ Carol & \begin{cases} bids(chair, 300) & \triangleq bCc \\ bids(table, 500) & \triangleq bCt \end{cases} \end{aligned} \tag{6.36}$$

With these, the argumentation framework that describes the current state of the negotiation is the one from Figure 6.10.

The grayed arguments from the figure form the only complete extension of the framework which, in addition, satisfies the outcome consistency constraint from the previous subsection. We say that this extension is a valid outcome of the framework. Should this be the final state of the negotiation process, it would mean that Brian is bound to pay 400 for the chair and Carol should pay 500 for the table. The resulting state after these actions would be the one also described by the extension, where the winning bidders actually get the items they pay for. But more about commitment and the outcome of the auction later on.

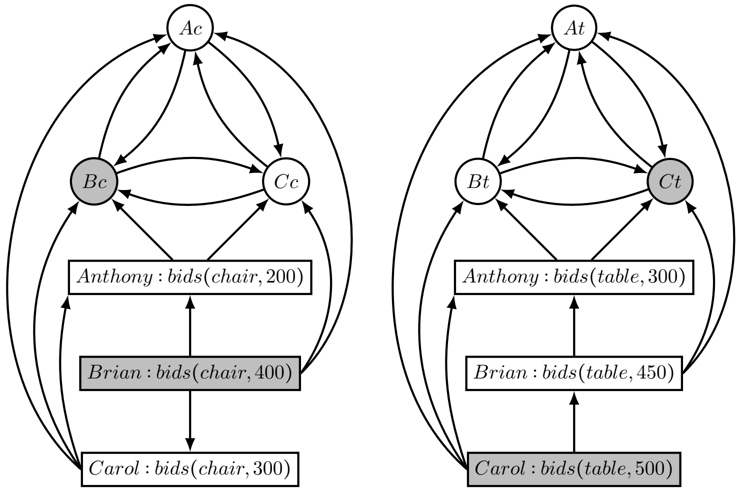



**Figure 6.10:** *Negotiation status after the first round. Grayed arguments form the only complete extension of the framework.*

### 6.3.4 Beliefs, desires, intentions

In this subsection we explore the role that beliefs, desires and intentions play in our scenario. While there is no one-to-one correspondence, we have several mechanisms that can help. First of all, let us see that the action arguments can be seen as intentions, or even plans, of the agents. They turn into actual actions only once an extension is chosen as the outcome of the negotiation.

Another mechanism consists in restricting the action arguments that are available to each agent. This can help us express the desires of the agent (the agent only considers the actions it would do) but it may also describe the abilities of the agent (some agents may have a smaller set of abilities with respect to others). In our case, we have that Brian's actions range from $bids(chair, 1)$ to $bids(chair, 400)$ and from $bids(table, 1)$ to $bids(table, 600)$, while Carol's actions range from $bids(chair, 1)$ to $bids(chair, 300)$ and from $bids(table, 1)$ to $bids(table, 500)$.

We also know that Carol can only afford one item. In order to say this, we will add attacks between Carol's bid for the chair and her bid for the table. Furthermore, Brian wants both items. In order to enforce this, we can add attacks from outcome arguments where Brian does not own the chair to Brian's bid for the table and from outome arguments where Brian does not own the table to Brian's bid for the chair.

The goals of the agents can also be expressed by means of constraints. Brian's desire to buy either both items or none can be expressed by $\varphi_{Brian} = Bc \wedge Bt \vee \neg Bc \wedge \neg Bt$. Carol's wish to buy one of the items can be expressed by $\varphi_{Carol} = Cc \vee Ct$. Anthony's wish to sell both items can be written as $\varphi_{Anthony} = \neg Ac \wedge \neg At$.

We can see the resulting argumentation framework in Figure 6.11, where we also show in gray background the elements of one of its complete extensions. Note

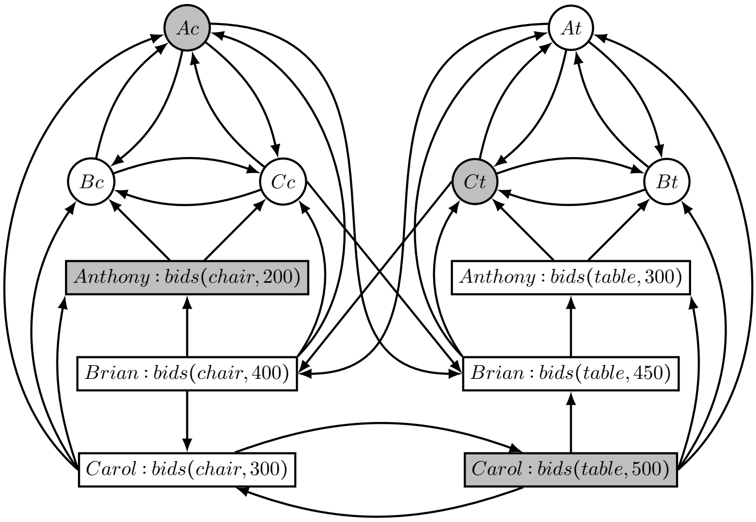



**Figure 6.11:** *Complete bid scenario extended from Figure 6.10. Grayed arguments form one of the three complete extensions of the framework.*

that the framework has the following three complete labelings:

$$\begin{aligned}&(\{Cc, bCc, At, bAt\}, \{Ac, bAc, Bc, bBc, Bt, bBt, Ct, bCt\}, \varnothing)\\&(\{Ac, bAc, Ct, bCt\}, \{Bc, bBc, Cc, bCc, At, bAt, Bt, bBt\}, \varnothing)\\&(\varnothing, \varnothing, \{Ac, bAc, Bc, bBc, Cc, bCc, At, bAt, Bt, bBt, Ct, bCt\})\end{aligned} \quad (6.37)$$

Note that only the first two of them satisfy the outcome consistency constraint $\varphi_o$. Both of them also satisfy $\varphi_{Carol}$ and $\varphi_{Brian}$, but not $\varphi_{Anthony}$. In the next subsection we will see what an agent can do in order to bring about the satisfaction of its own goals.

### 6.3.5 The result of the negotiation

We have seen that the state of the negotiation game can be described using an argumentation framework and constraints. We will now discuss the result of the negotiation and also how the rules that lead to this result can be used throughout the game in order to decide the actions of the agents.

The game is turn-based. During its turn, each agent can change the current argumentation framework so that the instantaneous result of the negotiation (a complete labeling, for example) matches its goals, i.e. satisfies the corresponding constraint. However, there are limits to what an agent can change during a valid move.

First of all, an agent can add any number of action arguments and also attacks between them in order to encode preferences. However, an agent cannot add attacks toward the actions of other agents. On the other hand, we have seen in our scenario that a higher bid for an item implicitly attacks a lower bid of another agent for the same item. Such attacks are automatically added when both involved arguments are in the framework describing the current state of the negotiation. They cannot be removed and are known and accepted by all participants.

The attacks toward outcome arguments are also implicit and are part of the rules of the particular negotiation, known in advance to all participants. In our case, the implicit attacks come from a bid to the arguments that give the corresponding item to other agents than the bidder.

On the other hand, agents are free to add attacks from other agents' actions or from the outcome arguments to their own actions. We have used this when we encoded the fact that Brian wants to buy both items or none of them. In more complex scenarios, such attacks can encode a promise (if a particular outcome is selected, the agent will execute some action or not) or even a threat (if one agent executes some action, another will respond with an action that is not favorable to the first agent).

The outcome of the negotiation uses complete labelings and a set of rules known in advance to all participants. The rules should state whether a particular kind of complete labeling is used, such as preferred, grounded or stable, and also what will be the outcome in case several labelings exist. Only labelings that arg-satisfy the outcome consistency constraint are taken into account. If no such labeling exists, it may make sense to use constrained open labelings instead, as they guarantee the existence of labelings for any reasonable constraint.

The agents can use the knowledge of these rules in order to decide their actions during their turn. Indeed, an agent can compute a kind of constrained open labeling that fits the outcome rules (decided labelings if stable labelings are used for the outcome, unique labelings instead of the grounded labeling, etc.) and satisfies its own goals. The agent can then choose one of these labelings and defeat the corresponding set of ignored arguments in order to change the outcome of the negotiation to fit its goals.

We consider again the configuration from Figure 6.11 and discuss it for each agent. We have already seen that only the first two complete labelings from (6.37) satisfy the outcome consistency constraint. No matter what rule is used for choosing between the two, Carol's goals are satisfied. If, however, Carol prefers the chair to the table, she can remove the attack that goes from her bid for the table to her bid for the chair.

Brian's goal is satisfied, as he gets none of the items irrespective of the chosen labeling. If, however, Brian decides that he prefers to try to get both items, he may compute the constrained labelings that ensure this, then attack the corresponding ignored arguments. For example. Brian can see that if Carol's bid for the table is defeated, then he may get both items. To this end, he may provide a higher bid for the table.

Anthony's goals are not satisfied by either of the two possible outcomes, as he cannot sell both items. Anthony can either pass, hoping that either Brian or Carol will bid higher, or he may lower the minimum price for one of the items.

In addition to action arguments, agents may also use epistemic arguments to reason for example about the quality of the items and negotiate a price reduction. We will include such epistemic arguments in the formal model, presented in Section 6.4.

### 6.3.6 End of the game and commitment

We have seen that the game proceeds in turn-based fashion, each agent changing the current configuration to better fulfill its goals. Whenever an agent is satisfied with the current configuration, or has no more practical or epistemic arguments to put forward for producing a favorable change, the agent will pass. The game ends when none of the agents can or wishes to change the argumentation framework that describes the current configuration.

Once the game has reached its final configuration, all agents are committed to fulfill their intentions by executing the corresponding actions. Suppose that the configuration from Figure 6.11 is reached at the end of the negotiation and the complete (in fact even stable) labeling depicted in the figure is chosen as the outcome. Then Carol will take the table, but must also pay for it, while Anthony will keep the chair.

### 6.3.7 Discussion

We end the section with a discussion of the relevance of the semantics introduced in Section 6.2 for the multi-agent scenario presented here. We have seen that constrained open labelings can be used for deciding what arguments need to be attacked in order for the goals of an agent to be satisfied. The actual type of labeling to be used depends on the rules that are used for deciding the outcome of the game. If an agent relies on constrained strict labelings then, upon attacking the corresponding set of ignored arguments, the resulting framework will have a single complete extension, which will also be stable. Thus, any kind of rule that is based on complete labelings will see the same outcome.

A problem that can appear is that it might not be possible to find any complete labeling that satisfies the outcome consistency constraint, in which case open labelings can be used instead of complete labelings for deciding the result of the negotiation. This also means that some of the agents' arguments might end up ignored, so the rules in this case must be very specific. A possible rule is to simply pick a random open labeling, which may act as an incentive for agents to maintain the outcome consistency constraint satisfied throughout the game, in order to avoid uncertainty of the result of the negotiation.

## 6.4 Argumentation-based Negotiation Game

In this section we turn the informal scenario presented in Section 6.3 into a formal model for a turn-based negotiation game featuring agents and abstract argumentation. We first discuss the most important elements of the model separately, then provide the complete formal definitions. We end the section with a discussion of the approach.

### 6.4.1 Outcome arguments

Since we are modeling a negotiation game, an important element we have to represent is the object of negotiation. We will use a set $\mathcal{A}_o$ of outcome arguments to encode possible results of the negotiation. The attack relation between outcome arguments is fixed and comes from intrinsic incompatibilities or dependencies between the possible outcomes.

While the attacks between the arguments can ensure that no two incompatible outcomes can be labeled `in` at the same time, it is not always possible to ensure that there is at least one `in`-labeled argument (or a set of compatible arguments labeled `in` at the same time). For this, just as we did for the auction scenario, we will use a propositional formula $\varphi_o$ that will be referred to as the outcome consistency constraint.

The fixed attacks between the outcome arguments are part of a larger implicit attack relation $\mathcal{R}$ that also includes attacks between other types of arguments, to be introduced in the following subsections.

For the auction scenario we have the following elements:

$$\begin{aligned}\mathcal{A}_o =& \{has(\alpha, chair), has(\alpha, table) \mid \alpha \in \{Anthony, Brian, Carol\}\}\\ =& \{Ac, At, Bc, Bt, Cc, Ct\}\\ \mathcal{R} \cap (\mathcal{A}_o \times \mathcal{A}_o) =& \{(Ac, Bc), (Bc, Ac), (Bc, Cc), (Cc, Bc), (Cc, Ac), (Ac, Cc),\\ & (At, Bt), (Bt, At), (Bt, Ct), (Ct, Bt), (Ct, At), (At, Ct)\}\\ \phi_o =& (Ac \vee Bc \vee Cc) \wedge (At \vee Bt \vee Ct)\end{aligned} \tag{6.38}$$

where the arguments and the attacks are the same as those depicted in Figure 6.9.

### 6.4.2 Practical arguments

We have used practical, or action, arguments in the previous section, in order to describe the intentions of agents. We have used first-order logic predicates, annotated with the name of the agent executing the action, as in $Carol : bids(table, 350)$. For the formal model we are not interested in the actual structure of the practical arguments, however, and we will simply consider a set of abstract practical arguments $\mathcal{A}_p$.

We also consider that there is an implicit attack relation between practical arguments, encoding dependencies or incompatibilities between actions. Furthermore, there are also attacks between practical arguments and outcome arguments. All these attacks are part of the larger implicit attack relation $\mathcal{R}$.

For the auction scenario from the previous section we can write the elements introduced here as follows:

$$\begin{aligned}Agents =& \{Anthony, Brian, Carol\}\\ \mathcal{A}_p =& \{\alpha : bids(chair, x), \alpha : bids(table, x) \mid \alpha \in Agents, x \in \mathbb{N}\}\\ \mathcal{R} \cap (\mathcal{A}_p \times \mathcal{A}_p) =& \{\alpha : (bids(item, x), \beta : bids(item, y)) \mid item \in \{chair, table\},\\ & \alpha, \beta \in Agents, \alpha \neq \beta, x \in \mathbb{N}, y \in \mathbb{N}, x > y\}\\ \mathcal{R} \cap (\mathcal{A}_p \times \mathcal{A}_o) =& \{(\alpha : bids(item, x), has(\beta, item)) \mid item \in \{chair, table\}\\ & \alpha, \beta \in Agents, \alpha \neq \beta, x \in \mathbb{N}\}\end{aligned} \tag{6.39}$$

The implicit attacks presented in (6.39) come from the fact that higher bids are preferred to lower bids for the same item and a bid for an item, if labeled `in`, should guarantee that the bidding agent will get the item.

### 6.4.3 Epistemic arguments

In addition to outcome and practical arguments, the agents can also argue using epistemic arguments. Such arguments are based on the agents' knowledge about the world or about each other. In the formal setting we assume a fixed set of available epistemic arguments (not all of them known to all agents).

The attacks between epistemic arguments are fixed and are based on the actual logical content of the arguments. In the formal model we only work with abstract arguments, so we assume that the attack relation is given. Furthermore we consider that there can also be implicit attacks between epistemic arguments and outcome or practical arguments. All these are included in the implicit attack relation $\mathcal{R}$.

Let us consider the following epistemic arguments for the auction scenario:

- $nC$ – "Carol should not be allowed to buy any of the items because several of the antiques she owns barely escaped from a fire last year and many of them were damaged."
- $nC'$ – "The fire was not Carol's fault, so she can still be trusted with the items."
- $nA$ – "The table should be sold for a starting price lower than 200, because the table is not authentic."

Then we can add the following elements to the formal model of the auction scenario:

$$
\begin{aligned}
\mathcal{A}_e &= \{nC, nC', nA\} \\
\mathcal{R} \cap (\mathcal{A}_e \times \mathcal{A}_e) &= \{(nC, nC'), (nC', nC)\} \\
\mathcal{R} \cap (\mathcal{A}_e \times \mathcal{A}_o) &= \{(nC, has(Carol, chair)), (nC, has(Carol, table))\} \\
\mathcal{R} \cap (\mathcal{A}_e \times \mathcal{A}_p) &= \{(nA, Anthony : bids(table, x)) \mid x \geq 200\}
\end{aligned}
\tag{6.40}
$$

### 6.4.4 The agents

We will use $\mathcal{A}g$ to refer to the set of agents. Each agent, or player, is characterized by the actions it can perform, its knowledge and its goals. Thus, we can represent an agent $\alpha_i$ as a tuple $(\mathcal{A}_i, \mathcal{K}_i, \varphi_i)$ consisting of the following elements:

- $\mathcal{A}_i \subseteq \mathcal{A}_p$ - the set of actions the agent can (or is willing to) do
- $\mathcal{K}_i \subseteq \mathcal{A}_e$ - the set of epistemic arguments that the agent knows
- $\varphi_i \in PL(\mathcal{A}_o)_{NNF}$ - the goals of the agent as a propositional formula that can be used as a constraint

For the auction scenario, we have the three agents *Anthony*, *Brian* and *Carol*. We will subscript the components of their corresponding tuple with their names instead of numbers. Based on the scenario presented in the previous section, agent

*Anthony* is given by:

$$
\begin{aligned}
\alpha_{Anthony} &= (\mathcal{A}_{Anthony}, \mathcal{K}_{Anthony}, \varphi_{Anthony}) \\
\mathcal{A}_{Anthony} &= \{Anthony : bids(chair, x) \mid x \in \mathbb{N}, x \geq 200\} \\
&\cup \{Anthony : bids(table, x) \mid x \in \mathbb{N}, x \geq 300\} \\
\mathcal{K}_{Anthony} &= \varnothing \\
\varphi_{Anthony} &= \neg Ac \wedge \neg At
\end{aligned} \tag{6.41}
$$

For *Brian* we have the following:

$$
\begin{aligned}
\alpha_{Brian} &= (\mathcal{A}_{Brian}, \mathcal{K}_{Brian}, \varphi_{Brian}) \\
\mathcal{A}_{Brian} &= \{Brian : bids(chair, x) \mid x \in \mathbb{N}, x \leq 400\} \\
&\cup \{Brian : bids(table, x) \mid x \in \mathbb{N}, x \leq 600\} \\
\mathcal{K}_{Brian} &= \{nC\} \\
\varphi_{Brian} &= Bc \wedge Bt \vee \neg Bc \wedge \neg Bt
\end{aligned} \tag{6.42}
$$

Finally, for *Carol*'s agent we can write:

$$
\begin{aligned}
\alpha_{Carol} &= (\mathcal{A}_{Carol}, \mathcal{K}_{Carol}, \varphi_{Carol}) \\
\mathcal{A}_{Carol} &= \{Carol : bids(chair, x) \mid x \in \mathbb{N}, x \leq 300\} \\
&\cup \{Carol : bids(table, x) \mid x \in \mathbb{N}, x \leq 500\} \\
\mathcal{K}_{Carol} &= \{nC', nA\} \\
\varphi_{Carol} &= Cc \wedge \neg Ct \vee \neg Cc \wedge Ct
\end{aligned} \tag{6.43}
$$

Note that here we preferred to write a more explicit goal for Carol, one that does not assume any particular attacks between Carol's bids, such as the ones we have used in the previous section. Those attacks will probably appear in the framework anyway, in order to ensure that the goal is satisfied, but Carol is free to find other ways for satisfying her goal.

### 6.4.5 The rules of the game

We are now ready to specify the rules of the negotiation game. As mentioned already, it is a turn-based game. The current configuration of the game consists of an argumentation framework that encodes the possible outcomes and all the arguments put forward by the participating agents.

The initial configuration is given by $(\mathcal{A}_o, \mathcal{R}_o)$. During its turn, an agent $\alpha_i$ can change the framework using an arbitrary long sequence of actions from the following:

- add one of its own practical arguments $a \in \mathcal{A}_i$; implicit attacks between $a$ and arguments that were already in the framework are automatically added as well
- remove one of its own practical arguments from the framework

- add one of its own epistemic arguments $a \in \mathcal{K}_i$; implicit attacks are automatically added in this case as well; epistemic arguments cannot be removed from the framework once they have been added
- add an attack towards one of its own practical arguments that are already included in the framework
- remove an attack against one of its practical arguments, but only if the attack was previously added to the framework by the same agent (it is not an implicit attack)

In addition, the agent can choose to do nothing (pass). Once all agents choose to pass, the result of the game can be computed. The result is defined by an outcome strategy, which uses a kind of minimal labeling (open, decided, unique or strict) constrained by $\varphi_o$, specified by the rules of the game and denoted by $\mathcal{L}_\mathcal{S}$. If the chosen semantics provides more than one labeling, one is randomly chosen.

An agent is considered a winner of the game if its goals are satisfied by all the labelings provided by $\mathcal{L}_\mathcal{S}$. If only some of the labelings satisfy its goals, the agent possibly wins the game. If none of the labelings satisfies its goals, the agent has lost the game. Note that there can be several winners at the same time and also it is possible that all agents lose the game.

### 6.4.6 Putting it all together

We are now ready to formally define argumentation-based negotiation games.

**Definition 99.** *An **argumentation-based negotiation game** is defined as a tuple*

$$G = (\mathcal{A}_o, \mathcal{A}_p, \mathcal{A}_e, \mathcal{R}, \varphi_o, \mathcal{A}g, \mathcal{L}_\mathcal{S}) \tag{6.44}$$

*where $\mathcal{A}_o$ is a set of **outcome** arguments, $\mathcal{A}_p$ is a set of **practical (action)** arguments, $\mathcal{A}_e$ is a set of **epistemic (knowledge)** arguments, $R$ is the **implicit binary attack relation** on $\mathcal{A}_o \cup \mathcal{A}_a \cup \mathcal{A}_e$, $\varphi_o \in PL(\mathcal{A}_o)_{NNF}$ is a propositional formula encoding the **outcome consistency constraint** and $\mathcal{A}g$ is a set of **agents (players)**. Each agent $\alpha_i \in \mathcal{A}g$ is a tuple $(\mathcal{A}_i, \mathcal{K}_i, \varphi_i)$, where $\mathcal{A}_i \subseteq \mathcal{A}_p$ contains the **abilities** of the agent (the actions it can take), $\mathcal{K}_i \subseteq \mathcal{A}_e$ denotes the **knowledge** of the agent and $\varphi_i \in PL(\mathcal{A}_o)_{NNF}$ is a propositional formula that encodes the **goals** of the agent. The last element of the game tuple, $\mathcal{L}_\mathcal{S}$, is the **winning rule** and stands for a constrained argumentation labeling, $\mathcal{S} \in \{\mathcal{KO}, \mathcal{KD}, \mathcal{KU}, \mathcal{KS}, \mathcal{MKO}, \mathcal{MKD}, \mathcal{MKU}, \mathcal{MKS}\}$.*

We will consider a simple example, given by the following values:

$$
\begin{aligned}
\mathcal{A}_o &= \{yy, yn, ny, nn\} \\
\mathcal{A}_p &= \{y_1, n_1, y_2, n_2\} \\
\mathcal{A}_e &= \{d\} \\
\mathcal{R} &= \{(a,b) \mid a, b \in \mathcal{A}_o, a \neq b\} \\
&\quad \cup \{(y_1, ny), (y_1, nn), (n_1, yy), (n_1, yn), (y_2, yn), (y_2, nn), (n_2, yy), (n_2, ny)\} \\
&\quad \cup \{(d, n_2)\} \\
\varphi_o &= yy \vee yn \vee ny \vee nn \\
\mathcal{L}_{\mathcal{S}} &= \mathcal{L}_{\mathcal{MKD}} \\
\mathcal{A}g &= \{\alpha_1, \alpha_2\} \\
\alpha_1 &= (\{y_1, n_1\}, \{d\}, ny) \\
\alpha_2 &= (\{y_2, n_2\}, \varnothing, yn)
\end{aligned}
\tag{6.45}
$$

In words, we have two agents that can choose “yes” or “no” as an action, leading to four possible outcomes. We know that each agent prefers the outcome where its own choice was “no” and the other agent’s choice was “yes”. Furthermore, $\alpha_1$ knows some information $d$ that invalidates the result of $\alpha_2$’s “no” action. The outcome of the game is decided using minimal constrained decided labelings.

**Definition 100.** *A* ***game configuration*** *consists of an argumentation framework $(\mathcal{A}_c, \mathcal{R}_c)$, with $\mathcal{A}_c \subseteq \mathcal{A}_o \cup \mathcal{A}_p \cup \mathcal{A}_e$ and $\mathcal{R}_c \subseteq \mathcal{A}_c \times \mathcal{A}_c$, such that the following relations hold:*

$$
\begin{aligned}
&\mathcal{A}_o \subseteq \mathcal{A}_c \\
&\mathcal{R} \cap \mathcal{R}_c = \mathcal{R} \cap (\mathcal{A}_c \times \mathcal{A}_c)
\end{aligned}
\tag{6.46}
$$

*The* ***initial configuration*** *is $(\mathcal{A}_o, \mathcal{R} \cap (\mathcal{A}_o \times \mathcal{A}_o))$.*

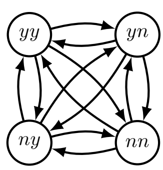



**Figure 6.12:** *Initial configuration for the game from (6.45).*

Thus, for our simple game, the initial configuration is the one from Figure 6.12. Next, we define the legal moves for our game.

**Definition 101.** *Given a game configuration $(\mathcal{A}_c, \mathcal{R}_c)$, a* ***valid move*** *for an agent $\alpha_i$ consists in generating a new configuration $(\mathcal{A}'_c, \mathcal{R}'_c)$ such that the following* *relations hold:*

$$\begin{aligned} &\mathcal{A}'_c \setminus \mathcal{A}_c \subseteq \mathcal{A}_i \cup \mathcal{K}_i \\ &\mathcal{A}_c \setminus \mathcal{A}'_c \subseteq \mathcal{A}_i \\ &\mathcal{R} \cap (\mathcal{A}'_c \times \mathcal{A}'_c) \subseteq \mathcal{R}'_c \\ &\mathcal{R}'_c \setminus (\mathcal{R}_c \cup \mathcal{R}) \subseteq \mathcal{A}'_c \times \mathcal{A}_i \\ &\mathcal{R}_c \setminus (\mathcal{R}'_c \cup \mathcal{R}) \subseteq \mathcal{A}'_c \times \mathcal{A}_i \end{aligned} \tag{6.47}$$

In words, the first condition from (6.47) requires that any added arguments are from the agent's own abilities or knowledge. The second condition requires that removed arguments are from the agent's own abilities (epistemic arguments cannot be removed). The third condition requires that the implicit attacks are not altered in any way. The last two conditions impose that any added or removed attacks target the epistemic arguments of the agent performing the move.

Suppose that $\alpha_1$ is the first agent to move. One valid move according to Definition 101 is to add the practical argument $n_1$ to the current configuration. We then assume that $\alpha_2$ adds $n_2$, so the current configuration becomes the one from Figure 6.13.

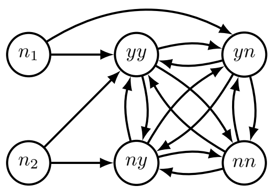



**Figure 6.13:** *Game configuration after the first move of both agents.*

The last concept we need to formally define is the result of the game, once all participants pass.

**Definition 102.** *Given the end game configuration $F_c = (\mathcal{A}_c, \mathcal{R}_c)$ and an agent $\alpha_i$, we distinguish the following cases:*

- $\alpha_i$ ***(surely) wins*** *the game iff $\mathcal{L}_\mathcal{S}(F_c, \varphi_o)$ contains a single labeling $\mathcal{L}$ and $\mathcal{L} \Vdash \varphi_i$*
- $\alpha_i$ ***possibly wins*** *the game iff $\mathcal{L}_\mathcal{S}(F_c, \varphi_o)$ contains more than one labeling and there exists at least one labeling $\mathcal{L}$ such that $\mathcal{L} \Vdash \varphi_i$*
- $\alpha_i$ ***loses*** *the game iff there is no labeling $\mathcal{L} \in \mathcal{L}_\mathcal{S}(F_c, \varphi_o)$ such that $\mathcal{L} \Vdash \varphi_i$*

For the configuration from Figure 6.13 the only $\mathcal{MKD}$ labeling is $(\{n_1, n_2, nn\}, \{yy, ny, yn\}, \varnothing, \varnothing)$. However, this labeling does not satisfy the goals of either agent. Let us see what $\alpha_1$ can choose as a move in order to satisfy its goals. For this, we need to compute $\mathcal{MKD}$ labelings constrained by $\varphi_1 = ny$. One such labeling is $(\{n_1, ny\}, \{yy, yn, nn\}, \varnothing, \{n_2\})$. so $\alpha_1$ can use the epistemic argument $d$ to defeat the ignored argument $n_2$.

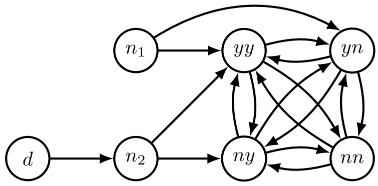



**Figure 6.14:** *Game configuration after $\alpha_1$'s second move.*

The resulting configuration is depicted in Figure 6.14 and its minimal decided labelings constrained by $\varphi_o$ are:

$$\begin{aligned}&(\{d, n_1, ny\}, \{n_2, yy, yn, nn\}, \varnothing, \varnothing)\\&(\{d, n_1, nn\}, \{n_2, yy, yn, ny\}, \varnothing, \varnothing)\end{aligned} \tag{6.48}$$

The first one satisfies $\varphi_1$, but the other does not. Nevertheless, agent $\alpha_1$ improved its status after its move, as now it possibly wins the game.

### 6.4.7 Discussion

In what follows we discuss three important issues related to the argumentation-based negotiation game that we have proposed in this section. First of all, we reiterate the importance of open labelings for deciding the agents' actions during the game. Furthermore, we discuss the implementation of the game as a multi-agent system that uses abstract argumentation for knowledge representation. Last, but not least, we compare our approach to existing literature on the use of abstract argumentation for multi-agent systems.

We start by discussing the use of conditional open labelings in the game. We have seen that the winners of the game can be decided by using such labelings and also that such labelings can help agents identify the arguments that should be defeated in order to achieve their goals. If the moving agent wants to be sure of victory, it may choose to use a unique labeling even if the winners are not decided as such.

Furthermore, it is important that the agents also account for the outcome consistency constraint. If, through their moves, the condition is always satisfied, the result of the game will be decided without ignoring any argument, thus by means of a constrained complete labeling. If, in addition, the agents use unique constrained labelings while computing the set of arguments to defeat, they will ensure that the resulting framework will have a single complete labeling, so, if that labeling does satisfy the goals of the agent, the agent will surely win the game. Otherwise, only the possibility of winning can be obtained.

This turn-based game can easily be turned into an implementation of a multi-agent system that uses argumentation for knowledge representation. The important aspect to note here is that the game was defined using abstract argumentation, so the model imposes no restriction on the logic to be used for knowledge representation in an actual implementation.

We have seen that in the theoretical model all arguments and attacks are somehow given in advance. In a real scenario, this will no longer be the case. The agents will work with knowledge bases from which they will have to build arguments. However, the attacks between arguments will be implicit in the sense that the other agents will also understand them, as they will be based on the logical content of the arguments.

Furthermore, the goals of the agents may change in time. The game can still be used whenever there is need for agreement between agents, for example for planning, but also for auctions, legal deliberation or persuasion.

An important feature of a system based on the game we have presented in this section is the fact that the model relies only on abstract argumentation. Most approaches that use argumentation for multi-agent systems only use it as a tool for a specific part of the agent behavior. The use of argumentation in artificial intelligence is discussed in (Rahwan and Simari, 2009), where several chapters are dedicated to its use for multi-agent systems.

Many works have outlined the benefits of using argumentation for various tasks in a multi-agent system, such as persuasion dialogues (Prakken, 2004), decision making (Amgoud, 2005) or belief revision (?Cayrol et al., 2008; Coste-Marquis et al., 2007), but there is little work on using argumentation as a base for the agent architecture and behavior. A step in this direction has been taken in (Kakas et al., 2011), where the authors propose a fully integrated argumentation-based agent architecture.

In the model proposed in (Kakas et al., 2011), an agent consists of several modules, each of them responsible for a specific task. A local argumentation theory in each module manages the task of the model by providing decisions. The inter-module and inter-agent communication is also based on argumentation theories that are included in each module. The approach relies on value-based argumentation (Bench-Capon, 2003) for building the argumentation theories.

The possible outcomes or decisions of a module are considered separately, they are not seen as arguments. On the other hand, arguments that support distinct decisions are taken to be conflicting. Furthermore, each argument is characterized by some parameters for which a partial ordering is available. In our approach we consider the possible outcomes as arguments and assume that arguments that support one outcome attack the outcomes that are incompatible with it.

We argue that the model proposed in (Kakas et al., 2011) is more complex in terms of the information that is used for building the argumentation theories (outcomes, arguments, parameters, partial ordering of parameters, arguments supporting a particular outcome etc.), whereas our approach only needs abstract arguments, attacks and constraints. On the other hand, our approach is not modular, which may lead to frameworks that are too large to work with. The use of unique decided labelings, which are related to the grounded extension, can help deal with this problem.

## 6.5 Chapter Summary

In this chapter we have turned to the use of logic with argumentation frameworks as a tool for specifying constraints for argument labelings. We proposed several labeling-based semantics and discussed their use for an argumentation-based negotiation game that can easily be translated to a multi-agent system based only on abstract argumentation frameworks and propositional logic constraints.

In Section 6.1 we have discussed the main works that have inspired our approach: argument labelings (Caminada, 2006a), constrained argumentation frameworks (Coste-Marquis et al., 2006) and conditional labelings (Boella et al., 2011).

We have introduced 16 novel labeling-based semantics in Section 6.2. They rely on ignoring some of the arguments from the framework, then doing a complete labeling of the remaining arguments. Several constraints can be applied to such labelings and still lead to universally-defined semantics. We have compared our approach to the works that inspired it, including the enhanced preferred semantics, which was presented in Section 2.8.

In Section 6.3 we discussed a negotiation scenario for selling two items, modeled as a turn-based game whose current configuration can be described with an argumentation framework. The approach relies on distinguishing three kinds of arguments: outcome arguments for the possible results of the negotiation, practical arguments for the actions that the participants can execute (their intentions) and epistemic arguments for their knowledge. The result of the game is decided using labelings and the special labelings that we have proposed can be used throughout the game for deciding the right arguments to attack in order to satisfy the goals of the agents, which are expressed as propositional logic constraints.

In Section 6.4 we have turned the negotiation scenario from Section 6.3 into a formal argumentation-based negotiation game. We have discussed the types of arguments and their meaning, the representation of agents' abilities and knowledge as arguments, as well as the mechanics of the game. We provided the formal representation of the negotiation scenario, but also another example to exhibit the applicability of open labelings to this game. We have also discussed the game as a possible foundation for the implementation of a multi-agent system based on abstract argumentation and constraints only.

*Chapter 7*

# Conclusions

The introductory section has put forward several research questions. Here we will see how these questions have been answered in the thesis, emphasizing our contributions and providing ideas for future research.

## 7.1 Properties of Argumentation Semantics

### Contributions

In Chapter 2, along with the introduction to abstract argumentation, we have provided an extensive survey of argumentation semantics, covering a total of 43 semantics (but also including several auxiliary concepts, such as ideal sets, that we have promoted as semantics). To the best of our knowledge, a similar survey (with respect to the number of considered semantics) is not available in the literature.

In Chapter 3 we have discussed several properties of argumentation semantics and also the relations between semantics, also providing an intuitive graphical representation.

First of all, we have considered cardinality properties (unique status vs multiple status, universally defined), conflict-freeness, admissibility, reinstatement (strong, weak and $\mathcal{CF}$), non-interference and directionality. We have shown that enforcing the satisfaction various combinations of admissibility and the three kinds of reinstatement leads to 8 principle-based semantics, out of which 3 are the well known $\mathcal{CF}$, $\mathcal{AS}$ and $\mathcal{CO}$, while the remaining 5 are novel semantics: $\mathcal{CF}^{cr}$, $\mathcal{CF}^{wr}$, $\mathcal{CF}^{cw}$, $\mathcal{AS}^{wr}$. We have discussed the satisfaction of the mentioned properties for all the $43+5=48$ semantics. A summary of the results is provided in Table 3.3 and Table 3.4.

For some of the properties (cardinality properties, conflict-freeness, admissibility), the satisfaction was already available in the argumentation literature for all semantics, for the other properties we had to fill in the gaps by providing our own proofs. For example we proved that $\mathcal{CF}1$ satisfies $\mathcal{CF}$-reinstatement and that $\mathcal{EPS}$ satisfies both weak and $\mathcal{CF}$-reinstatement, but not strong reinstatement.

We have also introduced a novel property, which we called additivity. The relevance of this property is practical, for the computation of extensions for large frameworks based on weakly connected components, but also theoretical, as we have related additivity to the normal form of modal formulas that describe argumentation semantics. We have proved that all the considered argumentation semantics satisfy this property, except the enhanced preferred semantics. Furthermore, we have shown that any argumentation semantics that is universally defined and satisfies additivity also satisfies non-interference.

For the satisfaction of the directionality property we had to fill several gaps in the existing literature. The important contribution in this context is the result from Theorem 2, where we show that whenever an argumentation semantics $\mathcal{S}em$ satisfies directionality, the semantics $\mathcal{S}em^S$, $\mathcal{S}em^{ids}$, $\mathcal{S}em^{cids}$ and $\mathcal{S}em^{id}$ also satisfy directionality.

We have discussed the inclusion relations between argumentation semantics, based mostly on existing literature, and have provided a map representation of these relations, but also of the satisfaction of the principles we have considered. The complete map of argumentation semantics, containing all the 48 semantics, is presented in Figure 3.18.

In addition, for all but the complete map, we have provided argumentation frameworks that distinguish each semantics from every non-ancestor semantics that is also on the map, thus providing an example-based characterization of semantics. This characterization can be used for selecting suitable semantics for practical applications, based on their behavior. The approach can also be used in multi-agent scenarios for recognizing the semantics used by other agents when the agents may use different terminologies or may not want to disclose the semantics they are using.

## Future work

In future work we will mainly focus on issues related to strong validation. First of all, we will provide argumentation frameworks that strongly validate the extended map of argumentation semantics presented in Figure 3.18. The challenge in this case consists in the large number of semantics that need to be distinguished, which is expected to lead to argumentation frameworks significantly larger than those provided for the strong validation of the compact map in Subsection 3.2.4.

For the case when the enhanced preferred and the preferred semantics coincide (i.e. we can assume that $\mathcal{EPS}$ is removed from the map), we can rely on the additivity property and satisfy parts of the strong validation using reasonably sized frameworks, then consider their disjoint union. For the case when the two semantics are distinct, we must account for the fact that the corresponding framework can have no non-empty admissible set and identify the groups of argumentation semantics that coincide, just as we did in Subsection 3.2.4.

Another issue that we wish to address in this context is the validation of the inner inclusion relation, which was not enforced here, but is also relevant especially in the context of judgement aggregation as discussed in (Caminada and Pigozzi, 2011).

Furthermore, we wish to discuss the satisfaction of other evaluation principles from (Baroni and Giacomin, 2007) and (Baroni et al., 2011a) with respect to all the argumentation semantics that were discussed in this thesis. While for the more commonly used semantics such results are already available in the literature, for many of them we will have to decide and prove the satisfaction ourselves, just as was the case here for the directionality property, but also for some of the others.

## 7.2 Modal Logic and Argumentation

### Contributions

In Chapter 4 we have provided several important results related to the normal form and the satisfaction of modal formulas. These results were used in Chapter 5 for relating modal formulas to principles of argumentation semantics, but they are also relevant for the modal logic community.

We have discussed a disjunctive normal form equivalent to that proposed in (Fine, 1975), but written in a more compact form, in order to allow formal proofs for longer formulas. We have also defined a disjunctive normal form for the converse modal language. We had to apply constraints on the modal minterms in order to filter out the ones that are not satisfiable. To the best of our knowledge, such a normal form has not been previously proposed in the literature.

We have also introduced a normal form for the global modalities, based on the approach suggested in (Areces and Gorin, 2010) for extracting the global modalities from within the scope of other modalities. We have combined the global normal form of $ML(\Diamond, \mathbf{E})$ with the disjunctive normal form of $ML(\Diamond)$ in order to provide one of the most important results of Chapter 4, namely the construction of models that satisfy global modal formulas based on a syntactical analysis of their normal form. We have proved a similar result for $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ as well.

In Chapter 5 we have discussed the use of $ML(\Diamond, \mathbf{E})$ and $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ for argumentation, using also the results from Chapter 4.

We have proved that if a modal formula from $ML(\Diamond, \mathbf{E})$ or $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ describes an argumentation semantics that is additive, then the formula is equivalent to $\mathbf{A}\psi$, where $\psi \in ML(\Diamond, \mathbf{E})$, respectively $\psi \in ML(\Diamond, \overline{\Diamond}, \mathbf{E})$. This result can be extended to other logics, provided that the global modality can be extracted from within the scope of all the operators of that logic.

We have extended the results from (Grossi, 2010) with a thorough analysis of first degree modal formulas and also with an analysis based on attack and defense, following the approach proposed in (Gratie and Florea, 2010). The latter has also provided us with two novel argumentation semantics $\mathcal{CF}^{da}$ and $\mathcal{AS}^{da}$.

We have used the satisfiable model construction algorithm for $ML(\Diamond, \mathbf{E})$ in order to prove that the only argumentation semantics that provides complete extensions, is universally defined, additive and can be described with a global modal formula is the complete semantics. This limitative result tells us that the global modal language is not expressive enough to satisfy several evaluation principles at the same time.

We have also provided a full picture of the use of the global modal language for describing argumentation semantics in Figure 5.9, where we have distinguished three categories:

- semantics that can be described with $ML(\Diamond, \mathbf{E})$ formulas (the ones covered in (Grossi, 2010) plus $\mathcal{CF}^{sr}$)

- semantics that cannot be described with $ML(\Diamond, \mathbf{E})$ either based on the limitative result mentioned above or based on total bisimulation examples that we have provided
- semantics that are yet undecided: we assume that no global modal formula can describe them, but a total bisimulation proof cannot be provided because they can be described using $\mu$-calculus extended with the global modality, which is total bisimulation invariant

We have also discussed the added value of using the converse modality for argumentation. The most important advantage is the fact that we can regard the model as an argumentation framework and we no longer need to use the converse of the accessibility relation. We have also identified other advantages, including a description for the $\mathcal{CF}^{cr}$ semantics.

On the other hand, we have managed to extend the limitative result from $ML(\Diamond, \mathbf{E})$ to cover $ML(\Diamond, \overline{\Diamond}, \mathbf{E})$ as well. The main challenge in this case was the number of minterms in the normal form of the formula for the complete semantics. We have seen that although the translation of a formula can have 4 billion minterms, filtering out the ones that are not satisfiable and reducing the modal degree by 1 resulted in just 14 minterms., which shows that the approach that we have proposed is indeed worthwhile.

We have also covered the remaining semantics with total converse bisimulation examples or argued that such examples cannot be available because the corresponding semantics can be described in $\mu$-calculus extended with converse and global modalities.

## Future work

In future work we will extend the results of this thesis by analyzing the modal definability of argumentation semantics with respect to other logics. Already from the results presented here we can see that $\mu$-calculus, extended with the global and converse modalities, can describe several argumentation semantics that can otherwise not be described. Furthermore, the bisimulation examples that we have used in Subsection 5.3.4 can also be applied to this language in order to show that the corresponding semantics cannot be captured. What is left to do in this case is to discuss the argumentation semantics that were covered in this thesis by the limitative results from Theorem 7 and Theorem 8 and not by bisimulation-based proofs.

On the other hand, a language that is only slightly less expressive than $\mu$-calculus is Propositional Dynamic Logic. For example, the iteration operator from PDL can be used instead of the fixpoint operator for describing the grounded semantics. Thus, it is interesting to compare the expressive power of PDL and $\mu$-calculus with respect to the modal definability of argumentation semantics.

Furthermore, we will extend the results related to the satisfiability of modal formulas that involve the global modalities to other logics, such as the aforementioned $\mu$-calculus and PDL, and also try to relate modal formulas to other evaluation principles of argumentation semantics.

## 7.3 Argumentation Semantics for MAS

### Contributions

In Chapter 6 we have taken a different approach and worked with propositional logic within argumentation frameworks, based on the constrained argumentation frameworks approach proposed in (Coste-Marquis et al., 2006).

We have introduced 16 novel labeling-based semantics based on the idea that arguments can be temporarily ignored in order to obtain a satisfactory labeling of the other arguments, then they should be attacked in order to enforce such a labeling. We have proved that our approach ensures that satisfaction of every reasonable constraint, where a reasonable constraint is one that does not require conflicting arguments to be accepted at the same time.

Furthermore, we have used these semantics within an auction scenario which is a refinement of the one we proposed in (Gratie and Florea, 2012a).

We have also turned the scenario into a formal model of an argumentation-based negotiation game. The game relies on abstract argumentation for describing possible outcomes, the intentions of the agents and persuasion arguments. Furthermore, constraints are used for the goals of the agents. The current state of the game is an argumentation framework that can be interpreted in order to decide the winners of the game, based on whether their goal is satisfied or not. We showed that the novel labeling-based semantics that we proposed can help the agents identify favorable moves in this game.

We have also suggested that the game can be turned into a multi-agent system based on abstract argumentation and we have argued that such an approach, while not as modular as the one from (Kakas et al., 2011), is closer to Dung's abstract argumentation frameworks, as it only adds propositional constraints.

### Future work

In future work, we will primarily focus on the properties of the labeling based semantics that we have introduced in this thesis. We are interested in the computational complexity of these semantics and in the development of efficient algorithms for computing at least a subset of the corresponding labelings, possibly based on the substitution approach used in (Boella et al., 2011) for conditional labelings.

We have seen in the thesis that our labeling-based semantics can be used in order to find sets of arguments that should be attacked in order for some constraint to be satisfied. We can also use them for updating the labelings when new arguments are added to the framework.

Furthermore, we are going to discuss the properties of the argumentation-based negotiation game from a game-theoretical perspective. For example we will discuss strategies and possible winners for such games.

We have suggested in the thesis that the negotiation game can also be used for designing an argumentation based multi-agent system. We are going to provide an actual implementation of such a system and perform an experimental evaluation of the approach.

## Publications

During the work on this thesis I have published the following papers:

Gratie, C. and Florea, A. M. (2010). Generic representation for extension-based semantics of argumentation frameworks. In *Proceedings of the 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2010)*, pages 182–187, Washington, DC, USA. IEEE Computer Society.

Gratie, C. and Florea, A. M. (2012a). Argumentation semantics for agents. In Cossentino, M., Weiss, G., Tuyls, K., and Kaisers, M., editors, *European Workshop on Multi-Agent Systems (EUMAS 2011), Maastricht, The Netherlands, 14–15 November 2011, selected and revised papers*. Springer.

Gratie, C. and Florea, A. M. (2012b). Fuzzy labeling for argumentation frameworks. In McBurney, P., Parsons, S., and Rahwan, I., editors, *Proceedings of the 8th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2011), Taipei, Taiwan, 3 May, 2011*, volume 7543 of *Lecture Notes in Computer Science*, pages 1–8. Springer-Verlag.

Gratie, C. and Florea, A. M. (2012c). SCC-recursiveness revisited. In *Proceedings of the 9th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2012), Valencia, Spain, 4 June, 2012.*

Gratie, C., Florea, A. M., and Meyer, J.-J. (2012a). Full hybrid $\mu$-calculus, its bisimulation and application to argumentation. In *Proceedings of the 13th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA 2012)*, volume 7486 of *Lecture Notes in Artificial Intelligence*, pages 181–194. Springer-Verlag.

Gratie, C., Florea, A. M., and Meyer, J.-J. (2012b). General directionality and the local behavior of argumentation semantics. In *Proceedings of the 1st International Conference on Agreement Technologies (AT 2012), Dubrovnik, Croatia, 15-16 October, 2012.*

Olaru, A., Gratie, C., and Florea, A. M. (2010a). Context-aware emergent behaviour in a MAS for information exchange. *Scalable Computing: Practice and Experience (SCPE)*, 11(1):33–42.

Olaru, A., Gratie, C., and Florea, A. M. (2010b). Emergent properties for data distribution in a cognitive MAS. *Computer Science and Information Systems (COMSIS)*, 7(3):643–660.

# Extension-based argumentation semantics for agent interactions – *Summary*

Argumentation plays an important part in human interaction, whether it is used for persuasion, negotiation or simply for sharing one's point of view on a certain topic. Even at the abstract level, choosing the acceptable arguments from a given set of conflicting arguments is a challenging problem, one that was given multiple solutions in the argumentation literature, in the form of argumentation semantics.

An abstract argumentation framework is a directed graph that encodes the attacks between arguments. The content of the arguments is only used for deriving the attack relation and plays no further role in the abstract setting. Instead, the semantics are only defined based on the attacks between arguments. For example, one of the most intuitive constraints that can be imposed is that there should be no conflict between accepted arguments. Another popular requirement is that the accepted arguments are able to defend themselves against attacks coming from rejected arguments. Such constraints can yield more extensions, i.e. sets of accepted arguments, for the same framework.

On the other hand, it is difficult to say which of these constraints are the most appropriate, and this is why the argumentation literature contains a large array of semantics to choose from. It is commonly agreed that there is no "best" argumentation semantics and that, instead, each of them has unique properties that make it more appropriate for some application domain or another. For example, in some applications it might be important that the set of accepted arguments is as large as possible, reclaiming, say, the use of the preferred semantics, while for other applications it might be more desirable to compute the extensions faster, leading to the use of the very restrictive grounded semantics.

The first part of the thesis (Chapter 2 and Chapter 3) provides an extensive survey of existing argumentation semantics and their properties, including both the mainstream proposals and the ones that have received less attention in the literature so far. In particular, the inclusion relations between semantics are captured in a map that can provide useful information for easily comparing novel semantics with existing ones.

As already stated, the selection of an argumentation semantics for a given application can be a challenging task. One can decide on a set of properties that the extensions need to satisfy in the context of the application, then choose between semantics that satisfy those properties. As an alternative approach, we propose the use of argumentation frameworks that can distinguish between a given set of semantics, in the sense that for any two semantics the set difference between the set of extensions given by the first one and that yielded by the second one is non-empty, unless the former is always included in the latter. Thus, for choosing the appropriate semantics, one only needs to analyze a limited number of frameworks and decide which extensions are appropriate with respect to the given application (i.e. no knowledge of the complex properties of the semantics is required).

Argumentation frameworks and the Kripke models from modal logic are similar in the sense that they both give some meaning to a directed graph. This intuition has lead to the use of modal logic for describing argumentation semantics, first proposed in (Grossi, 2010), where the global modal language was used for capturing several semantics. This link between modal logic and abstract argumentation allows for the transfer of algorithms and properties from one domain to the other. In particular, it can help a rational agent to reason about semantics and about the acceptability of sets of arguments.

However, there are many argumentation semantics that cannot be captured by global modal formulas. In recent work, monadic second order logic has been proposed as a language that is expressive enough to capture argumentation semantics. On the other hand, it is not clear whether the global modal language can capture any other semantics aside from those included in Grossi's paper. Furthermore, it is possible that a small increase in the expressive power (while at the same time retaining decidability) might be enough for capturing several additional semantics. The second part of the thesis (Chapter 4 and Chapter 5) deals with these issues.

We discuss the global modal language and the global converse modal language. For both of them we propose an algorithm that constructs models for satisfiable formulas, based on their normal form. Using this approach, we relate the formulas that describe argumentation semantics to properties that are satisfied by those semantics. This leads to a negative result showing that even a small set of properties cannot hold at the same time as modal definability using the chosen languages. For semantics that are not covered by this general result we provide either formulas that describe them or bisimulation-based proofs that such formulas do not exist.

The last part of the thesis (Chapter 6) deals with the use of abstract argumentation in dynamic systems, for modeling the intentions and goals of agents, as well as the state of the environment where they are situated. The proposed model relies on the use of argumentation frameworks together with constraints on the acceptability of certain arguments. We show that the use of traditional extension-based semantics in this context is not appropriate and we propose a novel class of semantics, based on temporarily ignoring a set of arguments in order to satisfy the given constraints. Upon choosing an extension, the corresponding ignored set is to be attacked with new arguments, so as to bring about the desired extension. We use this approach to create a formal model of argumentation-based negotiation and we explain the use of the novel semantics as strategies for the agents.

# Extensie-gebaseerde semantiek van argumentatie voor agent-interacties – *Samenvatting*

Argumentatie speelt een belangrijke rol in menselijke interactie, of het nu wordt gebruikt voor overtuigen, onderhandelen of eenvoudigweg voor het delen van gezichtspunten over een bepaald onderwerp. Zelfs op een abstract niveau is het kiezen van acceptabele argumenten uit een gegeven verzameling van conflicterende argumenten een uitdagend probleem, een dat meerdere oplossingen heeft opgeleverd in de literatuur op het gebied van argumentatie, in de vorm van argumentatiesemantieken.

Een abstract argumentatie-raamwerk is een gerichte graaf dat de aanvallen tussen argumenten codeert. De inhoud van de argumenten wordt alleen gebruikt voor het afleiden van de aanvalsrelatie en speelt geen verdere rol in het abstracte raamwerk. In plaats hiervan wordt de semantiek alleen gebaseerd op de aanvallen tussen argumenten. Bijvoorbeeld, een van de meest intuÃŕtieve randvoorwaarden die kunnen worden opgelegd is dat er geen conflict mag zijn tussen geaccepteerde argumenten. Een andere populaire eis is dat de geaccepteerde argumenten zichzelf kunnen verweren tegen aanvallen vanuit afgewezen argumenten. Zulke randvoorwaarden kunnen meer extensies, d.w.z. verzamelingen geaccepteerde argumenten, opleveren voor hetzelfde raamwerk.

Aan de andere kant is het moeilijk om te zeggen welke van deze randvoorwaarden de meest geschikte zijn en dit is waarom de argumentatie-literatuur een groot arsenaal aan semantieken bevat waaruit men kan kiezen. Er wordt algemeen aangenomen dat er geen 'beste' argumentatie-semantiek bestaat en dat in plaats hiervan elke semantiek unieke eigenschappen heeft die deze semantiek meer geschikt maken voor een bepaald toepassingsgebied. Bijvoorbeeld, in sommige toepassingen zou het belangrijk kunnen zijn dat de verzameling geaccepteerde argumenten zo groot mogelijk is, wat neerkomt op het gebruik van geprefereerde semantiek, terwijl voor andere toepassingen het meer wenselijk zou zijn om de extensies sneller te berekenen, hetgeen leidt tot het gebruik van de zeer restrictieve 'grounded' semantiek.

Het eerste deel van de thesis (Hoofdstukken 2 en 3) geeft een uitgebreid overzicht van bestaande argumentatie-semantieken en hun eigenschappen, inclusief de dominante voorstellen in de literatuur en degene die tot nu toe minder aandacht hebben gekregen. In het bijzonder worden de inclusie-relaties tussen de semantieken gevat in een soort kaart die bruikbare informatie kan verschaffen voor het gemakkelijk vergelijken van nieuwe semantieken met bestaande.

Zoals hierboven gesuggereerd kan de keuze voor een argumentatie-semantiek voor een gegeven toepassing een uitdagende taak zijn. Men kan beslissen aan de hand van een verzameling eigenschappen waaraan de extensies moeten voldoen in de context van de toepassing en dan kiezen tussen de semantieken die aan deze eigenschappen voldoen. Als een alternatieve benadering stellen we het gebruik van argumentatie-raamwerken voor die tussen een verzameling van semantieken kunnen onderscheiden in de zin dat voor willekeurig twee semantieken het verzamelingstheoretisch verschil tussen de extensie gegeven door de eerste en die gegeven door de tweede niet-leeg is, tenzij de eerstgenoemde altijd is bevat in de ander. Aldus is het voor het kiezen van de geschikte semantiek alleen maar nodig om een beperkt aantal raamwerken te analyseren en te beslissen welke extensies geschikt zijn met betrekking tot de gegeven toepassing (d.w.z. geen kennis van de ingewikkelde eigenschappen van de semantieken is vereist).

Argumentatieraamwerken en de Kripkemodellen van modale logica zijn gelijksoortig in de zin dat deze beide een betekenis geven aan een gerichte graaf. Deze intuÃŕtie heeft geleid tot het gebruik van modale logica voor het beschrijven van argumentatie-semantiek, voor het eerst voorgesteld in (Grossi, 2010), waar een globale modale taal werd gebruikt voor het beschrijven van verschillende semantieken. Deze connectie tussen modale logica en abstracte argumentatie maakt het mogelijk dat algoritmen en eigenschappen van het ene domein naar het andere kunnen worden overgevoerd. I.h.b. kan het een rationele agent helpen bij het redeneren over semantiek en over de aanvaardbaarheid van verzamelingen argumenten.

Echter, er zijn vele argumentatie-semntieken die niet kunnen worden gevat door globale modale formules. In recent werk is monadisch tweede-orde logica voorgesteld als een taal die expressief genoeg is om argumentatie-semantiek te kunnen beschrijven. Aan de andere kant is het niet duidelijk of dÃľ globale modale taal andere semantieken kan beschrijven dan die in Grossi's artikel. Verder is het mogelijk dat een kleine verhoging van de expressieve uitdrukkingskracht (met behoud van beslisbaarheid) genoeg zou zijn om verschillende andere semantieken te kunnen behandelen. Het tweede deel van de thesis (Hoofdstukken 4 en 5) gaan hierover.

We bespreken de globale modale taal en de globale 'conversen' modale taal. Voor beide stellen we een algoritme voor dat modellen construeert voor vervulbare formules, gebaseerd op hun normaalvorm. Door gebruikmaking van deze aanpak relateren we de formules die argumentatie-semantieken beschrijven met eigenschappen waaraan deze semantieken voldoen. Dit leidt tot een negatief resultaat dat aantoont dat zelfs een kleine verzameling eigenschappen niet kan gelden tegelijk met de modale definieerbaarheid gebruikmakend van de gekozen talen. Voor semantiek dat niet worden gedekt door dit algemene resultaat geven we hetzij formules die ze beschrijven of bewijzen gebaseerd op bisimulatie dat zulke formules niet bestaan.

Het laatste deel van de thesis (Hoofdstuk 6) gaat over het gebruik van abstracte argumentatie in dynamische systemen, voor het modelleren van de intenties en doelen van agenten, en de toestand van de omgeving waarin ze zijn gesitueerd. Het voorgestelde model rust op het gebruik van argumentatie-raamwerken samen met de randvoorwaarden op de aanvaardbaarheid van bepaalde argumenten. We tonen aan dat het gebruik van de traditionele extensie-gebaseerde semantiek in deze context niet geschikt is en we stellen een nieuwe klasse van semantieken voor, gebaseerd op het tijdelijk negeren van een verzameling argumenten, om aan de gegeven randvoorwaarden te voldoen. Bij het kiezen van de actuele extensie en de corresponderende genegeerde verzameling, moet deze laatste worden aangevallen met nieuwe argumenten, zodat de gewenste extensie resulteert. We gebruiken deze aanpak om een formeel model te maken van argumentatie-gebaseerde onderhandeling en we leggen het gebruik van de nieuwe semantiek uit als strategieÃńn voor de agenten.

# Semantici bazate pe extensii ale sistemelor de argumentare pentru interacțiunile dintre agenți – *Rezumat*

[Rezumat]

Argumentarea joacă un rol important în interacţiunea dintre oameni, fie că este folosită pentru persuasiune, negociere sau pur şi simplu pentru exprimarea unei opinii cu privire la un anumit subiect. Chiar şi la nivel abstract, alegerea argumentelor acceptabile dintr-o mulţime dată, care poate conţine şi conflicte, este o problemă dificilă, care a primit în literatura de specialitate diverse soluţii prezentate ca semantici de argumentare.

Un sistem abstract de argumentare este un graf orientat care codifică relaţia de atac dintre argumente. Conţinutul logic al argumentelor este folosit doar pentru a determina relaţia de atac şi nu joacă niciun rol ulterior în modelul abstract. În schimb, semanticile sunt definite numai în funcţie de atacurile dintre argumente. De exemplu, una dintre cele mai intuitive restricţii care poate fi impusă este ca între argumentele acceptate să nu existe conflicte. O altă restricţie utilizată frecvent este ca argumentele acceptate să se poată apăra împotriva atacurilor care vin din partea argumentelor respinse. Astfel de constrângeri pot genera mai multe extensii (mulţimi de argumente acceptate) pentru acelaşi sistem de argumentare.

Pe de altă parte, este greu de precizat care constrângeri sunt cele mai potrivite, acesta fiind şi motivul pentru care literatura de specialitate conţine un număr mare de semantici între care se poate alege. Este în general acceptat faptul că nu există o "cea mai bună semantică" şi că, în schimb, fiecare semantică are proprietăţi unice care o fac potrivită pentru anumite domenii de aplicare. De exemplu, pentru anumite aplicaţii poate fi important ca numărul de argumente acceptate să fie cât mai mare, conducând de exemplu la utilizarea semanticii "preferred", în vreme ce pentru alte aplicaţii ar putea fi mai important ca extensiile să fie calculate într-un timp cât mai scurt, cerinţă satisfăcută de semantica "grounded".

Prima parte a tezei (Capitolul 2 şi Capitolul 3) prezintă o trecere în revistă a semanticilor propuse până acum în literatura de specialitate şi proprietăţile acestora,

incluzând atât semanticile "de masă" cât şi pe cele care au primit o atenţie limitată până acum. În particular, relaţia de incluziune dintre semantici este prezentată sub forma unei hărţi care poate furniza informaţii utile pentru compararea rapidă a semanticilor noi cu cele deja existente.

După cum s-a precizat anterior, alegerea semanticii de argumentare pentru o anumită aplicaţie este o problemă dificilă. O soluţie posibilă constă în identificarea unor proprietăţi pe care extensiile trebuie să le satisfacă în contextul aplicaţiei, urmată de selecţia unei semantici care garantează acele proprietăţi. Ca metodă alternativă, teza propune utilizarea unor sisteme de argumentare capabile să diferenţieze semanticile între ele, în sensul că pentru oricare două semantici diferenţa dintre mulţimea de extensii generată de prima semantică şi cea generată de a doua este nevidă, cu excepţia cazului în care avem de-a face cu o relaţie de incluziune general valabilă. Astfel, pentru alegerea semanticii potrivite, este suficient să se analizeze un număr restrâns de sisteme de argumentare şi să se decidă care sunt extensiile potrivite ale acestora în raport cu aplicaţia dorită (cu alte cuvinte, nu mai este necesară cunoaşterea proprietăţilor complexe ale semanticilor).

Sistemele de argumentare şi modelele Kripke din logica modală se aseamănă prin faptul că ambele asociază o semnificaţie specială unui graf orientat. Această intuiţie a dus la utilizarea logicii modale pentru descrierea semanticilor de argumentare, propusă pentru prima dată în (Grossi, 2010), unde limbajul modal global a fost folosit pentru a descrie mai multe semantici. Această legătură între logica modală şi argumentarea abstractă permite transferul algoritmilor şi proprietăţilor de la un domeniu la celălalt. În particular, este posibil ca un agent inteligent să raţioneze despre semantici şi despre acceptabilitatea argumentelor.

Cu toate acestea, există multe semantici care nu pot fi descrise cu formule modale. În articole recente, logica monadică de ordinul 2 a fost propusă ca limbaj suficient de expresiv pentru descrierea semanticilor de argumentare. Pe de altă parte, nu este clar dacă logica modală globală poate descrie şi alte semantici decât cele incluse în lucrarea lui Grossi. De asemenea, este posibil ca o creştere mică a expresivităţii limbajului (dar păstrând decidabilitatea) să fie suficientă pentru a descrie mai multe semantici. Partea a doua a tezei (Capitolul 4 şi Capitolul 5) tratează aceste probleme.

Teza prezintă limbajul modal global şi limbajul modal global convers. Pentru ambele limbaje, este propus un algoritm care permite construirea de modele pentru formulele satisfiabile, bazate pe forma normală a acestora. Folosind această abordare, se stabileşte o legătură între formulele care descriu semantici şi proprietăţile acestor semantici, legătură ce determină şi un rezultat negativ conform căruia există o muţime restrânsă de proprietăţi care nu pot fi satisfăcute simultan cu descrierea modală folosind limbajele analizate în teză. Pentru semanticile care nu sunt acoperite de acest rezultat general sunt prezentate fie formule care le descriu, fie demonstraţii bazate pe bisimulare că astfel de formule nu există.

Ultima parte a tezei (Capitolul 6) se ocupă cu utilizarea argumentării abstracte în sisteme dinamice, pentru modelarea intenţiilor şi scopurilor agenţilor inteligenţi, precum şi a stării mediului în care sunt situaţi agenţii. Modelul propus are la bază utilizarea sistemelor de argumentare împreună cu restricţii asupra acceptabilităţii anumitor argumente. Se arată în teză că utilizarea semanticilor de argumentare tradiţionale în acest context nu produce rezultate adecvate şi se propune o clasă nouă de semantici, bazate pe ignorarea temporară a unor argumente pentru a putea satisface restricţiile date. După alegerea unei extensii, argumentele care au fost ignorate trebuie să fie atacate folosind argumente noi, aşa încât extensia dorită să poată fi obţinută prin aplicarea semanticii complete. Această abordare este folosită pentru construcţia unui model formal de negociere bazat pe argumentare abstractă şi care foloseşte semanticile propouse ca strategii pentru agenţii din sistem.

# Bibliography

Amgoud, L. (2005). A general argumentation framework for inference and decision-making. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI 2005)*, pages 26–33.

Areces, C. and Gorin, D. (2010). Coinductive models and normal forms for modal logics. *Journal of Applied Logic*, 8(4):305–318.

Areces, C. and ten Cate, B. (2007). Hybrid logics. In *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 821–868. Elsevier.

Baroni, P., Caminada, M., and Giacomin, M. (2011a). An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410.

Baroni, P., Cerutti, F., Giacomin, M., and Guida, G. (2009). Encompassing attacks to attacks in abstract argumentation frameworks. In *Proceedings of 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009)*, pages 83–94, Verona, Italy.

Baroni, P., Dunne, P. E., and Giacomin, M. (2011b). On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175(3–4):791–813.

Baroni, P. and Giacomin, M. (2007). On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700.

Baroni, P. and Giacomin, M. (2008). Resolution-based argumentation semantics. In *Proceedings of the 2nd Conference on Computational Models of Argument (COMMA 2008)*, pages 25–36, Tolouse, France.

Baroni, P., Giacomin, M., and Guida, G. (2004). Towards a formalization of skepticism in extension-based argumentation semantics. In *Proceedings of the 4th Workshop on Computational Models of Natural Argument (CMNA 2004)*, pages 47–52, Valencia, Spain.

Baroni, P., Giacomin, M., and Guida, G. (2005). SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1–2):162–210.

Bench-Capon, T. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448.

Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal Logic*, volume 27 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.

Boella, G., Gabbay, D., Perotti, A., van der Torre, L., and Villata, S. (2011). Conditional labelling for abstract argumentation. In *Proceedings of the 1st Workshop on Theory and Applications of Formal Argumentation (TAFA 2011)*.

Bradfield, J. and Stirling, C. (2007). Modal $\mu$-calculi. In *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, chapter Modal mu-calculi, pages 721–756. El.

Caminada, M. (2006a). On the issue of reinstatement in argumentation. In Fisher, M., van der Hoek, W., Konev, B., and Lisitsa, A., editors, *Proceedings of the 10th European Conference on Logics in Artificial Intelligence (JELIA 2006)*, volume 4160 of *Lecture Notes in Computer Science*, pages 111–123, Liverpool, UK. Springer.

Caminada, M. (2006b). Semi-stable semantics. In Dunne, P. E. and Bench-Capon, T. J. M., editors, *Computational Models of Argument: Proceedings of COMMA 2006*, volume 144 of *Frontiers in Artificial Intelligence and Applications*, pages 121–130. IOS Press.

Caminada, M. (2007). Comparing two unique extension semantics for formal argumentation: Ideal and eager. In *Proceedings of the 19th Benelux Conference on Artificial Intelligence (BNAIC 2007)*, pages 81–87.

Caminada, M. (2010). An algorithm for stage semantics. In Baroni, P., Cerutti, F., Giacomin, M., and Simari, G. R., editors, *Proceedings of the 2010 conference on Computational Models of Argument: Proceedings of COMMA 2010*, pages 147–158. IOS Press.

Caminada, M. and Gabbay, D. M. (2009). A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145.

Caminada, M. and Pigozzi, G. (2011). On judgment aggregation in abstract argumentation. *JAAMAS*, 22(1):64–102.

Cayrol, C., de Saint Cyr, F. D., and Schiex, M. C. L. (2008). Revision of an argumentation system. In *Proceedings of KR 2008*, pages 124–134.

Cayrol, C. and Lagasquie-Schiex, M.-C. (2005). On the acceptability of arguments in bipolar argumentation frameworks. In *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, pages 378–389. Springer-Verlag.

Coste-Marquis, S., Devred, C., Konieczny, S., Lagasquie-Schiex, M.-C., and Marquis, P. (2007). On the merging of dungâĂŹs argumentation systems. *Artificial Intelligence*, 171:730–753.

Coste-Marquis, S., Devred, C., and Marquis, P. (2005). Prudent semantics for argumentation frameworks. In *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2005)*, pages 568–572, Hong-Kong, China.

Coste-Marquis, S., Devred, C., and Marquis, P. (2006). Constrained argumentation frameworks. In *Proceedings of the 10th International Conference on Practices in Knowledge Representation and Reasoning*, pages 112–122.

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

Dung, P. M., Mancarella, P., and Toni, F. (2007). Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674.

Dunne, P. E., Hunter, A., and McBurney, P. (2009). Inconsistency tolerance in weighted argument systems. In *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2009)*, pages 851–858, Budapest, Hungary.

Dvorak, W., Dunne, P. E., and Woltran, S. (2011). Parametric properties of ideal semantics. In Walsh, T., editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 851–856.

Dvorak, W. and Gaggl, S. A. (2012). Incorporating stage semantics in the SCC-recursive schema for argumentation semantics. In *Proceedings of the 14th International Workshop of Non-Monotonic Reasoning (NMR 2012)*, Rome, Italy.

Dvorak, W., Szeider, S., and Woltran, S. (2012). Abstract argumentation via monadic second order logic. Technical report, Technische Universitat Wien.

Fine, K. (1975). Normal forms for modal logic. *Notre Dame Journal of Formal Logic*, 16(2):305–318.

Gabbay, D. (2009a). Modal provability foundations for argumentation networks. *Studia Logica*, 93(2–3):181–198.

Gabbay, D. M. (2009b). Fibring argumentation frames. *Studia Logica*, 93(2–3):231–295.

Gratie, C. and Florea, A. M. (2010). Generic representation for extension-based semantics of argumentation frameworks. In *Proceedings of the 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2010)*, pages 182–187, Washington, DC, USA. IEEE Computer Society.

Gratie, C. and Florea, A. M. (2012a). Argumentation semantics for agents. In Cossentino, M., Weiss, G., Tuyls, K., and Kaisers, M., editors, *European Workshop on Multi-Agent Systems (EUMAS 2011), Maastricht, The Netherlands, 14–15 November 2011, selected and revised papers.* Springer.

Gratie, C. and Florea, A. M. (2012b). Fuzzy labeling for argumentation frameworks. In McBurney, P., Parsons, S., and Rahwan, I., editors, *Proceedings of the 8th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2011), Taipei, Taiwan, 3 May, 2011*, volume 7543 of *Lecture Notes in Computer Science*, pages 1–8. Springer-Verlag.

Gratie, C. and Florea, A. M. (2012c). SCC-recursiveness revisited. In *Proceedings of the 9th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2012), Valencia, Spain, 4 June, 2012.*

Gratie, C., Florea, A. M., and Meyer, J.-J. (2012a). Full hybrid $\mu$-calculus, its bisimulation and application to argumentation. In *Proceedings of the 13th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA 2012)*, volume 7486 of *Lecture Notes in Artificial Intelligence*, pages 181–194. Springer-Verlag.

Gratie, C., Florea, A. M., and Meyer, J.-J. (2012b). General directionality and the local behavior of argumentation semantics. In *Proceedings of the 1st International Conference on Agreement Technologies (AT 2012), Dubrovnik, Croatia, 15-16 October, 2012.*

Grossi, D. (2010). On the logic of argumentation theory. In van der Hoek, W., Kaminka, G., Lesperance, Y., Luck, M., and Sandip, S., editors, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 409–416. IFAAMAS.

Grossi, D. (2011). An application of model checking games to abstract argumentation. In van Ditmarsch, H., Lang, J., and Ju, S., editors, *Proceedings of the 3rd International Workshop on Logic, Rationality and Interaction (LORI'11)*, number 6953 in LNAI, pages 74—86. Springer.

Jakobovits, H. and Vermeir, D. (1999). Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9:215–261.

Janssen, J., De Cock, M., and Vermeir, D. (2008). Fuzzy argumentation frameworks. In *Proceedings of the 12th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2008)*, pages 513–520.

Kakas, A., Amgoud, L., Kern-Isberner, G., Maudet, N., and Moraitis, P. (2011). Aba: Argumentation based agents. In *ArgMAS 2011*.

Modgil, S. (2006). Hierarchical argumentation. In *Proceedings of the 10th European Conference on Logics in Artificial Intelligence (JELIA 2006)*, pages 319–332, Liverpool, UK.

Olaru, A., Gratie, C., and Florea, A. M. (2010a). Context-aware emergent behaviour in a MAS for information exchange. *Scalable Computing: Practice and Experience (SCPE)*, 11(1):33–42.

Olaru, A., Gratie, C., and Florea, A. M. (2010b). Emergent properties for data distribution in a cognitive MAS. *Computer Science and Information Systems (COMSIS)*, 7(3):643–660.

Prakken, H. (2004). *Commonsense reasoning (lecture notes)*.

Rahwan, I. and Simari, G., editors (2009). *Argumentation in Artificial Intelligence*. Springer.

Sattler, U. and Vardi, M. (2001). The hybrid $\mu$-calculus. In Gore, R., Leitsch, A., and Nipkow, T., editors, *Automated Reasoning*, volume 2083 of *Lecture Notes in Computer Science*, pages 76–91. Springer Berlin / Heidelberg.

Verheij, B. (1996). Two approaches to dialectical argumentation: admissible sets and argumentation stages. In Meyer, J.-J. C. and van der Gaag, L. C., editors, *Proceedings of the 8th Dutch Conference on Artificial Intelligence (NAIC 1996)*, pages 357–368, Utrecht, Netherlands.

Zhang, Z. and Lin, Z. (2010). Enhancing Dung's preferred semantics. In *Proceedings of the 6th International Symposium on Foundations of Information and Knowledge Systems (FoIKS 2010)*, pages 58–75, Sofia, Bulgaria.