Informatics

Unpacking The Argument

A Claim-Centric View On Abstract Argumentation

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktorin der Technischen Wissenschaften

by

Anna Rapberger, MSc
Registration Number 01058006

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Dipl.-Ing. Dr.techn. Stefan Woltran
Second advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Christian Fermüller

The dissertation has been reviewed by:

Francesca Toni

Gerhard Brewka

Vienna, 20th February, 2023

Anna Rapberger

Erklärung zur Verfassung der Arbeit

Anna Rapberger, MSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 20. Februar 2023

Anna Rapberger

Acknowledgements

First of all, I would like to thank my supervisor Stefan Woltran for his outstanding support throughout the last years. I am very grateful to him for encouraging me to pursue a PhD and for all the invaluable discussions which helped me shape my research. I furthermore want to thank my co-supervisor and supervisor of my master's thesis Chris Fermüller, who was the first who introduced me to the world of argumentation. I want to express my gratitude to Francesca Toni and Gerhard Brewka for agreeing to review this thesis and for their valuable feedback.

During the years of my PhD I had the pleasure of meeting, discussing and collaborating with so many nice and inspiring people. I am deeply grateful for this opportunity, and I want to thank all my co-authors and colleagues for all the lively discussions and fruitful collaborations. In particular, thanks to Gerhard Brewka, Ringo Baumann, and Markus Ulbricht who invited me to Leipzig and made a great research stay possible for me, despite the rising Covid wave end of 2020; thanks to Wolfgang Dvořák for his support and help, especially at the beginning of my PhD; and thanks to Markus for being my Wolfgang from time to time. Thanks to Matthias König, Michael Bernreiter, Jan Maly, and Adrian Haret for being great office mates; and thanks to all further colleagues from DBAI, visitors of the institute, and collaborators, including Martin Lackner, Johannes Wallner, Juliane Auerböck, Oliviero Nardi, Giovanni Buraglio, Beatrix Buhl, Yannis Dimopoulos, Leendert van der Torre, and many more. I furthermore would like to thank my fellow PhD students from the LogiCS DK for all the social activities and entertaining coffee breaks, including Sanja Lukumbuzya, Thekla Hamm, Rafael Kiesel, Davide Longo, Maya Olszewski, Kees van Berkel, and many more; in particular, I would like to express my thanks to Anna Prianichnikova for her continuous support and effort for the DK. Thanks to Toni Pisjak who was always welcoming and helpful.

My greatest thanks go to my parents for supporting me and helping me in any possible way. Furthermore, I would like to thank Melchior for many things, in particular for his inspirational passion for mathematics, but also for supporting me in all life decisions, even quitting smoking in the final months of my PhD. Finally, I want to thank my friends for being the best *Wahlfamilie* one could ask for.

This research was supported by the Austrian Science Fund (FWF) through the projects P32830 and W155-N23 and the Vienna Science and Technology Fund (WWTF) through the project ICT19-065. I want to thank both funding agencies for their financial support.

Kurzfassung

Argumentationstheorie ist ein zentraler Bereich in der KI Forschung. Dabei ist Phan Minh Dungs Ansatz, die Argumente als abstrakte Entitäten aufzufassen und deren Akzeptierbarkeit anhand der Beziehungen der Argumente zueinander zu bestimmen, von wesentlicher Bedeutung. Seine abstrakten Argumentationsmodelle erleichtern die Berechenbarkeit der Akzeptanz der Argumente; ihre Darstellung als gerichteter Graph ist außerdem einfach verständlich und kann dadurch insbesondere bei der Analyse von Debatten mit vielen Argumenten hilfreich sein. In der KI Forschung verstehen wir Argumente im Allgemeinen als komplexe Strukturen, die aus verschiedenen Komponenten bestehen. Ein Argument kann auf *Prämissen* aufbauen, *Fakten* und *Evidenzen* miteinbeziehen und logische Schlüsse anwenden; manche dieser Schlüsse können auch anfechtbar sein. Obwohl verschiedene Argumente sich in ihrem Aufbau stark voneinander unterscheiden können, haben sie doch eines gemeinsam: Jedes Argument hat eine *Konklusion*, also eine Behauptung, die von dem Argument unterstützt wird. Die Konklusionen der Argumente nehmen eine zentrale Rolle in der Argumentationstheorie ein. Oft ist es das Ziel einer argumentativen Analyse, bestimmte Konklusionen auf ihre Plausibilität hin zu überprüfen. Außerdem beeinflussen Konklusionen die Struktur eines Argumentationsmodells wesentlich, da Konflikte zwischen Argumenten von deren Konklusionen determiniert werden. In der Argumentationsforschung wird der konklusions-fokusierten Analyse von Argumentationsmodellen allerdings oft eine sekundäre Rolle zugesprochen. Ihr Einfluss in einem System von Argumenten wird oft unterschätzt. Außerdem wird die Akzeptanz von Konklusionen oft als Nebenprodukt der Akzeptanz von Argumenten wahrgenommen, wodurch sich die Argumentationsforschung typischerweise auf die Argumente in einem Modell konzentriert. Im Gegensatz zu Argumenten können Konklusionen jedoch auch mehrmals in einem Argumentationsmodell auftreten. Dadurch ist die Analyse der Konklusionen nicht hinreichend durch die Argumentationsanalyse abgedeckt.

In dieser Arbeit widmen wir uns zentralen Fragen, die rund um die Auswertung der Konklusionen auftreten. Wir arbeiten die zentrale Bedeutung der Konklusionen in anderen nicht-monotonen Theorien heraus. Wir entwickeln ein hybrides Auswertungsverfahren für Argumentationsmodelle, indem wir Konklusionen in die Auswertung miteinbeziehen. Des Weiteren beschäftigen wir uns mit Eigenschaften von verschiedenen Auswertungsverfahren für Konklusionen. Wir charakterisieren ihre Ausdrucksstärke und bestimmen die Komplexität zu entscheiden, ob eine Konklusion akzeptierbar ist, sowie die Komplexität anderer zentraler Entscheidungsprobleme. Wir widmen uns außerdem Problemen in dynamischen Modellen, in denen die Wissensbasen b.z.w. die Argumentationsmodelle um zusätzliche Informationen erweitert werden. Dabei erforschen wir das Äquivalenzverhalten verschiedener Akzeptanzauswertungsverfahren. Wir charakterisieren die sogenannte Starke Äquivalenz, die, in Anlehnung an Äquivalenz in monotonen Logiken, aussagt, dass zwei Argumentationsmodelle, die den gleichen Output liefern, auch nach Hinzufügen weiterer Informationen bezüglich ihrer akzeptierbaren Konklusionen weiterhin übereinstimmen. Wir erforschen außerdem, wie eine Wissensbasis erweitert werden muss, um eine bestimmte Konklusion gültig zu machen. Wir zeigen, dass die starke Äquivalenz sowie das Akzeptanzerzwingen einer Konklusion für unsere Modelle in Polynomialzeit entscheidbar sind. Damit identifizieren wir Fragmente verwandter Formalismen, für die ebenjene Probleme komplexitätstheoretisch schwer lösbar, d.h. mindestens NP-schwer oder coNP-schwer, sind. Zusammenfassend stellt diese Arbeit eine umfangreiche Analyse der Akzeptanz von Konklusionen in Argumentationsmodellen, deren struturellen und komplexitätstheoretischen Eigenschaften und deren Verhalten in dynamischen Situationen dar.

Abstract

The representation of conflicting scenarios in terms of abstract arguments and attacks has been considerably promoted by the work of Dung; his abstract argumentation frameworks (AFs) are a key formalism in AI research nowadays. Claims are an inherent part of each argument; they substantially determine the structure of the abstract representation. Nevertheless, a claim-based analysis is often considered secondary. While fundamental properties of argument acceptability are well understood, only little is known about structural and computational aspects of claim acceptability. However, since a claim can be supported by several arguments, the identification of acceptable claims poses additional challenges that go beyond argument acceptance. Moreover, the strict focus on arguments in the abstract representation restricts the modeling capacities of AFs to problems that do not involve claims in the evaluation. In this thesis, we address these issues by unpacking the abstract arguments: to conduct a thorough analysis of argumentative reasoning processes from a claim-centered view, we utilize extensions of the abstract model that keep track of parts of the inner structure of arguments. A prominent role in this thesis play claim-augmented argumentation frameworks (CAFs), which extend AFs by assigning a claim to each argument. As we will see, this minimal modification will be sufficient to analyze claim acceptance at a very general level.

We analyze fundamental principles of claim semantics, examine their expressiveness, and study the computational complexity of conclusion-focused reasoning. Inspired by certain shortcomings of traditional claim assessment methods, we propose a hybrid approach to evaluate claim acceptance by shifting certain evaluation steps on the level of claims. We furthermore consider dynamics in argumentative settings; thereby, we focus on strong equivalence and on the enforcement of claims. We introduce claim and vulnerability augmented AFs (cvAFs) to capture knowledge base expansions on the abstract level; this formalism extends AFs by identifying arguments with pairs consisting of a claim and a set of vulnerabilities. With our characterizations of strong equivalence and claim enforcement for cvAFs, we obtain tractable fragments for related non-monotonic knowledge representation languages for which these problems are, in general, intractable. In summary, this thesis provides a thorough analysis of fundamental properties of abstract argumentation semantics from the perspective of the claims of the arguments.

Contents

CHAPTER 1

Introduction

> **claim** /ˈklām/ *n* **2 b :** an assertion open to challenge
>
> *The Merriam-Webster Dictionary*

The identification of plausible statements in the presence of inconsistent and conflicting information is an argumentative act. Rational thinking undoubtedly plays a central role in the evaluation of the acceptability of claims. It allows for conclusions to be drawn from a set of evidences or propositions (premises); new findings can be derived from prior knowledge by putting together the right pieces. According to Walton, inferences of these kind are the building blocks of reasoning [182]. Argumentative reasoning, however, goes beyond that: in light of inconsistencies, premises can be challenged, different standpoints compete with each other. The controversial exchange is often considered as an intrinsic part of argumentation. Van Emeren et al. [176] argue that argumentation involves reasoning but is not a distinctive form of it. They identify the possibility to change the acceptability of a claim as a characteristic element of argumentation.

The ability to revise claim acceptability, for instance, by putting forward another argument, distinguishes argumentation from classical logic in a central aspect: in classical logic, the addition of new information does not affect the validity of former beliefs. If a formula φ is a logical consequence of a set of formulae Δ it is impossible to change this by adding new formulae to Δ, i.e., if $\Delta \vdash \varphi$, then $\Delta \cup \{\psi\} \vdash \varphi$ for each formula ψ. In contrast to classical logic, raising a new argument might affect the acceptability of statements: conclusions which have already been verified can get invalid once new information comes up. A statement can be considered plausible unless we learn otherwise. This renders argumentation as a member of non-monotonic theories.

The trustworthy assessment of claims plays an essential role, in particular in light of worldwide information on demand, fake news, and an increasing amount of available data. Thereby, the development of formal reasoning models which are capable of dealing with with non-monotonic inferences is a key aspect. This thesis provides a theoretical analysis of claim acceptability based on Dung's abstract model of argumentation [77] which constitutes one of the most prominent approaches in computational argumentation.

1.1 Argumentation in Artificial Intelligence

Argumentation theory connects several different research areas such as philosophy, psychology, and computer science. Broadly speaking, it is concerned "*with how assertions are proposed, discussed, and resolved in the context of issues upon which several diverging opinions may be held*" [35]. Driven by the rising demand for human-interpretable and explainable intelligent systems, argumentation theory has emerged as a distinct subfield of artificial intelligence (AI) in recent years [35, 154, 10, 17].

Argumentation research in AI aims to provide methods for automated reasoning in the presence of inconsistencies and conflicts. It spans from the development of formal reasoning models and argument structures [44, 173] over the extraction of arguments from text [129] to the design of efficient methods for conflict resolution [66]. Computational argumentation is closely connected to other non-monotonic reasoning paradigms and provides an orthogonal view to logic programming [77, 61]. Moreover, interpreting defaults as argumentative inferences gives a descriptive interpretation of the underlying mechanisms of non-monotonic reasoning in default theories. The evaluation of defeasible statements via the extraction of arguments and conflicts appears in several different settings; the whole procedure is referred to as *argumentation process* [17].

The Argumentation Process. The identification of acceptable statements is often considered as stepwise process. It consists of the following main components:

1. the identification of arguments and conflicts between them;
2. the determination of jointly acceptable arguments; and
3. the identification of justified statements.

Figure 1.1 provides a graphical illustration of the process.

Research regarding the identification of arguments and conflicts encompasses several different aspects. It includes the systematic construction of arguments from a given knowledge base (we call such processes *instantiation procedures*), but also argument mining techniques which are concerned with the extraction of arguments from natural language text [52, 129].

The construction of arguments following logical or inferential reasoning lies in the field of *structured argumentation*. First studies on this matter can be traced back to Aristotle's *topoi* (topics) which provide systematic rules for defeat and inference; researchers put effort in developing argument schemes [173, 181] and reasoning systems [137, 44, 116].

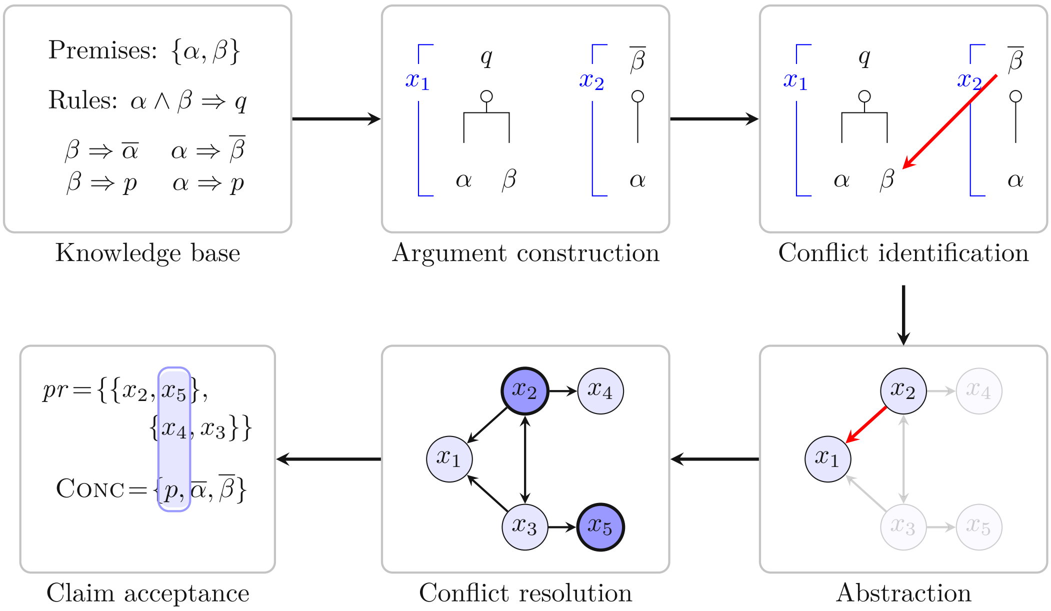


Figure 1.1: The argumentation process.

Broadly speaking, an *argument* is understood as a piece of information which supports, defends, or justifies a particular claim. Logic-based approaches identify the claim of an argument with a logical formula and the support with a set of formulae which derive the claim [41, 6]. In rule-based approaches, the claim is a sentence of a formal language while the support consists of assumptions and facts that infer the claim based on a deductive system [137, 70]. Other systems also consider evidence-based support; examples include juridical argumentation [36, 33] or decision-making procedures in medicine [123]. Conflicts between arguments, often referred to as *attacks*, are often asymmetric. They depend on the contradictions between claims and the *defeasible* elements of an argument (for instance, premises or defeasible inference rules).

After having constructed all arguments and attacks, we arrive at the abstraction step in the process. In his seminal paper [77], Phan Minh Dung reshaped the view on argumentation theory by demonstrating that argumentative settings can be interpreted as graph-like structures. An *abstract argumentation framework (AF)* consists of a set of arguments, treated as atomic objects (the nodes of the graph), and attacks between them (the arcs). The key observation is that argument acceptability can be decided by looking only at the conflicts between arguments. Dung formulated several *argumentation semantics* which are criteria to determine argument acceptance. Evaluating an argumentation framework in regard to a semantics yields different sets of jointly acceptable arguments.These so-called *extensions* reflect the different perspectives which coexist within the argumentation framework. In subsequent works, the initial set of semantics got extended in various ways [177, 53, 60, 79, 55]. To this day, researchers develop new semantics and fine-tune existing ones; also, the study of theoretical, structural and computational properties of argumentation semantics is a highly active research field, including studies on expressiveness [84], their dynamic behavior [18, 131, 179], and computational complexity [89], just to name a few.

In the final step of the process, the acceptance of claims is determined. Typically, this is done by extracting the conclusions of the acceptable arguments. Claim acceptance receives less attention in the literature. It is often treated as a byproduct of argument acceptance, leaving aside crucial differences between both concepts.

1.2 Claim Acceptance

Claim acceptance and argument acceptance are closely intertwined. There is, however, an essential difference: while each argument appears only once in an argumentation framework, a claim can be the conclusion of several arguments. This leads to interesting situations. For instance, an accepted claim can be supported by a rejected argument without affecting the acceptance status of it. Also, a claim can appear as a conclusion of each possible outcome of an argumentation framework although in the intersection of all extensions there is no argument that supports the claim. This is the case if it appears as a conclusion of *some* argument in each set of acceptable arguments. Such claims are also called *floating conclusions* [136]. Intuitively, it is not necessary to make a single argument win; it suffices to find the right arguments under each possible viewpoint. On the other hand, a claim is not rejected as long as not *all* arguments having this claim are defeated.

The difference between claim and argument acceptance has evoked several discussions around the turn of the millennium in the non-monotonic research community [121, 166, 120, 150]. Specifically regarding the treatment of floating conclusions, researchers have shed light on the many different ways to compute claim acceptance. Thereby, it can be observed that formalisms which incorporate a syntactic method when evaluating inconsistent knowledge bases (for instance, by putting the main emphasis on the proofs or arguments) tend to adapt a rather skeptical approach. Often, claims are considered skeptically accepted only if it is the claim of an argument that appears in all possible solutions (hence floating conclusions are not accepted within this approach). This can be observed in the context of inheritance networks [122] and is also the standard approach to claim acceptance in ASPIC+ (although a more generous approach is discussed as a valid alternative [137, Def. 2.18 and below]). Several researchers have criticized this approach as being overly skeptical (cf. [136, 150, 169]). To overcome these obstacles, they proposed alternative ways to compute claim acceptance, drawing certain inspirations from formalisms that put their emphasis on the models (sets of acceptable claims) of the knowledge base. As Lynn Stein points out [169], outlining certain drawbacks of skeptical evaluation methods of inheritance networks, "*if we wish to determine what is true in all possible worlds, we cannot avoid this kind of reasoning.*"

In light of these differences, we believe that the study of claims within argumentation as an independent subject matter is crucial to improve the overall argumentation process in several ways. First, argument and claim evaluation admit several differences. So far, it is not clear how these differences affect fundamental characteristics of argumentation semantics. A successful argumentation process should be able to provide a trustworthy assessment of the statements in question; hence it is evident that a thorough study on that matter is crucially needed. Second, acceptance criteria are often formulated only on the argument level, leaving aside the main differences between the evaluation of arguments and claims. Nevertheless, due to the differences outlined above, it is often not possible to transfer criteria of argumentation semantics to the claim level only by evaluating the arguments involved in the process. Here, we believe that genuine adaptions of argumentation semantics to the claim level can help to preserve the original spirit. Third, claims shape the structure of argumentation frameworks. In the vast majority of instantiation procedures, claims admit a distinct role within arguments as they are responsible for the outgoing attacks. From this observation, we can derive a fundamental property of the attack relation: usually, arguments with the same claim attack the same arguments. This property is referred to as *well-formedness*. Hence we observe that the structure of an argumentation framework strongly depends on the claims of the arguments. We believe that this observation is essential to determine static but also dynamic characterizations of argumentative reasoning.

1.3 Contributions

Due to the strong focus on argument acceptance within the community, it turns out that surprisingly little is known about claim characteristics. Breaking with the tradition of argument-focused research in argumentation,

we view arguments as means to assess the acceptability of claims.

Thereby, we focus on fundamental properties of claim acceptance, computational complexity, and dynamical aspects of reasoning. We tackle these problems by unpacking the arguments in Dung's abstract frameworks:

- We consider the arguments' *claims* explicitly in the abstract representation. We utilize *claim-augmented argumentation frameworks (CAFs)* [92] which extend AFs by assigning a claim to each argument. We assume no particular structure of claims; in the spirit of Dung, we treat claims as atomic objects. As we will see, this minimal generalization allows for analyzing claim acceptance at a very general level.
- To model changes in the underlying knowledge base at the abstract level, we furthermore consider the *vulnerabilities* of arguments as part of the abstract framework (in Chapter 8). With this, we are able to capture hidden weaknesses of the arguments.

In this thesis, we contribute to a thorough understanding of the role of claims in argumentation in several ways.

- In Chapter 3, we survey the role of claims in structured argumentation, logic programming, and their connection to collective attacks. We observe that the notion of well-formedness can be found in many non-monotonic reasoning paradigms. Among our findings is that for complete, grounded, preferred, and stable semantics, each set of acceptable claims corresponds to a *unique* extension for well-formed CAFs. Moreover, we show that well-formedness allows to *merge arguments* with the same claim into a single argument in the abstract representation. Intuitively, merging arguments $a_1, \dots, a_n$ with claim c yields a new argument a with disjunctive support $supp(a_1) \vee \cdots \vee supp(a_n)$ (which combines all arguments for c) where the logical OR belongs to the meta-language. Merging arguments comes at a cost: instead of binary attacks, we obtain collective attacks in the abstract representation, reflecting the observation that a single argument might be too weak to successfully refute an argument that has several independent justifications.

  In this chapter, we show a fundamental correspondence between CAFs and structured argumentation formalisms, logic programs, and AFs with collective attacks (SETAFs) [141]. This correspondence is obtained by evaluating the underlying AF and extract the claims of the successful arguments in the final step (thus following the flow of the argumentation process outlined in Section 1.1). However, it turns out that this procedure does not yield satisfactory results for all semantics. As already observed by Caminada et al. in the context of semi-stable semantics, claim-based and argument-based maximization might result in a different outcome [61, 62]. We extend this result to SETAFs for a larger class of semantics.

- In Chapter 4, we address the mismatch of conclusion- and argument-focused evaluation methods and develop a class of semantics that puts claims into a stronger position when determining the acceptance of arguments and claims. In their conventional treatment, claim semantics are derived from the argument-based evaluation of an AF. On the downside, this shows that several fundamental concepts of argumentation semantics, however, are poorly understood or not even conceptualized on the level of claims (e.g., defeat of claims). Moreover, several semantics from related formalisms cannot be captured on the abstract level under standard instantiation procedures as observed in the previous chapter. We develop genuine notions for claim defeat and claim-set maximization and show that the semantics based on these concepts indeed capture a broader range of conclusion-focused reasoning methods; in particular, we show that L-stable logic programming semantics corresponds to the so-called h-semi-stable semantics for CAFs.

We have identified two different evaluation methods: *inherited* and *hybrid semantics*. The former class of semantics bases the evaluation on argument acceptance and extracts claims in the final step while the latter class considers claims within the evaluation process. The following two chapters are dedicated to a thorough analysis of both classes of conclusion-focused reasoning.

- In Chapter 5, we examine fundamental properties by adapting the *principle-based approach* to argumentation semantics [11, 175]. On the one hand, such a classification yields theoretical insights into the nature of the different semantics and on the other hand, can help to guide the search for suitable semantics appropriate in different scenarios. We introduce novel principles for claim semantics and study well-known properties of argumentation semantics such as e.g., I-maximality, naivety, and reinstatement. We compare hybrid semantics and inherited semantics as well as general CAFs and well-formed CAFs with respect to these properties. Our analysis complements similar studies on classical argumentation semantics and sheds light on the different levels of arguments and claims. For instance, we show that although preferred semantics satisfies the central principle of *I-maximality*, i.e., $\subseteq$-maximality of its extensions, it is not necessarily the case that preferred semantics in terms of claims satisfies this principle.

  We furthermore examine the expressiveness of claim semantics in terms of *signatures* [84]. The characterization of the signature of a semantics, i.e., the set of all possible extension-sets a framework can possess under the given semantics, is key to understand its expressive power. Besides theoretical insights, knowing which extensions can jointly be modeled within a single framework under a given semantics, for instance, is crucial in dynamic scenarios [27]. Among our findings is that claim semantics are in general very expressive, in particular when dropping any structural restrictions on the attack relation.

- In Chapter 6, we investigate the *computational complexity of reasoning*. This chapter complements previous studies on the complexity of inherited semantics presented in [92]. We settle the computational complexity of all the hybrid semantics, i.e. stable, naive, preferred, semi-stable, and stage semantics, for the main decision problems of credulous and skeptical acceptance, verification, and testing for non-empty extensions. Among our findings is that for naive semantics, the hybrid variant is harder than its inherited counterpart, while for preferred semantics, it is the inherited variant that shows higher complexity.

  We furthermore determine the complexity of the *concurrence problem*, i.e. whether for a given CAF and a semantics, the inherited and hybrid variant of that semantics coincide. Note that showing this problem to be easy would suggest that there are relatively natural classes of CAFs which characterize whether or not the two variants collapse. However, as we will see, concurrence can be surprisingly hard, up to the third level of the polynomial hierarchy.

So far, we have conducted a *static* analysis of the semantics by investigating fundamental properties. In the following two chapters, we consider *dynamic* scenarios instead. Thereby, we focus in particular on *strong equivalence*: given a knowledge base $\mathcal{K}$, is it possible to replace a subset $\mathcal{H}$ of $\mathcal{K}$ by an equivalent one, say $\mathcal{H}'$, without changing the meaning of $\mathcal{K}$? Within the KR community it is folklore that this is usually not the case when considering non-monotonic formalisms. Driven by this observation, the notion of strong equivalence has been proposed, developed and examined in various contexts [132, 142].

The adaption of the concept to conclusion-focused reasoning passes through a number of stages.

- First, inspired by the classical treatment within abstract argumentation, we define strong equivalence for CAFs as follows: we say that two CAFs are *strongly equivalent* to each other if and only if they yield the same outcome under all possible expansions (i.e., the addition of new arguments and attacks).

  In Chapter 7, we provide characterization results of strong equivalence between CAFs via so-called *kernels*, i.e., semantics-dependent sub-frameworks, for each CAF semantics we consider in this work. Moreover, we discuss ordinary equivalence for CAFs and present dependencies between semantics for this weaker equivalence notion. We furthermore present a rigorous complexity analysis of these concepts. We show that deciding ordinary equivalence can be computationally hard, up to the third level of the polynomial hierarchy.

- When considering the proposed definition in relation to knowledge bases, we, however, run into an issue. Consider a scenario in which we instantiate a knowledge base following our standard argumentation procedure twice, using two different argument naming schemes (e.g., in the first instantiation, arguments are called $a_1, \dots, a_n$ and in the second instantiation, arguments are called $b_1, \dots, b_n$). We obtain two frameworks which are topologically equivalent but disagree on their argument names. The addition of new arguments and attacks will result in different modifications; hence the two frameworks are not strongly equivalent to each other although they encode the same knowledge base.

  In order to overcome this issue, we relax the initial definition as follows: two CAFs are *renaming strongly equivalent* to each other with respect to a semantics if and only if it is possible to find an appropriate argument renaming such that they are strongly equivalent with respect to this semantics. We discuss these novel equivalence concepts in Chapter 7.

- Finally, we consider strong equivalence with respect to a concrete formalism that allows for instantiating a knowledge base as an abstract framework, namely assumption-based argumentation (ABA). We make the fundamental observation that not only the claim but also the *hidden weaknesses* of an argument play a crucial role when determining strong equivalence of two knowledge bases by considering only the abstract level. We extend CAFs accordingly: we assign to each argument a set of vulnerabilities which encodes all possible attack points (so-called *cvAFs*).

  In Chapter 8, we study enforcement and strong equivalence for ABA. We show that both problems lie on the first level of the polynomial hierarchy and are thus intractable. Furthermore, we show that both problems are tractable for cvAFs. Based on results from Chapter 3 in which we identified a fragment of ABA which is in one-to-one correspondence with CAFs, we obtain a fragment of ABA for which enforcement and strong equivalence is tractable. In a similar manner, we obtain tractability results for logic programs as well.

We conclude in Chapter 9 with an overview of our results and discussion of related work.

1.4 Publications

The results presented in this thesis are based on the following publications (we note that authors are listed alphabetically):

[103] Wolfgang Dvořák, Alexander Greßler, Anna Rapberger, and Stefan Woltran. The complexity landscape of claim-augmented argumentation frameworks. *Artificial Intelligence*, page 103873, 2023. ISSN 0004-3702.

[29] Ringo Baumann, Anna Rapberger, and Markus Ulbricht. Equivalence in argumentation frameworks with a claim-centric view - classical results with novel ingredients. In *36th AAAI Conference on Artificial Intelligence (AAAI'22), Proceedings*, pages 5479–5486. AAAI Press, 2022.

[158] Anna Rapberger and Markus Ulbricht. On dynamics in structured argumentation formalisms. In *19th International Conference on Principles of Knowledge Representation and Reasoning (KR'22), Proceedings*, pages 288–298, 2022.

[127] Matthias König, Anna Rapberger, and Markus Ulbricht. Just a matter of perspective: Intertranslating expressive argumentation formalisms. In *Computational Models of Argument (COMMA'22), Proceedings*, pages 212–223. IOS Press, 2022.

[99] Wolfgang Dvořák, Alexander Greßler, Anna Rapberger, and Stefan Woltran. The complexity landscape of claim-augmented argumentation frameworks. In *35th AAAI Conference on Artificial Intelligence (AAAI'21), Proceedings*, pages 6296–6303. AAAI Press, 2021.

[98] Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. Argumentation semantics under a claim-centric view: Properties, expressiveness and relation to SETAFs. In *17th International Conference on Principles of Knowledge Representation and Reasoning (KR'20), Proceedings*, pages 341–350, 2020.

[97] Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. On the relation between claim-augmented argumentation frameworks and collective attacks. In *24th European Conference on Artificial Intelligence (ECAI'20), Proceedings*, volume 325 of *FAIA*, pages 721–728. IOS Press, 2020.

[157] Anna Rapberger. Defining argumentation semantics under a claim-centric view. In *9th European Starting AI Researchers' Symposium (STAIRS'20), Proceedings*, volume 2655 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020.

Chapter 3 composes work presented in [127, 97]. In the present work, we extend these results by identifying fragments of the formalisms under consideration which are in one-to-one correspondence.

Results from Chapter 4 have been published in [98, 157]. An extended version has been submitted to *Artificial Intelligence* and is currently under review (available under [102]).

The complexity analysis presented in Chapter 6 has been published in [99, 103].

The principle-based analysis of the semantics presented in Chapter 5 is part of an extended version of [98] (cf. [102]). This version furthermore contains expressiveness results for all considered semantics which have been covered only partially in the conference version.

Most of the results presented in Chapter 7 have been published in [29]. The present version extends the conference paper by novel results regarding well-formed CAFs. A long version of this paper has been submitted to *Journal of Artificial Intelligence Research*. Chapter 8 has been published in [158], an extended version has been accepted for publication (with minor revision) in the *Journal of Artificial Intelligence Research*.

Remark 1.4.1. *In the present work, we revised notation and names of the CAF semantics. Originally, in [157, 98], we introduce hybrid semantics under the name* claim-level semantics*. In [157, 98] and subsequent work, we denote hybrid (former: claim-level) semantics by cl-σ and inherited semantics by σ_c. Accordingly, we have adapted the notation: in the present work, we write σ_i instead of σ_c to denote the inherited variant of the semantics σ; likewise, we write σ_h instead of cl-σ to denote the hybrid variant.*

The author of this thesis furthermore co-authored the following papers.

[39] Michael Bernreiter, Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. The Effect of Preferences in Abstract Argumentation Under a Claim-centric View. In *37th AAAI Conference on Artificial Intelligence (AAAI'23), Proceedings*, to appear.

[159] Anna Rapberger, Markus Ulbricht, and Johannes Peter Wallner. Argumentation Frameworks Induced by Assumption-Based Argumentation: Relating Size and Complexity. In *20th International Workshop on Non-Monotonic Reasoning (NMR'22), Proceedings*, volume 3197 of *CEUR Workshop Proceedings*, pages 92–103. CEUR-WS.org, 2022.

[38] Michael Bernreiter, Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. The Effect of Preferences in Abstract Argumentation Under a Claim-centric View. In *20th International Workshop on Non-Monotonic Reasoning (NMR'22), Proceedings*, volume 3197 of *CEUR Workshop Proceedings*, pages 27–38. CEUR-WS.org, 2022.

[119] Thekla Hamm, Martin Lackner, and Anna Rapberger. Computing Kemeny rankings from d-euclidean preferences. In *7th International Conference on Algorithmic Decision Theory (ADT'21), Proceedings*, volume 13023 of *LNCS*, pages 147–161. Springer, 2021.

[100] Wolfgang Dvořák, Matthias König, Anna Rapberger, Johannes Peter Wallner, and Stefan Woltran. ASPARTIX-V - a solver for argumentation tasks using ASP. In *Workshop on Answer Set Programming and Other Computing Paradigms (ASPOCP'21)*, 2021.

[94] Wolfgang Dvořák, Sarah Alice Gaggl, Anna Rapberger, Johannes Peter Wallner, and Stefan Woltran. The ASPARTIX system suite. In *Computational Models of Argument (COMMA'20), Proceedings*, volume 326 of *FAIA*, pages 461–462. IOS Press, 2020.

[96] Wolfgang Dvořák, Anna Rapberger, Johannes Peter Wallner, and Stefan Woltran. ASPARTIX-V19 - an answer-set programming based system for abstract argumentation. In *Foundations of Information and Knowledge Systems - 11th International Symposium (FoIKS'20), Proceedings*, volume 12012 of *LNCS*, pages 79–89. Springer, 2020.

[104] Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. On the different types of collective attacks in abstract argumentation: equivalence results for SETAFs. *Journal of Logic and Computation*, 30(5):1063–1107, 06 2020.

[95] Wolfgang Dvořák, Anna Rapberger, and Johannes Peter Wallner. Labelling-based algorithms for SETAFs. In *3rd International Workshop on Systems and Algorithms for Formal Argumentation (SAFA'20), Proceedings*, volume 2672 of *CEUR Workshop Proceedings*, pages 34–46. CEUR-WS.org, 2020.

[93] Wolfgang Dvořák, Anna Rapberger, and Stefan Woltran. Strong equivalence for argumentation frameworks with collective attacks. In *42nd German Conference on AI (KI'19), Proceedings*, volume 11793 of *LNCS*, pages 131–145. Springer, 2019.

The work most closely related to the present thesis is [38, 39], which continues the line of claim-centric research in argumentation by studying the effect of preferences in this context. In this matter, we furthermore want to highlight our work on assumption-based argumentation [159] and on argumentation frameworks with collective attacks (SETAFs) [104, 97, 93]. As we will discuss in Chapter 3, they admit a close connection to claim-augmented argumentation. We note that [104] is an extended version of [93].

The work in [119] lies in the research area of computational social choice and focuses on the Kemeny rank aggregation function.

In [100, 96, 94], we present (and further develop) the ASPARTIX system (`https://www.dbai.tuwien.ac.at/proj/argumentation/systempage`).

CHAPTER 2

Background

In this chapter, we introduce abstract and claim-augmented argumentation frameworks; moreover, we fix notations used throughout the thesis. We recall other formalisms and concepts from the literature on the fly. Below we give pointers to the respective sections.

We assume familiarity with basic concepts from classical propositional logic. Below, we clarify notations used in this work.

Classical propositional logic. We use the standard connectives such as logical OR ($\vee$), AND ($\wedge$), and negation ($\neg$). Propositional formulae range over a set of propositional variables which we denote by lower case roman letters at the end of the alphabet (i.e., x, y, z, u, v). An *atomic formula* is a formula without any logical connectives (hence the simplest well-formed formulae in the logic). *Literals* are formulae of the form x and $\neg x$ for propositional variables x. We use Greek lower case letters to denote formulae. We evaluate formulae according to the standard semantics of propositional logic. An interpretation $I : X \to \{true, false\}$ (or $\{T, F\}$, or $\{0, 1\}$) assigns truth values to all propositional variables X of a formula φ; it is a *model* of φ if assigns *true* to the formula. We associate a model I of a formula φ with the set of atoms that are set to true under I.

Let L denote a set of literals and let $C \subseteq 2^L$. A formula is in *conjunctive normal form (CNF)* if it is of the form $\bigwedge_{c \in C} \bigvee_{x \in c} x$; it is in *disjunctive normal form (DNF)* if it is of the form $\bigvee_{c \in C} \bigwedge_{x \in c} x$.

2.1 Abstract Argumentation

We introduce abstract argumentation frameworks [77]; for a comprehensive introduction, we refer the interested reader to [17]. We fix U as countable infinite domain of arguments. As done in many other works on abstract argumentation, we focus on frameworks with a *finite* set of arguments.

Definition 2.1.1. *An* argumentation framework (AF) *is a pair* (A, R) *where* $A \subseteq U$ *is a finite set of arguments and* $R \subseteq A \times A$ *is the attack relation. By* $\mathfrak{A}$ *we denote the class of all AFs.*

AFs can be represented as directed graphs where the nodes correspond to the arguments and the arcs correspond to the attack relation. Figure 2.1 gives an example of an AF.

Definition 2.1.2. *Let* $F = (A, R)$ *be an AF. We say that* $a \in A$ attacks $b \in A$ *iff* $(a, b) \in R$*. We call* b *an* attacker *of* a*. A set of arguments* $E \subseteq A$ attacks b *iff* $(a, b) \in R$ *for some* $a \in E$*.* E *attacks the set* $D \subseteq A$ *iff there is an* $a \in D$ *such that* E *attacks* a*.*

Definition 2.1.3. *Let* $F = (A, R)$ *be an AF. We write* $a_F^+ = \{b \mid (a, b) \in R\}$ *and* $a_F^- = \{b \mid (b, a) \in R\}$ *to denote the set of all arguments attacked by resp. attacking the argument* $a \in A$*. For a set of arguments* $E \subseteq A$*, we write* $E_F^+ = \bigcup_{a \in E} a_F^+$ *and* $E_F^- = \bigcup_{a \in E} a_F^-$*. We call* $E_F^\oplus = E \cup E_F^+$ *the* range *of* E *in* F*. If no ambiguity arises, we drop the subscript* F*.*

The *range* of a set of arguments E consists of all arguments which are accepted and rejected by E. For instance, considering our example AF F from Figure 2.1, the set of arguments $\{x_1, x_3\}$ attacks the arguments x_0, x_2, and x_4, hence its range is $\{x_0, x_1, x_2, x_3, x_4\}$.

A crucial notion in abstract argumentation is *defense*. If an argument it challenged, it is possible to reinstate its plausibility by finding appropriate counter-arguments.

Definition 2.1.4. *Let* $F = (A, R)$ *be an AF. A set of arguments* $E \subseteq A$ defends *an argument* $a \in A$ *iff* E *attacks each attacker of* a *(equivalently,* $a_F^- \subseteq E_F^+$*).*

In our running example, the argument x_1 defends the argument x_3 against the attacker x_2.

Equally central is the notion of *conflict-freeness* which formalizes the reasonable assumption that jointly acceptable arguments should not contradict each other.

Definition 2.1.5. *Let* $F = (A, R)$ *be an AF. We call a set of arguments* $E \subseteq A$ conflict-free *(in* F*) iff* E *does not attack itself. Otherwise, we call* E conflicting *(in* F*). We write* $cf(F)$ *to denote the set of conflict-free sets in* F*.*

Next we turn to *argumentation semantics*. In his seminal paper [77], Dung introduced several semantics which serve as criteria to determine argument acceptance in an AF. Formally, they are defined as functions $\sigma : \mathfrak{A} \to 2^{2^U}$ which assign to each AF $F = (A, R)$

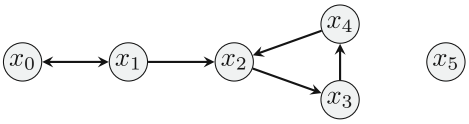


Figure 2.1: An example AF $F = (A, R)$.

a set $\sigma(F) \subseteq 2^A$ of extensions. Conflict-freeness already serves as example of an argumentation semantics. Apart from this fundamental yet a little bit naive semantics, Dung considered *admissible*, *grounded*, *preferred*, and *stable semantics*. Subsequent work extends this initial set of semantics in various ways. In this thesis, we will furthermore consider *naive*, *semi-stable* [177, 53, 60] and *stage semantics* [177]. For an overview over these and many more argumentation semantics, we refer the reader to [16].

Definition 2.1.6. *Let $F = (A, R)$ be an AF and consider a set $E \in cf(F)$. We say that*

- *E is* admissible *(in F) ($E \in ad(F)$) iff iff each $a \in E$ is defended by E in F;*
- *E is* complete *(in F) ($E \in co(F)$) iff if $E \in ad(F)$ and each $a \in A$ defended by E in F is contained in E;*
- *E is* grounded *(in F) ($E \in gr(F)$) iff E is $\subseteq$-minimal in $co(F)$;*
- *E is* preferred *(in F) ($E \in pr(F)$) iff E is $\subseteq$-maximal in $ad(F)$;*
- *E is* stable *(in F) ($E \in stb(F)$) iff $E_F^{\oplus} = A$;*
- *E is* naive *(in F) ($E \in na(F)$) iff E is $\subseteq$-maximal in $cf(F)$;*
- *E is* semi-stable *(in F) ($E \in ss(F)$) iff $E \in ad(F)$ and there is no $D \in ad(F)$ with $E_F^{\oplus} \subset D_F^{\oplus}$;*
- *E is* stage *(in F) ($E \in stg(F)$) iff there is no $D \in cf(F)$ with $E_F^{\oplus} \subset D_F^{\oplus}$.*

For a semantics σ, we call E a σ-extension of F. If we want to emphasize that the extensions contain arguments, then we speak of argument-extensions*.*

Example 2.1.7. *The AF F from Figure 2.1 admits the following extensions:*

- $cf(F) = \{\emptyset, \{x_0\}, \{x_0, x_2\}, \{x_0, x_2, x_5\}, \{x_0, x_3\}, \{x_0, x_3, x_5\}, \{x_0, x_4\}, \{x_0, x_4, x_5\}, \{x_0, x_5\}, \{x_1\}, \{x_1, x_3\}, \{x_1, x_3, x_5\}, \{x_1, x_4\}, \{x_1, x_4, x_5\}, \{x_1, x_5\}, \{x_2\}, \{x_2, x_5\}, \{x_3\}, \{x_3, x_5\}, \{x_4\}, \{x_4, x_5\}, \{x_5\}\}$
- $ad(F) = \{\emptyset, \{x_0\}, \{x_1\}, \{x_5\}, \{x_0, x_5\}, \{x_1, x_3\}, \{x_1, x_5\}, \{x_1, x_3, x_5\}\}$
- $co(F) = \{\{x_5\}, \{x_0, x_5\}, \{x_1, x_3, x_5\}\}$

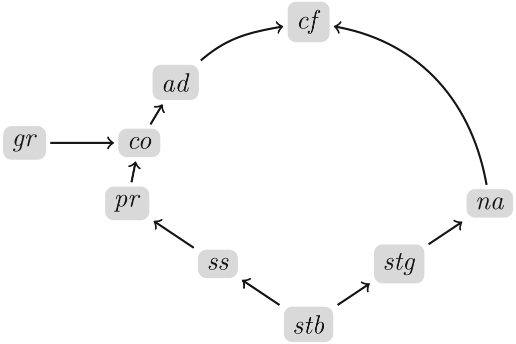


Figure 2.2: Relations between AF semantics. An arrow from σ to τ indicates that $\sigma(\mathcal{F}) \subseteq \tau(\mathcal{F})$ for each AF F.

- $gr(F) = \{\{x_5\}\}$
- $pr(F) = \{\{x_0, x_5\}, \{x_1, x_3, x_5\}\}$
- $stb(F) = ss(F) = stg(F) = \{\{x_1, x_3, x_5\}\}$
- $na(F) = \{\{x_0, x_2, x_5\}, \{x_0, x_3, x_5\}, \{x_0, x_4, x_5\}, \{x_1, x_3, x_5\}, \{x_1, x_4, x_5\}\}$

We recall that for each AF F,

$$stb(F) \subseteq stg(F) \subseteq na(F) \subseteq cf(F) \text{ and } stb(F) \subseteq ss(F) \subseteq pr(F) \subseteq ad(F);$$

also $stb(F) = ss(F) = stg(F)$ in case $stb(F) \neq \emptyset$. Figure 2.2 gives an overview over the relations between all argumentation semantics considered in this work.

We furthermore note that semantics $\sigma \in \{na, pr, stb, stg, ss\}$ deliver *incomparable* sets, i.e. for all $E, D \in \sigma(F)$, $E \subseteq D$ implies $E = D$; the property is also referred to as *I-maximal*.

We introduce the *characteristic function* of a set of arguments.

Definition 2.1.8. *For an AF $F = (A, R)$ and a set of arguments $E \subseteq A$, we let $\Gamma_F(E) = \{a \in A \mid a_F^- \subseteq E_F^+\}$. If no ambiguity arises, we drop the subscript F.*

We note that complete extensions correspond to the fixed points of Γ_F; i.e., a set of arguments is complete iff $\Gamma_F(E) = E$. The grounded extension is unique for each AF F; it is the least fixed point of the characteristic function. Hence we can compute the grounded extension by iterative application of Γ_F, starting from the empty set. For each AF F, there is some $k \in \mathbb{N}$ such that $\Gamma_F^k(\emptyset) = G$ for the grounded extension $G \in gr(F)$.

We will make use of certain modifications of AFs. We consider the union of two AFs (also called *expansion*) and the *deletion* of an argument or a set of arguments.

Definition 2.1.9. *Given two AFs $F = (A, R)$ and $G = (A', R')$, we write $F \cup G = (A \cup A', R \cup R')$ to denote their component-wise union.*

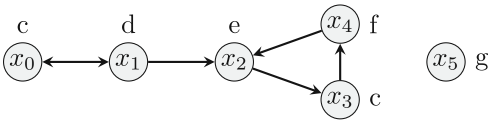


Figure 2.3: An example CAF $\mathcal{F} = (A, R, cl)$ with $cl(x_0) = cl(x_3) = c$, $cl(x_1) = d$, $cl(x_2) = e$, $cl(x_4) = f$, and $cl(x_5) = g$.

Definition 2.1.10. *For an AF $F = (A, R)$ and a set of arguments $A' \subseteq A$, we define the deletion of A' from F as $F \setminus A' = (A \setminus A', R \cap (A \setminus A')^2)$.*

2.2 Claim-augmented Argumentation

In this section, we extend AFs by adding claims to the abstract representation. We note that there are similar approaches that consider claims in the form of logical formulae in AFs [67, 68]. In this thesis, we stick to the approach considered in [92] in which claims are considered atomic. In the spirit of Dung's argumentation frameworks, we do not assume any particular structure or logic; our abstract claims could be a logical formula but also a statement given in natural language obtained from argument mining techniques.

Below, we define *claim-augmented argumentation frameworks (CAFs)* [92]. We fix a countable infinite domain C of claims.

Definition 2.2.1. *A* claim-augmented argumentation framework (CAF) *is a triple $\mathcal{F} = (F, cl) = (A, R, cl)$ where $F = (A, R)$ is an AF and $cl : A \to C$ is a function which assigns a claim to each argument in A. The claim-function is extended to sets in the following way: For a set $E \subseteq A$, $cl(E) = \{cl(a) \mid a \in E\}$.*

Let us point out a conceptual advantage of CAFs: with CAFs it is possible to capture situations in which two arguments represent the same conclusion, a scenario which cannot be formalized with standard argumentation frameworks without further assumptions.

Figure 2.3 presents an example of a CAF. Claims are depicted next to the arguments. The example extends our running example depicted in Figure 2.1 by assigning claims to each argument. Observe that claim c appears twice; it is the claim of the argument x_0 and of the argument x_3.

Definition 2.2.2. *Let $\mathcal{F} = (A, R, cl)$ be a CAF and let $c \in cl(A)$. We call an argument $a \in A$ with $cl(a) = c$ an* occurrence *of claim c (in $\mathcal{F}$).*

Definition 2.2.3. *Let $\mathcal{F} = (A, R, cl)$ be a CAF. For a set of claims $S \subseteq cl(A)$, we call a set of arguments $E \subseteq A$ with $cl(E) = S$ a* realization of S in $\mathcal{F}$*. If the realization E of S has property p (e.g., if E is admissible), we say that E is a p realization of S.*

Example 2.2.4. *Consider the CAF $\mathcal{F}$ from Figure 2.3. The claim-set $S = \{c, g\}$ has three realizations in $\mathcal{F}$, namely $E_1 = \{x_0, x_5\}$, $E_2 = \{x_3, x_5\}$, and $E_3 = \{x_0, x_3, x_5\}$. All of these realizations are conflict-free, however, only the realization E_1 is admissible.*

In [92], semantics of CAFs are defined based on the standard semantics of the underlying AF. The extensions are interpreted in terms of the claims of the arguments. We call this variant *inherited semantics* (i-semantics).

Definition 2.2.5. *For a CAF* $\mathcal{F} = (F, cl)$ *and a semantics* σ*, we define* inherited variant *of* σ *as* $\sigma_i(\mathcal{F}) = \{cl(E) \mid E \in \sigma(F)\}$*. We call a set* $E \in \sigma(F)$ *with* $cl(E) = S$ *a* σ_i*-realization of* S *in* $\mathcal{F}$*. We call a set* $S \in \sigma_i(\mathcal{F})$ *a* σ_i*-extension or a* σ_i*-claim-set. To emphasize that the extensions contain claims we also call them* claim-extensions.

Example 2.2.6. *Consider the CAF* $\mathcal{F}$ *from Figure 2.3. To evaluate* $\mathcal{F}$ *under complete semantics, we compute the complete extensions of the underlying AF and extract the claims in the next step. From Example 2.1.7, we have* $co((A, R)) = \{\{x_5\}, \{x_0, x_5\}, \{x_1, x_3, x_5\}\}$*. After applying the claim-function we obtain the claim-sets* $\{g\}$*,* $\{c, g\}$*, and* $\{d, c, g\}$*.*

The CAF $\mathcal{F}$ *accepts the following claim-sets under the considered semantics:*

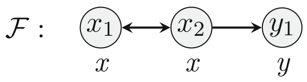
- $cf_i(\mathcal{F}) = \{\emptyset, \{c\}, \{d\}, \{e\}, \{f\}, \{g\}, \{c, e\}, \{c, g\}, \{c, f\}, \{f, g\}, \{d, c\}, \{d, f\}, \{d, g\}, \{e, g\}, \{c, e, g\}, \{c, f, g\}, \{d, c, g\}, \{d, f, g\}\}$
- $ad_i(F) = \{\emptyset, \{c\}, \{d\}, \{g\}, \{c, g\}, \{d, c\}, \{d, g\}, \{d, c, g\}\}$
- $co_i(F) = \{\{g\}, \{c, g\}, \{d, c, g\}\}$
- $gr_i(F) = \{\{g\}\}$
- $pr_i(F) = \{\{c, g\}, \{d, c, g\}\}$
- $stb_i(F) = ss_i(F) = stg_i(F) = \{\{d, c, g\}\}$
- $na_i(F) = \{\{c, e, g\}, \{c, g\}, \{c, f, g\}, \{d, c, g\}, \{d, f, g\}\}$

Basic relations between different semantics carry over from AFs, i.e. for any CAF $\mathcal{F}$,

$$stb_i(\mathcal{F}) \subseteq stg_i(\mathcal{F}) \subseteq na_i(\mathcal{F}) \subseteq cf_i(\mathcal{F}) \text{ and } stb_i(\mathcal{F}) \subseteq ss_i(\mathcal{F}) \subseteq pr_i(\mathcal{F}) \subseteq ad_i(\mathcal{F});$$

moreover, if $stb(\mathcal{F}) \neq \emptyset$ then $stb_i(\mathcal{F}) = ss_i(\mathcal{F}) = stg_i(\mathcal{F})$. Otherwise, we observe that we lose fundamental properties of semantics like I-maximality of preferred, naive, stable, semi-stable, and stage semantics:

Example 2.2.7. *Consider a CAF* $\mathcal{F}$ *given as follows:*



The underlying AF has two stable extensions: $\{x_2\}$ *and* $\{x_1, y_1\}$*. The resulting i-stable claim-sets are* $\{x\}$ *and* $\{x, y\}$*. Hence i-stable claim-sets are not necessarily I-maximal. Observe that* $na_i(\mathcal{F}) = stb_i(\mathcal{F}) = ss_i(\mathcal{F}) = stg_i(\mathcal{F}) = pr_i(\mathcal{F})$ *in this case, thus the same observation also holds for i-preferred, i-naive, i-stage, and i-semi-stable semantics.*

We will furthermore consider isomorphisms between CAFs. Graph-theoretically speaking, our CAF isomorphisms are arc- and labelling-preserving bijections.

Definition 2.2.8. *A bijective function* $f : A_F \rightarrow A_G$ *between two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *is an isomorphism if* f *is attack-preserving i.e.,* $(x, y) \in R_F$ *iff* $(f(x), f(y)) \in R_G$ *for all* $x, y \in A_F$*, and claim-preserving, i.e.,* $cl(x) = cl(f(y))$ *for all* $a \in A_F$*.* $\mathcal{F}$ *and* $\mathcal{G}$ *are isomorphic to each other (*$\mathcal{F} \cong \mathcal{G}$*) iff there is an isomorphism* $f : A_F \rightarrow A_G$*.*

We extend deletions of arguments in AFs to CAFs by appropriate restrictions of the claim-function.

Definition 2.2.9. *For a CAF* $\mathcal{F} = (A, R, cl)$ *and a set of arguments* $A' \subseteq A$*, we define the deletion of* A' *from* $\mathcal{F}$ *as* $F \setminus A' = (A \setminus A', R \cap (A \setminus A')^2, cl|_{A \setminus A'})$ *(where* $cl|_{A \setminus A'} : A \setminus A' \rightarrow cl(A)$*, as usual).*

Moreover, since we want to modify CAFs also by taking claims into account, we make use of the following definition and remove all arguments associated to a particular claim.

Definition 2.2.10. *For a CAF* $\mathcal{F} = (A, R, cl)$ *and a set of claims* $S \subseteq cl(A)$*, we define the deletion of claims* S *from the CAF* $\mathcal{F}$ *by* $\mathcal{F} \setminus S = (A \setminus A', R \cap (A \setminus A')^2, cl|_{A \setminus A'})$ *where* $A' = \{a \in A \mid cl(a) \in S\}$*.*

Well-formed CAFs. We consider a class of frameworks that appears in many different contexts: *well-formed CAFs* incorporate the basic observation that attacks typically depend on the claim of the attacking argument.

Definition 2.2.11. *A CAF* (A, R, cl) *is called* well-formed *if* $a^+_{(A,R)} = b^+_{(A,R)}$ *for all* $a, b \in A$ *with* $cl(a) = cl(b)$*.*

In well-formed CAFs we can speak of claims attacking arguments.

Definition 2.2.12. *Let* $\mathcal{F} = (A, R, cl)$ *be a well-formed CAF. We say that a claim* $c \in cl(A)$ attacks *an argument* $a \in A$ *if* $(x, a) \in R$ *for each argument* $x \in A$ *with claim* c*. Likewise, we say that* $S \subseteq cl(A)$ *attacks* $a \in A$ *if there is* $c \in S$ *that attacks* a*.*

2.3 Terminology and Notation

We give an overview over terminology, notation, and conventions used in this work.

Notation 2.3.1. *We use italic capital letters to denote AFs: typically, we use the letters* F*,* G*, and* H*. We use the corresponding calligraphic capital letters to denote CAFs:* $\mathcal{F}$*,* $\mathcal{G}$*, and* $\mathcal{H}$*. For a CAF* $\mathcal{F}$*, we write* $A_{\mathcal{F}}$*,* $R_{\mathcal{F}}$*,* $cl_{\mathcal{F}}$ *to indicate the affiliations; moreover, we write* F *to denote the AF corresponding to* $\mathcal{F}$ *(analogously for CAFs* $\mathcal{G}$ *and* $\mathcal{H}$*). If no ambiguity arises, we occasionally drop the subscript* $\mathcal{F}$ *(*$\mathcal{G}$*,* $\mathcal{H}$*, respectively).*

Notation 2.3.2. *We make use of the following abbreviations:* if and only if (iff), without loss of generality (w.l.o.g.), respectively (resp.), *and* with respect to (w.r.t.).

Statements, claims, and conclusions. All of these notions are closely related. There are, however, subtle differences. A *statement* can be any kind of clear or formal expression, not necessarily connected to an argument. A *claim* is is the distinct element of an argument whose merit must be established. We use the term *conclusion* in two ways. First, we call a claim a conclusion if it is the claim of an accepted argument (we speak, for instance, about the conclusions of an argumentation system and refer to the set of (cautiously or skeptically) accepted claims). Second, adapting terminology of structured argumentation in which arguments are defeasible proofs, we call a claim of an argument a conclusion if we put the focus on a particular argument and if we want to highlight that it is the result of a (logical or rule-based) defeasible proof. We note that each claim is a statement and each conclusion is a claim, but not vice versa.

Claim semantics. We use this term to denote the category of semantics that return sets of claims. In the context of CAFs, we use claim semantics synonymous for *CAF semantics*. If we speak of *claim semantics* we refer to *all* semantics for CAFs considered in this work.

Argument-level and claim-level. We sometimes think of argumentation in layers. The argument-level is the layer in which claims are considered secondary. We speak of the argument-level if we are interested on evaluation in terms of arguments or properties of argument-extensions. We speak of the claim-level if we go one step further and focus on the evaluation or on properties of claim-extensions.

Argument-focused vs. conclusion-focused evaluation methods. In the broad area of knowledge representation and reasoning, we speak of an *argument-focused evaluation method* if the emphasis lies on the (defeasible) proofs or justifications. By *conclusion-focused* or *claim-focused evaluation methods* we denote evaluation methods that put their main emphasis on the outcome (acceptable atoms, sentences, claims) and not on the proof that justifies the conclusion. In brief: argument-focused evaluation methods output arguments (proofs, evidence) while conclusion-focused evaluation-methods output claims (atoms, sentences). We also say *argument-focused* resp. *claim-focused reasoning*.

Argument-based vs. claim-based semantics. A semantics is *argument-based* if the acceptance status (of arguments or claims) is solely determined on argument-level. AF semantics, for instance, decide acceptance by looking only at the arguments. Also, inherited semantics for CAFs are argument-based: acceptance of claims is determined by deciding acceptance of the corresponding arguments. We speak of *claim-based* or *claim-sensitive semantics* if claims are taken into account in the evaluation of the acceptance status (of arguments or claims).

CHAPTER 3

Claims in Non-Monotonic Reasoning Formalisms

Non-monotonicity is a core aspect of argumentation. In classical logic, the addition of a new formula ψ to a set of formulae Γ does not downsize the set of derivable formulae: the derivability operator $\vdash$ is *monotonic*. In contrast to classical logic, the addition of new information in non-monotonic theories can lead to a smaller set of acceptable statements. Many non-monotonic theories are closely connected to argumentation [77, 62]. The connection is established via so-called *instantiation procedures*. These semantics-preserving translations take instances of some non-monotonic reasoning formalism and express them in terms of (abstract) arguments and attacks. Although there are exceptions, the vast majority of these procedures considers AFs as target formalism. In this chapter, we argue why it is beneficial to consider CAF instantiations instead. We will look into several knowledge representation theories and discuss the role of claims. Our main focus in this chapter lies on the semantics introduced by Dung: complete, grounded, preferred, and stable semantics. As we will see, there are several surprising obstacles regarding other semantics. We will discuss them throughout the sections.

We start our survey with a closer inspection of the sub-class of CAFs which we will repeatedly encounter in relation with instantiation procedures: *well-formed* CAFs model a natural behavior of attacks; they satisfy that arguments with the same claim attack the same arguments. We discuss several properties of this class. We will furthermore identify redundancies in CAFs, giving rise to a normal form that preserves Dung's semantics.

Our first non-monotonic formalism under consideration is *logic programming* [135] which is closely related to argumentation [77, 184, 61]. Next, we will discuss the role of claims in structured argumentation with main focus on *ASPIC+* [137] and *assumption-based argumentation (ABA)* [70]. We furthermore consider a generalization of AFs which allow for *collective attacks (SETAFs)* [141]. As we will see, well-formedness is an integral part of the construction of attacks in all of the aforementioned formalisms. It turns out that *all* CAFs generated from logic programs, ABA frameworks, and SETAFs are well-formed. We identify fragments of these formalisms which are in one-to-one correspondence (up to isomorphism) to well-formed CAFs. We furthermore take a look beyond well-formedness and discuss preferences in ASPIC+. We show that preference incorporation yield frameworks which violate well-formedness; in particular, we show that each CAF can be modeled by an induced sub-graph of an ASPIC+ instance.

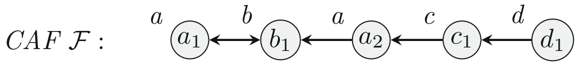
3.1 Well-formed CAFs and Normal Forms

In this section, we show that complete claim-extensions in well-formed CAFs admit a *unique realization* which implies that all semantics which return (a subset of) complete claim-extensions satisfy this property as well. Moreover, we establish the notion of *redundant arguments* which gives rise to an efficient procedure to reduce the number of arguments in a framework. Based on that notion, we define a *normal form* of CAFs.

Before we start to discuss unique realizations, let us recall that in well-formed CAFs, it makes sense to speak about *claims* attacking arguments since all arguments with the same claim have the same outgoing attacks.

In general, a claim-set in a (well-formed) CAF can have many realizations.

Example 3.1.1. *Consider the following well-formed CAF $\mathcal{F}$ and the claim-set $\{a, d\}$:*



The set $\{a, d\}$ has three realizations: $\{a_1, d_1\}$, $\{a_2, d_1\}$, and $\{a_1, a_2, d_1\}$. All realizations are admissible while only the latter one is complete: indeed, d_1 defends a_2 which in turn defends the argument a_1 thus $\{a_1, a_2, d_1\}$ is the unique complete realization of $\{a, d\}$.

Since each realization of a complete claim-set in a well-formed CAF attacks—and thus defends—the same arguments, it holds that each complete claim-set admits a unique co_i-realization. This property extends to all complete-based inherited semantics. We obtain an analogous result for i-naive semantics that extends to i-stage semantics.

Proposition 3.1.2. *Let $\mathcal{F}$ be a well-formed CAF and let $\sigma \in \{gr, co, pr, stb, ss, na, stg\}$. Each $S \in \sigma_i(\mathcal{F})$ admits a unique σ_i-realization in $\mathcal{F}$.*

Proof. First, we show that the union of two conflict-free realizations E, D ($E \neq D$) of the claim-set S is conflict-free. By well-formedness, we have that E and D attack the same arguments, i.e., $E^+ = D^+$. We show that $E \cup D$ is conflict-free: towards a contradiction, assume $E \cup D$ is conflicting. W.l.o.g., assume there is $x \in E$ that attacks $y \in D$. Then $y \in E^+ = D^+$ and thus D is not conflict-free, contradiction.

Naive (and stage) claim-extensions have a unique na_i-(stg_i-)realization since they are $\subseteq$-maximal conflict-free sets. For complete semantics, we furthermore observe that for every two co_i-realizations E and D of a complete claim-set S it holds that they defend each other (since they attack the same arguments). Since the union $E \cup D$ conflict-free, we obtain that S has a unique realization in $\mathcal{F}$. The result extends to all semantics that return complete sets, i.e., to grounded, preferred, stable, and semi-stable semantics. ☐

Let us discuss redundancies in CAFs. An argument is *redundant* if there is some other argument with the same claim that attacks the same arguments and has less attackers.

Definition 3.1.3. *Let $\mathcal{F}$ be a CAF. An argument $x \in A_\mathcal{F}$ is called* redundant *(in $\mathcal{F}$) w.r.t. argument $y \in A_\mathcal{F}$ if $x \neq y$, $cl(x) = cl(y)$, $x^+_\mathcal{F} = y^+_\mathcal{F}$, and $x^-_\mathcal{F} \supseteq y^-_\mathcal{F}$.*

We note that the definition of redundant arguments applies to general CAFs although it is inspired by the well-formedness property. We show next that such arguments can be removed without changing Dung's semantics. The following lemma will be useful.

Lemma 3.1.4. *Let $\mathcal{F}$ be a CAF, $x \in A_\mathcal{F}$ be redundant in $\mathcal{F}$ w.r.t. $y \in A_\mathcal{F}$, and let $\{x, y\}$ be conflicting in F. It holds that $(x, x), (y, x) \in R$.*

Proof. In case x is attacked by either x or y we obtain that both x and y attack x using $x^+ = y^+$. In case $(x, y) \in R$ we obtain $(x, x) \in R$ using $x \in y^- \subseteq x^-$. In case $(y, y) \in R$ we obtain $(x, y) \in R$ from $y \in y^- \subseteq x^-$. Hence $(x, x), (y, x) \in R$ in all cases. ☐

Proposition 3.1.5. *Let $\mathcal{F}$ be a CAF, $x \in A_\mathcal{F}$ be redundant in $\mathcal{F}$ w.r.t. $y \in A_\mathcal{F}$, and let $\mathcal{F}' = \mathcal{F} \setminus \{x\}$. Then, for $\sigma \in \{cf, ad, co, pr, stb\}$, it holds that $\sigma_i(\mathcal{F}) = \sigma_i(\mathcal{F}')$.*

We will prove this statement in the subsequent part of this section. For conflict-free and admissible semantics, the following observations will be useful.

Lemma 3.1.6. *Let $\mathcal{F}$ be a CAF, $x \in A_\mathcal{F}$ be redundant in $\mathcal{F}$ w.r.t. $y \in A_\mathcal{F}$, and $\mathcal{F}' = \mathcal{F} \setminus \{x\}$. For a set $E \subseteq A_\mathcal{F}$ with $x \in E$, let $E' = (E \setminus \{x\}) \cup \{y\}$. It holds that (a) $cl(E) = cl(E')$, $E^+_F = E'^+_F$, and $E^-_F \supseteq E'^-_F$. Moreover, if E is conflict-free, we have (b) $E'^-_{F'} = E'^-_F$ and (c) $E'^+_{F'} = E'^+_F$.*

Proof. (a) is by definition of E'. For (b), assume otherwise, then there is some $z \in E'$ such that $(x, z) \in R$; since $E \in cf(F)$, this implies that x attacks y, i.e., $z = y$ yielding $(x, x) \in R$ by Lemma 3.1.4. Regarding (c), assume $(z, x) \in R$ for some $z \in E'$; similar as above, this implies $(y, x) \in R$ and thus $(x, x) \in R$ by Lemma 3.1.4. ☐

Lemma 3.1.7. *$cf_i(\mathcal{F}) = cf_i(\mathcal{F}')$ for $\mathcal{F}, \mathcal{F}'$, $x \in A_\mathcal{F}$ defined as in Proposition 3.1.5.*

Proof. Let $S \in cf_i(\mathcal{F})$ and let E be a cf-realization of S in $\mathcal{F}$. If $x \notin E$, then E is a cf-realization of S in $\mathcal{F}'$ as well and thus $S \in cf_i(\mathcal{F}')$. In case $x \in E$, we let

$E' = (E \setminus \{x\}) \cup \{y\}$. By Lemma 3.1.6, we have $E_F^+ = E_F'^+ = E_{F'}'^+$ and $E_F^- \supseteq E_F'^- = E_{F'}'^-$, thus E' is conflict-free in F'. Since adding an argument (and attacks involving it) does not change the conflict-freeness of a set of arguments, we obtain $\mathit{cf}_i(\mathcal{F}) = \mathit{cf}_i(\mathcal{F}')$. $\square$

Lemma 3.1.8. *$ad_i(\mathcal{F}) = ad_i(\mathcal{F}')$ for $\mathcal{F}, \mathcal{F}'$, $x \in A_\mathcal{F}$ defined as in Proposition 3.1.5.*

Proof. Let $S \in ad_i(\mathcal{F})$ and let E denote an ad_i-realization of S in $\mathcal{F}$. First assume $x \notin E$. By Lemma 3.1.7, E is conflict-free in $\mathcal{F}'$, moreover, $E_{F'}^- = E_F^- \setminus \{x\} \subseteq E_F^+ \setminus \{x\} = E_{F'}^+$ (since $E_F^- \subseteq E_F^+$ by admissibility). In case $x \in E$, let $E' = (E \setminus \{x\}) \cup \{y\}$. By Lemma 3.1.7 E' is conflict-free in F', moreover, $E_{F'}'^- = E_F'^- \subseteq E_F^- \subseteq E_F^+ = E_F'^+ = E_{F'}'^+$ by Lemma 3.1.6. In both cases, E' defends itself in F'. To show the other direction, let $S \in ad_i(\mathcal{F}')$ and let E denote an ad_i-realization of S. By Lemma 3.1.7, E is conflict-free in F. Moreover, E defends itself in F: first, observe that E defends itself against all arguments $z \neq x$ since $E_F^+ = E_{F'}^+ \setminus \{x\}$. In case x attacks E in F we have y attacks E in F (since they have the same outgoing attacks) and thus E attacks x since it defends itself against the attack from y and since $x_F^- \supseteq y_F^-$. Hence E defends itself in F. $\square$

To show that $co_i(\mathcal{F}) = co_i(\mathcal{F}')$, we prove an even stronger result: we show that complete extensions of the underlying AFs are in one-to-one correspondence to each other.

Lemma 3.1.9. *Let $\mathcal{F}, \mathcal{F}'$, and $x \in A_\mathcal{F}$ be defined as in Proposition 3.1.5. Then $E \in co(F')$ iff $E \in co(F)$ or $E \cup \{x\} \in co(F)$ (and not both are contained in $co(F)$).*

Proof. First, let $E \in co(F')$. By Lemma 3.1.8, $E \in ad(F)$. First assume $E' = E \cup \{x\} \in ad(F)$. By Lemma 3.1.6, it holds that $E \cup \{y\} \in adm(F')$, i.e., $y \in E$ since $E \in co(F)$. Consider an argument $z \in A_\mathcal{F}$ such that $z_F^- \subseteq E_F'^+$ (i.e., E' defends z in F). In case $z = x$ we have E' defends x by assumption. In case $z \neq x$ we have $z_{F'}^- = z_F^- \setminus \{x\} \subseteq E_F'^+ \setminus \{x\} = E_F'^+ = E_{F'}^+$ thus E defends z in F'. Hence $z \in E$. Thus $E' \in co(F)$. In case $E \cup \{x\} \notin ad(F)$, it holds that $E \in co(F)$: Consider $z \in A_\mathcal{F}$ such that $z_F^- \subseteq E_F^+$. It holds that $z_{F'}^- = z_F^- \setminus \{x\} \subseteq E_F^+ \setminus \{x\} = E_{F'}^+$, thus $z \in E$. Hence $E \in co(F)$. For the other direction, consider a set $E \in co(F)$ and let $E' = E \setminus \{x\}$. Consider an argument $z \in A_\mathcal{F} \setminus \{x\}$ such that $z_{F'}^- \subseteq E_{F'}'^+$ (z is defended by E' in F'). Recall that $x \in E$ implies $y \in E$. Thus, by Lemma 3.1.6, $z_F^- \setminus \{x\} = z_{F'}^- \subseteq E_{F'}'^+ = E_F^+$, i.e., E defends z against all arguments in $A_\mathcal{F} \setminus \{x\}$ in F. Since $y_F^+ = x_F^+$ we obtain E defends z against x in case E defends z against y. Hence $z \in E \setminus \{x\}$ and thus $E' \in co(F')$. $\square$

It follows that there is a bijection between $co(F)$ and $co(F')$, mapping each set $E \in co(F')$ to E if E is complete in F or to $E \cup \{x\}$ otherwise. We obtain the following result.

Corollary 3.1.10. *$co_i(\mathcal{F}) = co_i(\mathcal{F}')$ for $\mathcal{F}, \mathcal{F}'$, $x \in A_\mathcal{F}$ defined as in Proposition 3.1.5.*

Lemma 3.1.11. *$pr_i(\mathcal{F}) = pr_i(\mathcal{F}')$ for $\mathcal{F}, \mathcal{F}'$, $x \in A_\mathcal{F}$ defined as in Proposition 3.1.5.*

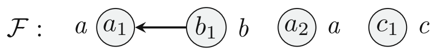
Proof. The proof is by Lemma 3.1.9 and since $x \in E$ implies $y \in E$ for each $E \in co(F)$. Thus, $D' \subset E'$ in F' implies $D \subset E$ in F for all $E', D' \in co(F')$ (here, D and E denote the complete extensions in F corresponding to D' and E'). Moreover, $D \subset E$ in F implies $D' \subset E'$ in F' (the last $\subset$-relation is indeed strict: otherwise, in case $D' = E'$ it holds that $x \in E \setminus D$, then $y \in E$ and $y \in D$, consequently $E^+_F = D^+_F$, thus both E and D defend x, contradiction). Hence we obtain $pr_i(\mathcal{F}) = pr_i(\mathcal{F}')$. ☐

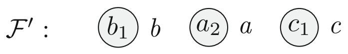
Lemma 3.1.12. *$stb_i(\mathcal{F}) = stb_i(\mathcal{F}')$ for $\mathcal{F}, \mathcal{F}'$, $x \in A_{\mathcal{F}}$ defined as in Proposition 3.1.5.*

Proof. Let $S \in stb_i(\mathcal{F})$, let E be a stb_i-realization of S in $\mathcal{F}$, and let $E' = E \setminus \{x\}$. Then $E^+_F = E'^+_{F'}$ and thus $E' \in stb(F')$. For the other direction, let $S \in stb_i(\mathcal{F}')$ and E a stb_i-realization of S in $\mathcal{F}'$. If $E \cup \{x\}$ is conflict-free we have $E \cup \{x\} \in stb(F)$. Otherwise, either (i) $(x, x) \in E$ or (ii) E attacks x or (iii) x attacks E. In case (i) we have $(y, x) \in R$ by Lemma 3.1.4. If $y \in E$ we have $E \in stb(F)$; otherwise $y \in E^+_F$ implies $x \in E^+_F$ thus $E \in stb(F)$. In case (ii), we have $E \in stb(F)$. In case (iii), we have y attacks E, thus E attacks y (since $E \in stb(F')$), thus E attacks x since $x^-_F \supseteq y^-_F$. Hence $E \in stb(F)$. ☐

We give counter-examples for the remaining semantics under consideration.

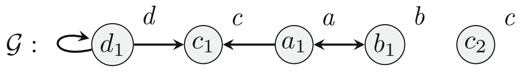
Example 3.1.13. *Consider the following well-formed CAF $\mathcal{F}$ and the CAF $\mathcal{F}' = \mathcal{F} \setminus \{a_1\}$:*

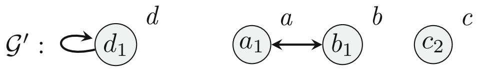




In $\mathcal{F}$, $\{a, c\}$ and $\{a, b, c\}$ are naive (since $\{a_1, a_2, c_1\}$ and $\{b_1, a_2, c_1\}$ naive in F). $\mathcal{F}'$, however, has a unique naive extension $\{a_2, b_1, c_1\}$ thus $\{a, c\}$ is not in $na_i(\mathcal{F})$ anymore.

For semi-stable and stage semantics, consider the well-formed CAF $\mathcal{G}$ and $\mathcal{G}' = \mathcal{G} \setminus \{c_1\}$:





In $\mathcal{G}$, c_1 is redundant w.r.t. c_2. The CAF $\mathcal{G}$ has a unique semi-stable (stage) claim-set: $\{a, c\}$, witnessed by $\{a_1, c_2\}$ with range $\{a_1, b_1, c_1, c_2\}$. $\mathcal{G}'$, however, has two semi-stable (and stage) claim-sets: $\{a, c\}$, realized by $\{a_1, c_2\}$, and $\{b, c\}$, realized by $\{b_1, c_2\}$.

We are ready to define a normal form for CAFs that preserve all Dung semantics.

Definition 3.1.14. *A CAF $\mathcal{F}$ is* in normal form *(also called* normalized*) iff there are no redundant arguments in $\mathcal{F}$.*

The following result is by repetitive application of Proposition 3.1.5.

Theorem 3.1.15. *Any CAF $\mathcal{F}$ can be transformed into an normalized CAF $\mathcal{F}'$, such that $\sigma_i(\mathcal{F}) = \sigma_i(\mathcal{F}')$ for $\sigma \in \{cf, ad, co, pr, stb\}$.*

We furthermore consider CAFs that contain no two arguments with the same claim, the same outgoing and incoming attacks. We call such arguments *copies* of each other.

Definition 3.1.16. *Let $\mathcal{F}$ be a CAF. An argument $x \in A_{\mathcal{F}}$ is a* copy *of $y \in A_{\mathcal{F}}$ iff $cl(x) = cl(y)$, $x^+_F = x^+_F$, and $x^-_F = x^-_F$. If $\mathcal{F}$ has no copies we call $\mathcal{F}$* copy-free.

Copies are a special case of redundant arguments, hence they can be removed without changing semantics (cf. Theorem 3.1.15).

3.2 Conclusions in Logic Programs

Logic programming belongs to the family of *declarative programming languages*. Its origin can be traced back to the early seventies, where the idea to consider programs as theories of a formal language was substantially developed [161, 128]. A logic program is a set of sentences in a logical language, typically written as clauses (Horn clauses if all sentences are atomic formulae). Semantics for logic programming often incorporate *negation-as-failure*: if some statement is not known to be true it is considered false. Gelfond and Lifschitz formalized *stable model semantics* [115] which incorporates this paradigm. Since both logic programming and abstract argumentation are considered two of the most influential non-monotonic reasoning paradigms, it is not surprising that their relation has been studied thoroughly in the literature: the first AF instantiation of a logic program has been already considered in Dung's seminal paper [77]; in subsequent works, further procedures relate several semantics of logic programs and argumentation (see, e.g., [184, 61, 114]). Interestingly, the role of claims in these procedures has often been neglected although they play a crucial role in constructing arguments and attacks between them. Moreover, atoms in the logic program correspond to claims in the argumentation framework, that is, they also play a crucial role in establishing the semantics correspondence.

In this section, we examine the close connection of logic programs and claim-focused argumentation. We discuss shortcomings of the classical instantiation and show how to extend the procedure to CAFs. As outlined above, the resulting framework will be well-formed, independently of the considered program. We furthermore present a translation from CAFs to logic programs and identify the fragments of both formalisms which are in one-to-one correspondence to each other.

Logic Programs in a Nutshell. We consider normal logic programs (LPs) [48] with default negation `not`. Such programs consist of rules r of the form

$$r:\ c \leftarrow a_1, \dots, a_n, \texttt{not}\ b_1, \dots, \texttt{not}\ b_m,$$

read as 'c if a_1 and ... and a_n and `not` b_1 and ... and `not` b_m'. Here, a_i, b_i and c are ordinary atoms; $\mathcal{L}(P)$ is the set of all atoms occurring in P. The atoms a_i are called *positive atoms*, denoted by $pos(r) = \{a_1, \dots, a_n\}$, and the atoms b_i are called *negative atoms*, we write $neg(r) = \{b_1, \dots, b_m\}$. We call $head(r) = c$ the *head* of r and $body(r) = \{a_1, \dots, a_n, \texttt{not}\ b_1, \dots, \texttt{not}\ b_m\}$ the *body* of the rule. The indices n, m are allowed to be equal to zero (that is, rules can have an empty body or only positive or negative atoms in the body). Given a set of atoms B, we write $\texttt{not}\ B = \{\texttt{not}\ b \mid b \in B\}$.

We introduce 3-valued model semantics [153] which generalize stable model semantics [115] by allowing for undefined atoms.

Definition 3.2.1. *A 3-valued Herbrand interpretation of an LP P is a tuple (T, F) with $T \cup F \subseteq \mathcal{L}(P)$ and $T \cap F = \emptyset$. An atom $a \in \mathcal{L}(P)$ is* true *iff $a \in T$,* false *iff $a \in F$, and* undefined *otherwise.*

Definition 3.2.2. *The* reduct *of an LP P with respect to a 3-valued Herbrand interpretation $I = (T, F)$ is the program P/I obtained by (i) removing each rule r from P with $T \cap neg(r) \neq \emptyset$, (ii) removing "$\texttt{not}\ b$" from each remaining rule whenever $b \in F$, and (iii) for each $a \notin T \cup F$, replacing each occurrence of "$\texttt{not}\ a$" by* **u**.

Given two 3-valued Herbrand interpretations $I = (T, F)$ and $I' = (T', F')$, we write $I \leq I'$ iff $T \subseteq T'$ and $F \supseteq F'$.

Definition 3.2.3. *A Herbrand interpretation $I = (T, F)$ is a* 3-valued model *of a program P iff I is a $\leq$-minimal model of P/I satisfying, for all atoms $a \in \mathcal{L}(P)$,*

(a) $a \in T$ iff there is a rule $r \in P/I$ with $a = head(r)$ and $pos(r) \subseteq T$, and

(b) $a \in F$ iff for each rule $r \in P/I$ with $a = head(r)$ we have $pos(r) \cap F \neq \emptyset$.

As P/I is a positive program, such a model exists and is unique [153]. We are ready to define the semantics of logic programs following [153, 61].

Definition 3.2.4. *For an LP P and a 3-valued interpretation $I = (T, F)$ of P, we call T*

- *partially stable (P-stable) if I is a 3-valued model of P;*
- *well-founded if I is a 3-valued model of P with $\subseteq$-minimal T;*
- *regular if I is a 3-valued model of P with $\subseteq$-maximal T;*
- *stable if I is a 3-valued model of P and $T \cup F = \mathcal{L}(P)$;*
- *L-stable if I is a 3-valued model of P and $T \cup F$ is $\subseteq$-maximal.*

Example 3.2.5. *Consider the following LP P containing rules*

$$r_0 : a \leftarrow \texttt{not}\ d. \qquad r_1 : d \leftarrow \texttt{not}\ a. \qquad r_2 : b \leftarrow d. \qquad r_3 : c \leftarrow d, \texttt{not}\ b.$$

Consider the interpretation $I = (\{a\}, \{b, c, d\})$. *We construct* P/I *and obtain the rules*

$$r'_0 : a \leftarrow . \qquad\qquad r'_2 : b \leftarrow d. \qquad\qquad r'_3 : c \leftarrow d.$$

It can be checked that I *is indeed a* $\leq$*-minimal model of* P/I. *We thus obtain that* I *is a 3-valued model of* P, *thus* $\{a\}$ *is partially stable, regular, and stable in* P. *Observe that* I *is not well-founded in* P *since* $(\emptyset, \mathcal{L}(P))$ *is a 3-valued model of* P *as well. Overall,* P *has three* P*-stable models:* $\emptyset$, $\{a\}$, *and* $\{d, b\}$*; the least one, i.e.,* $\emptyset$ *is well-founded while* $\{a\}$ *and* $\{d, b\}$ *are regular, stable and L-stable.*

We recall the translation from LPs into AFs following the procedure given in [61].

Translation 3.2.6. *For an LP* P, x *is an argument (in* P*) with*

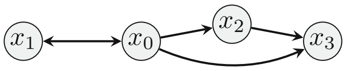
- $\textsc{Conc}(x) = c$,
- $\textsc{Rules}(x) = \bigcup_{i \leq n} \textsc{Rules}(x_i) \cup \{r\}$, *and*
- $\textsc{Vul}(x) = \bigcup_{i \leq n} \textsc{Vul}(x_i) \cup \{b_1, \dots, b_m\}$

iff there is a rule $r \in P$ *of the form* $r : \ c \leftarrow a_1, \dots, a_n, \mathtt{not}\ b_1, \dots, \mathtt{not}\ b_m$ *and arguments* $x_1, \dots, x_n$ *(in* P*) with* $\textsc{Conc}(x_i) = a_i$ *and* $r \notin \textsc{Rules}(x_i)$ *for all* $i \leq n$.

Given two arguments x *and* y, *we say* x *attacks* y *if* $\textsc{Conc}(x) \in \textsc{Vul}(y)$. *The corresponding AF is denoted by* $F_P = (A_P, R_P)$.

Example 3.2.7. *Consider again our program* P *from Example 3.2.5. We construct AF* F_P *using Translation 3.2.6: First, we obtain arguments* x_0, x_1 *for the rules* r_0 *and* r_1 *with* $\textsc{Conc}(x_0) = a$, $\textsc{Rules}(x_0) = \{r_0\}$, $\textsc{Vul}(x_0) = \{d\}$, *and* $\textsc{Conc}(x_1) = d$, $\textsc{Rules}(x_1) = \{r_1\}$, $\textsc{Vul}(x_1) = \{a\}$, *respectively. Using argument* x_1 *we can construct two further arguments* x_2 *and* x_3 *with* $\textsc{Conc}(x_2) = b$, $\textsc{Rules}(x_2) = \{r_1, r_2\}$, $\textsc{Vul}(x_2) = \{a\}$, *and* $\textsc{Conc}(x_3) = c$, $\textsc{Rules}(x_3) = \{r_1, r_3\}$, $\textsc{Vul}(x_3) = \{a, b\}$, *respectively. Attacks are based on the claims and vulnerabilities of the arguments, e.g.,* x_0 *attacks* x_1 *since* $\textsc{Conc}(x_0) = a \in \textsc{Vul}(x_1)$. *The resulting AF is depicted below:*

AF F_P :



The AF F_P *has three complete extensions:* $\emptyset$, $\{x_0\}$, *and* $\{x_1, x_2\}$. *The AF representation yields the models of the original logic program when extracting the atom in the rule head of the corresponding arguments. Indeed, we obtain the* P*-stable models* $\emptyset$, $\{a\}$, *and* $\{d, b\}$*; likewise, the stable extension corresponds to the stable model of* P. *In this example, L-stable semantics coincide with semi-stable semantics; we will see, however, that this is not always the case.*

As demonstrated by the above example, the transformation indeed preserves Dung semantics under an appropriate mapping. As shown in [61] this can be done by identifying functions *ConcLab2ArgLab* and *ArgLab2ConcLab* that map each P-stable (well-founded, regular, stable) 3-valued model to a complete (grounded, preferred, stable) argument-labelling and vice versa. An argument-labelling is a function *Lab* that assigns each argument a label in, out, or undec. Furthermore, let $\mathsf{in}(\mathit{Lab})$ denote the in-labelled arguments of an argument-labelling *Lab*. We obtain the following correspondence:

Proposition 3.2.8. *Let P be an LP and let $I = (T, F)$ be a 3-valued interpretation.*

- *T is P-stable iff $\mathsf{in}(\mathit{ConcLab2ArgLab}(I)) \in \mathit{co}(F_P)$;*
- *T well-founded iff $\mathsf{in}(\mathit{ConcLab2ArgLab}(I)) \in \mathit{gr}(F_P)$;*
- *T regular iff $\mathsf{in}(\mathit{ConcLab2ArgLab}(I)) \in \mathit{pr}(F_P)$;*
- *T stable iff $\mathsf{in}(\mathit{ConcLab2ArgLab}(I)) \in \mathit{stb}(F_P)$.*

The correspondence does not extend to L-stable semantics: Caminada et al. [61] show that L-stable semantics cannot be captured by classical AF semantics (under usual instantiation methods). There is, however, another weakness in the translation: establishing the correspondence between AFs and LPs requires an intermediate step (the re-interpretation of the arguments in terms of their claims) and is thus not directly given. We show how this detour can be circumvented by instantiating LPs directly as CAFs.

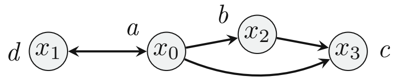
Representing LPs as CAFs. When defining the atom in the head of the respective rules to be the claims of the arguments, we obtain a CAF instantiation as follows:

Translation 3.2.9. *For an LP P, we obtain the associated CAF $\mathcal{F}_P = (F_P, cl_P)$ where $F_P = (A_P, R_P)$ is obtained as in Definition 3.2.6 and $cl_P(x) = \textsc{Conc}(x)$ for each $x \in A_P$.*

Translation 3.2.9 yields *well-formed* CAFs: when inspecting the attack construction in Translation 3.2.6, we see that arguments with the same claim attack the same arguments.

Example 3.2.10. *When assigning the corresponding claims to the arguments of the AF F_P associated to our LP P from Example 3.2.5 we obtain the following CAF $\mathcal{F}_P$:*

CAF $\mathcal{F}_P$:



As outlined above, the correspondence between LP and AF semantics is established via mappings that extract the claims. When instantiating an LP directly as CAF, this additional step can be avoided since the claims are part of the abstract representation.

Proposition 3.2.11. *Let P be a logic program and (T, F) be a 3-valued interpretation.*

- *T is P-stable iff $T \in co_i(\mathcal{F}_P)$;*
- *T is well-founded iff $T \in gr_i(\mathcal{F}_P)$;*
- *T is regular iff $T \in pr_i(\mathcal{F}_P)$;*
- *T is stable iff $T \in stb_i(\mathcal{F}_P)$.*

In line with observations in [61] we observe that the correspondence does not extend to L-stable semantics, as the following example demonstrates.

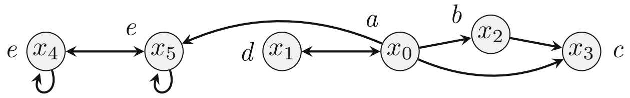
Example 3.2.12. *Consider our LP P from Example 3.2.5. We extend P with the rules*

$$r_4 : e \leftarrow \mathtt{not}\ e. \qquad\qquad r_5 : e \leftarrow \mathtt{not}\ a, \mathtt{not}\ e.$$

and let $P' = P \cup \{r_4, r_5\}$. The atom e stays undefined (it can be neither accepted nor rejected by design). Thus P' has no stable models. The P-stable, well-founded, and regular models of P and P' coincide; the sets $\{a\}$ and $\{d, b\}$ are both L-stable in P'.

Following Translation 3.2.9, we obtain the following CAF $\mathcal{F}_{P'}$ corresponding to P':

CAF $\mathcal{F}_{P'}$:



The CAF $\mathcal{F}_{P'}$ has only one semi-stable extension (and thus only one i-semi-stable claim-set): the argument x_0 attacks all but x_4 thus its range is maximal, yielding the unique semi-stable extension $\{x_0\}$; extracting the claim yields the claim-set $\{a\}$. Consequently, L-stable and semi-stable semantics do not necessarily yield the same outcome.

We have seen that *each* normal LP can be represented as CAF. Apart from L-stable semantics, the translation preserves the semantics; i.e., it is possible to express and evaluate each LP in terms of arguments and attacks without losing its original interpretation.

In the next part of the section, we dive deeper into the syntactical level of the translation. Which fragments of the LPs and CAFs, respectively, are in one-to-one correspondence to each other? What is the image of Translation 3.2.9 and what is its inverse? As a first observation, we recall that each CAF obtained from the translation is well-formed, that is, its image lies in the class of well-formed CAFs. Our next step to answer these questions is to identify a translation that maps a given well-formed CAF to an LP.

Representing CAFs as LPs. Starting by a well-formed CAF, we interpret each argument x as rule r with head $cl(x)$; moreover, each claim c that attacks x is a vulnerability of r, that is, '$\mathtt{not}\ c$' appears in the body of the rule.

Translation 3.2.13. *For a well-formed CAF $\mathcal{F} = (A, R, cl)$, we define the corresponding LP $P_{\mathcal{F}}$ by $P = \{c \leftarrow \mathtt{not}\ cl(x^-). \mid x \in A,\ cl(x) = c\}$.*

Arguments that are copies of each other yield the same rule: indeed, arguments $x, y \in A$ with $cl(x) = cl(y) = c$ and $x^- = y^-$ yield rule '$c \leftarrow \texttt{not}\ cl(x^-)$'. Moreover, no rule in the LP $P_{\mathcal{F}}$ obtained from Translation 3.2.13 has positive atoms. We call such rules *atomic*.

Definition 3.2.14. *A rule r of an LP P is called* atomic *if $pos(r) = \emptyset$. A program P is called* atomic *iff each rule in P is atomic.*

Next we show that atomic LPs and copy-free well-formed CAFs are in one-to-one correspondence to each other. The repetitive application of Translation 3.2.9 and 3.2.13 results in the same instance when starting from atomic LPs; likewise, when starting from copy-free well-formed CAFs, we obtain a CAF that is isomorphic to the original instance.

Proposition 3.2.15. *For each atomic LP P, it holds that $P = P_{\mathcal{F}_P}$. For each well-formed copy-free CAF $\mathcal{F}$, it holds that $\mathcal{F} \cong \mathcal{F}_{P_{\mathcal{F}}}$.*

Proof. Starting from an atomic program P, we obtain a well-formed CAF $\mathcal{F}_P$ in which each rule $r \in P$ yields an argument x with $cl(x) = head(r)$, $\textsc{Rules}(x) = \{r\}$, and $\textsc{Vul}(x) = \{b \mid \texttt{not}\ b \in body(r)\}$. Hence $x^- = \{y \mid \texttt{not}\ cl(y) \in body(r)\}$. Applying Translation 3.2.13 to argument x yields the rule $cl(x) \leftarrow cl(x^-)$. Since $cl(x^-) = \{b \mid \texttt{not}\ b \in body(r)\}$ we obtain that the rule coincides with r, showing that $P = P_{\mathcal{F}_P}$.

Given a copy-free well-formed CAF $\mathcal{F}$, we obtain a rule $cl(x) \leftarrow cl(x^-)$ for each argument $x \in A$ when applying Translation 3.2.13, which in turn yields an argument y with $cl(y) = head(r)$ which is attacked by $cl(x^-)$. Thus y has the same claim as x and the same outgoing and incoming attacks, showing that $\mathcal{F}$ and $\mathcal{F}_{P_{\mathcal{F}}}$ are indeed isomorphic. □

As a consequence we obtain that Translation 3.2.13 preserves the semantics: indeed, T is complete (grounded, preferred, stable) in a given well-formed CAF $\mathcal{F}$ iff T is complete (grounded, preferred, stable) in $\mathcal{F}'_{P_{\mathcal{F}'}}$ where $\mathcal{F}'$ denotes the CAF obtained by removing copies from $\mathcal{F}$ (removing copies does not change the semantics by Proposition 3.1.5; moreover, by Proposition 3.2.15, $\mathcal{F}'$ and $\mathcal{F}'_{P_{\mathcal{F}'}}$ are syntactically equivalent) iff T is P-stable (well-founded, regular, stable) in $P_{\mathcal{F}'}$ (by Proposition 3.2.11). As observed above, copies of arguments result in the same rule, thus we obtain $P_{\mathcal{F}'} = P_{\mathcal{F}}$.

Corollary 3.2.16. *Given a well-formed CAF $\mathcal{F}$ and the associated LP $P_{\mathcal{F}}$, it holds that T is complete (grounded, preferred, stable) in $\mathcal{F}$ iff there is a 3-valued interpretation $I = (T, F)$ of P such that T is P-stable (well-founded, regular, stable, respectively) in $P_{\mathcal{F}}$.*

We end this section with an interesting observation. Our results prove a result which is considered folklore: each logic program can be transformed into an atomic LP without changing the semantics. Given an LP P, we can compute the corresponding CAF and apply Translation 3.2.13 to obtain the atomic version of P. We note however that this correspondence does not extend to L-stable semantics since the intermediate step—switching to the corresponding CAF—is not semantics-preserving.

3.3 Claims in Structured Argumentation Formalisms

There is a prominent sub-area in argumentation where claims are naturally considered an integral part of each argument: in *structured argumentation*, which is concerned with the building blocks of arguments and their (supportive or opposing) relations, claims play an important role as distinctive part of each argument and are essential for attack determination. Over the past decades, a huge number of different approaches has been established; two influential directions are *rule-based* [137, 70] and *logic-based formalisms* [41]. Although there are several differences between these formalisms, they agree on the basic structure of arguments: Generally speaking, an argument consists of

- a *conclusion*, i.e., the *claim* of the argument, we can think of the claim, e.g., as logical formula ϕ or as atom or predicate p in some formal language; and
- a *support* from which the claim can be derived (assuming some deductive system). The support can be a set of formulae Γ which derives the formula ϕ (the claim); or a set of assumptions S together with a set of rules R such that S derives the predicate p (the claim) from rules in R.

Arguments contain *defeasible elements*; for instance, premises or defeasible rules. Researchers considered several different notions of conflict, e.g., undermining, undercutting, and rebutting attacks [41, 137]. They all are based on the same fundamental concept:

- an argument x attacks argument y iff the *claim* of x contradicts some *defeasible element* (usually part of the *support*) of y.

Attacks are often asymmetric which is due to the distinct role of the claims. We note that *all* formalisms confirming to this pattern satisfy well-formedness since arguments with the same claim attack the same arguments.

Structured and abstract argumentation formalisms are naturally closely related; while structured argumentation puts the emphasis on argument and attack construction, abstract formalisms allow for an alternative view on the argumentative setting by viewing arguments as abstract entities. In this way, both fields complement each other as they provide alternative representation and evaluation techniques of an argumentative setting.

In this section, we focus on two prominent examples of structured argumentation: *ASPIC with preferences (ASPIC+)* [137] and *assumption-based argumentation (ABA)* [70], both regarded as influential representatives of rule-based formalisms. One of the main differences between the formalisms is that ASPIC+ builds upon semantics of abstract argumentation (an AF instantiation is thus integral part of the formalism) while ABA has a native semantics based on assumption-sets (which are closely related to AF semantics).

We start our overview by reviewing a fragment of ASPIC+. We present an instantiation of ASPIC+ frameworks into CAFs that generalizes the AF instantiation. Although the attack construction satisfies well-formedness, it turns out that the incorporation of preferences yields non-well-formed CAFs. We show that each CAF can be modeled by an induced sub-graph of a framework generated by an ASPIC+ knowledge base.

The close correspondence of ABA and AF semantics is well-known [79, 70]; in this section, we discuss their semantics correspondence in terms of claims. Viewing instantiations in the tradition of translations between formalisms, we moreover discuss constructions to model (well-formed) CAFs in ABA. To do so, we focus on the *conclusion-extensions* of the ABA framework. We show that a particular class of ABA frameworks is in one-to-one correspondence with the class of (copy-free) well-formed CAFs.

3.3.1 Abstract Rule-Based Argumentation

ASPIC+ is a flexible framework with various extensions which exists in many different formulations [57, 151, 137]. In contrast to other structured argumentation formalisms, ASPIC+ puts the focus entirely on argument and attack construction. ASPIC instances are evaluated via instantiating an AF. The identification of acceptable claims is an important part of the evaluation; with CAFs, this evaluation can be performed already on abstract level which streamlines the process. It is important to mention that the conclusion-focused evaluation is not meant to replace the argument-based evaluation of the framework (naturally, the arguments are still available in the abstract representation); instead, our approach aims to enrich the traditional evaluation methods.

We note that we do not introduce ASPIC+ in its full generality; our goal is to demonstrate how to extend the typical instantiation procedure of ASPIC+ into AFs to CAFs and survey its advantages; moreover, we show that the resulting CAF is not necessarily well-formed when considering preferences as part of the theory.

Definition 3.3.1. *An* argumentation system *is a triple* $AS = (\mathcal{L}, \mathcal{R}, n)$ *where* $\mathcal{L}$ *is a logical language with unary negation symbol* $\neg$ *(we write* $\varphi = -\psi$ *in case* $\psi = \neg\varphi$ *or* $\varphi = \neg\psi$*);* $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *is a set of strict (*$\mathcal{R}_s$*) and defeasible (*$\mathcal{R}_d$*) rules of the form* $\varphi_1, \dots, \varphi_n \rightarrow \varphi_{n+1}$*, and* $\varphi_1, \dots, \varphi_n \Rightarrow \varphi_{n+1}$*, resp. (*$\varphi_i$ *are meta-variables ranging over wff in* $\mathcal{L}$*), and* $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$*; and* $n : \mathcal{R}_d \rightarrow \mathcal{L}$ *is a partial function. A* knowledge base *in* AS *is a set* $\mathcal{K} = \mathcal{K}_p \cup \mathcal{K}_n \subseteq \mathcal{L}$ *(with* $\mathcal{K}_p \cap \mathcal{K}_n = \emptyset$*) of premises (*$\mathcal{K}_p$*) and facts (*$\mathcal{K}_n$*). The tuple* $(AS, \mathcal{K})$ *is called* argumentation theory *(AT).*

The function $n : \mathcal{R}_d \rightarrow \mathcal{L}$ assigns defeasible rules logical formulae; building the necessary foundation to allow for attacks on rules (also called undercuts). In what follows, we identify $\mathcal{K}_p$ with rules of the form $\Rightarrow \varphi$.

Definition 3.3.2. *Given an AT* $\mathcal{A}$*,* A *is an argument (in* $\mathcal{A}$*) with*

- $\textsc{Conc}(A) = \varphi$*,*
- $\textsc{DefRules}(A) = \bigcup_{i \leq n} \textsc{DefRules}(A_i) \cup \{r\}$ *if* r *is defeasible (*$\textsc{DefRules}(A) = \bigcup_{i \leq n} \textsc{DefRules}(A_i)$ *otherwise),*

- $\textsc{TopRule}(A) = r$, *and*
- $\textsc{Sub}(A) = \bigcup_{i\leq n} \textsc{Sub}(A_i) \cup \{A_1, \dots, A_n\}$,

iff there is a rule $r \in \mathcal{R}$ *of the form* $r = \textsc{Conc}(A_1), \dots, \textsc{Conc}(A_n) \rhd \varphi$, $\rhd \in \{\rightarrow, \Rightarrow\}$, *and arguments* $A_1, \dots, A_n$ *(in* $\mathcal{A}$*).*

We deviate from the usual definition and omit $\textsc{Prem}(A)$ (i.e., the set of premises in A) because we identify premises with defeasible rules having empty body. This allows us to combine rebuts (attacks on conclusions of sub-arguments) and undermining attacks (attacks on premises) as follows:

Definition 3.3.3. *Argument* A *attacks argument* B *on* B' *(in an AT* $\mathcal{A}$*) iff* $\textsc{TopRule}(B')$ *is defeasible and* $\textsc{Conc}(A) = -\textsc{Conc}(B')$*.*

Attacks on rules, so-called *undercuts*, are defined as follows:

Definition 3.3.4. *Argument* A *undercuts* B *on* $r \in \textsc{DefRules}(B)$ *iff* $\textsc{Conc}(A) = -n(r)$*.*

Notice that attacks on sub-arguments as well as undercuts satisfy well-formedness; in both definitions, only the claim of the attacking argument is of importance. An instantiation with the attack notions that we have defined so far results in a well-formed CAF.

ASPIC+ allows for argument orderings specifying preferences between arguments. While undercutting attacks are independently successful of the preference ordering, the success of undermining and rebutting attacks crucially depends on the argument ordering.

Definition 3.3.5. *An argument* A successfully attacks *argument* B *iff* A *undercuts* B *or* A *attacks* B *on* B' *and* $B' \not\succ A$*. Successful attacks are also called defeats.*

The corresponding AF is obtained by constructing all arguments and successful attacks; we extend this instantiation to CAFs.

Translation 3.3.6. *Given an AT* $\mathcal{A}$*, then the AF* $F_\mathcal{A}$ *is obtained by constructing all arguments* A *and successful attacks* R *between arguments in* $\mathcal{A}$*. We define the associated CAF* $\mathcal{F}_\mathcal{A} = (F_\mathcal{A}, cl)$ *with* $cl(A) = \textsc{Conc}(A)$ *for each argument* A*.*

The resulting CAF has potentially infinitely many arguments. However, when focusing on ATs over a finite language, it suffices to construct only finitely many arguments. This result is often considered folklore and has been explicitly stated for other forms of structured argumentation [8, 159].

We evaluate an AT $\mathcal{A}$ by constructing AF $F_\mathcal{A}$ and apply the desired AF semantics. The instantiation of $\mathcal{A}$ as CAF simplifies the conclusion-focused evaluation of the knowledge base where the focus lies on justified formulae and acceptable conclusions (cf. [137, Definition 2.18], and below). Moreover, the CAF representation allows for a structural analysis of $\mathcal{A}$ that takes the claims into account.

As we will see, the resulting CAF is not necessary well-formed. We show that for each CAF $\mathcal{F}$, there is an argumentation system $\mathcal{A}$ such that $\mathcal{F}_\mathcal{A}$ has an induced subgraph G such that $\mathcal{F} = G$. That is, when focusing on a subset of arguments in an argumentation system, it not possible to find any structural constraints that are valid in general.

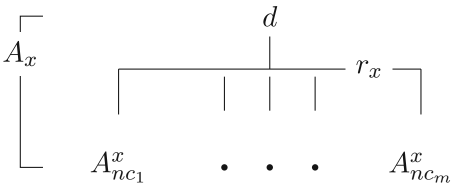
Proposition 3.3.7. *For each CAF $\mathcal{F}$, there is an AT $\mathcal{A}$ such that the associated CAF $\mathcal{F}_\mathcal{A}$ has an induced subgraph G satisfying that $\mathcal{F} = G$.*

Proof. First, we present an abstract construction that defines arguments and their sub-arguments and preference orderings between them. Second, we show how such a situation can be realized using the *last-link principle* [137].

Let $\mathcal{F} = (A, R, cl)$. For each argument $x \in A$, we introduce an ASPIC+ argument A_x with conclusion $\textsc{Conc}(A_x) = cl(x)$ and sub-arguments $\{A^x_{nc} \mid c \in cl(x^-)\}$ where $\textsc{Conc}(A^x_{nc}) = \neg c$, i.e., for each claim c that attacks x in $\mathcal{F}$, the argument A_x has one sub-argument with conclusion $\neg c$. For each such argument x we add rules $r^x_{nc} :\Rightarrow \textsc{Conc}(A^x_{nc})$ for all sub-arguments A^x_{nc} of A_x and a rule $r_x : \{\textsc{Conc}(A^x_{nc}) \mid c \in cl(x^-)\} \Rightarrow \textsc{Conc}(A_x)$ which is the toprule of A_x.

Below we depict the construction, given $x \in A$ with $cl(x) = d$ and $cl(x^-) = \{c_1, \dots, c_m\}$:

Argument A_x:



Now, without preferences, each argument with claim $c \in cl(x^-)$ attacks argument A_x on the respective sub-argument A^x_{nc} with conclusion $\neg c$.. In order to avoid unwanted attacks we define a preference ordering as follows: for all $y \in A$ with claim $c = cl(y) \in cl(x^-)$ but $(y, x) \notin R$, we let $A^x_{nc} \succ A_y$. The attack from A_y to A_x onto sub-argument A^x_{nc} is thus not successful. Restricting the CAF $\mathcal{F}_\mathcal{A}$ corresponding to $\mathcal{A}$ to the arguments A_x, $x \in A$, we obtain an induced subgraph G such that $\mathcal{F} = G$, as desired. We note that the ordering is a strict partial order, i.e., it is irreflexive, asymmetric, and transitive.

To give a construction using the last-link principle (formalizing that the last link, i.e., the ordering of the top-rules of arguments decides their relative strength), we define a preference-ranking on our rules instead of directly ranking the arguments: for all $y \in A$ with claim $c = cl(y) \in cl(x^-)$ but $(y, x) \notin R$, we let $r^x_{nc} \succ r_y$. Lifting the preference ordering to argument-level, we obtain the desired ranking between arguments, i.e., the argument A^x_{nc} is stronger than A_y if y has claim c but does not attack x in $\mathcal{F}$, yielding the induced subgraph G of $\mathcal{F}_\mathcal{A}$ that coincides with $\mathcal{F}$. $\square$

3.3.2 Assumption-based Argumentation

We assume a *deductive system* $(\mathcal{L}, \mathcal{R})$, where $\mathcal{L}$ is a formal language, i.e., a set of sentences, and $\mathcal{R}$ is a set of inference rules over $\mathcal{L}$. A rule $r \in \mathcal{R}$ has the form $a_0 \leftarrow a_1, \ldots, a_n$ with $a_i \in \mathcal{L}$ for all $i \leq n$. We denote the head of r by $head(r) = a_0$ and the body of r with $body(r) = \{a_1, \ldots, a_n\}$. If $n = 0$, i.e., if r has an empty body, we call r a *fact* and write $a_0 \leftarrow \top$ with $\top \notin \mathcal{L}$ or simply $a \leftarrow$ interchangeably.

Definition 3.3.8. *An ABA framework is a tuple* $(\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$*, where* $(\mathcal{L}, \mathcal{R})$ *is a deductive system,* $\mathcal{A} \subseteq \mathcal{L}$ *a set of assumptions, and* ${}^{-} : \mathcal{A} \rightarrow \mathcal{L}$ *is the contrary function mapping assumptions* $a \in \mathcal{A}$ *to sentences* $\mathcal{L}$*.*

We extend the contrary function to sets of assumptions S via $\overline{S} = \{\overline{a} \mid a \in S\}$.

Definition 3.3.9. *A deduction or* tree-derivation $S \vdash_R p$ *for a sentence* $p \in \mathcal{L}$ *supported by a set of assumptions* $S \subseteq \mathcal{A}$ *and a set of rules* $R \subseteq \mathcal{R}$ *is a finite rooted labeled tree such that the root is labeled with* p*, the set of labels for the leaves of the tree is equal to* S *or* $S \cup \{\top\}$*, and there is a surjective mapping from the set of internal nodes to* R *satisfying for each internal node* v *there is a rule* $r \in R$ *such that* v *labeled with* $head(r)$ *and the set of all successor nodes corresponds to* $body(r)$ *or* $\top$ *if* $body(r) = \emptyset$*.*

Assumption-based argumentation exists in many different variations. Below, we settle in which framework we are interested in.

Assumption 3.3.10. *In this work, we focus on* finite *ABA frameworks, i.e., we assume that the sets* $\mathcal{L}$*,* $\mathcal{R}$*, and* $\mathcal{A}$ *are finite. When studying computational aspects, we furthermore assume that each rule* $r \in \mathcal{R}$ *is stated explicitly, i.e., given as input. Also, we restrict our language* $\mathcal{L}$ *to* atomic formulae*, and assume that* $\mathcal{L}$ *is the union of all atoms that appear in* $\mathcal{R}$*,* $\mathcal{A}$*, and* $\overline{\mathcal{A}}$*. Moreover, we focus on* flat *ABA, assuming that each rule* r *in a given ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$ *satisfies* $head(r) \notin \mathcal{A}$*.*

Originally, flat frameworks have been defined by requiring that each set of assumptions is closed under deduction, i.e., $S \vdash_R a$ for all $S \subseteq \mathcal{A}$, $a \in S$. Apart from the computational advantages the alternative definition of flat frameworks is only syntactically stronger (by allowing rules of the form $a \leftarrow a$ for assumptions $a \in \mathcal{A}$ as well as rules r with $head(r) \in \mathcal{A}$ which contain body elements that are not derivable in D).

Arguments in ABA are based on tree-derivations. The literature considers different variants by taking, e.g., set of rules or the tree-derivation into account. In this work, we follow the definition proposed in, e.g., [80], and identify tree-derivations with each other if they agree on the set of assumptions and conclusion of the derivation.

Definition 3.3.11. *For an ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$*, an sentence* $p \in \mathcal{L}$*, and a set of assumptions* $S \subseteq \mathcal{A}$*, we write* $S \vdash p$ *iff there exists a set of rules* $R \subseteq \mathcal{R}$ *such that* $S \vdash_R p$ *in* D *and call* $S \vdash p$ *an* argument *with conclusion (or* claim*)* p *in* D*.*

Each assumption $a \in \mathcal{A}$ derives itself via $\{a\} \vdash_{\emptyset} a$ and thus induces an argument $\{a\} \vdash a$. We call such arguments *assumption-arguments* to distinguish them from the so-called *proper arguments* which correspond to derivation trees with strictly more than one node.

For an argument $x = S \vdash p$, we consider functions $cl(x) = p$ returning its conclusion resp. the assumptions. We extend these functions to sets of arguments: $cl(E) = \{cl(x) \mid x \in E\}$ and $asms(E) = \bigcup_{x \in E} asms(x)$ for a set of arguments E.

Definition 3.3.12. *For an ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ and a set of assumptions S, we let $Th_D(S) = \{p \in \mathcal{L} \mid \exists S' \subseteq S : S \vdash p\}$ be the set of all claims derivable by S in D.*

Each set of assumptions S is contained in $Th_D(S)$ since each assumption derives itself. We call $Th_D(S) \setminus S$ the set of *proper conclusions* of S.

The contrary relation determines *conflicts* between sets of assumptions. Using our notion of arguments, we can define attacks between sets of assumptions as follows.

Definition 3.3.13. *Let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ be an ABA framework. A set of assumptions $S \subseteq \mathcal{A}$* attacks *an assumption $a \in \mathcal{A}$ in D iff there is $S' \subseteq S$ such that $S' \vdash \overline{a}$. S attacks a set of assumptions $T \subseteq \mathcal{A}$ iff S attacks some $a \in T$.*

By $S_D^+ = \{a \in \mathcal{A} \mid S$ attacks a in $D\}$ we denote the set of all assumptions that are attacked by S in D. The set S is called *conflict-free* in D iff S does not attack itself; we say S *defends itself* in D iff it counter-attacks each set $T \subseteq \mathcal{A}$ which attacks S in D.

We next recall admissible, grounded, complete, preferred, stable, and semi-stable semantics for ABA (abbreviated *ad*, *gr*, *co*, *pr*, *stb*, and *ss*).

Definition 3.3.14. *Let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ be an ABA framework. Further, let $S \subseteq \mathcal{A}$ be a conflict-free set in D which is closed under deduction*[1].

- *$S \in ad(D)$ iff S defends itself against each attack;*
- *$S \in co(D)$ iff S is admissible and contains every assumption set it defends;*
- *$S \in gr(D)$ iff S is $\subseteq$-minimal in $co(D)$;*
- *$S \in pr(D)$ iff S is $\subseteq$-maximal in $ad(D)$;*
- *$S \in stb(D)$ iff S attacks each $\{x\} \subseteq \mathcal{A} \setminus S$;*
- *$S \in ss(D)$ iff S is admissible and $S \cup S^+$ is $\subseteq$-maximal in $ad(D)$.*

For a semantics σ, we call $\sigma(D)$ the σ-assumption-extensions*; $\sigma_{Th}(D) = \{Th_D(S) \mid S \in \sigma(D)\}$ the* σ-conclusion-extensions*; and $\sigma_{pTh}(D) = \{S \setminus \mathcal{A} \mid S \in \sigma_{Th}(D)\}$ the* proper σ-conclusion-extensions *of D.*

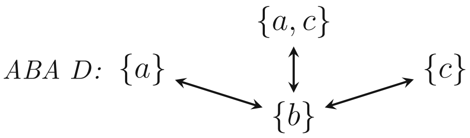
---

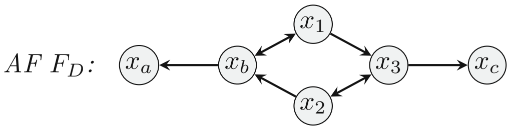
[1] We note that for flat ABA frameworks, the closure criteria can be omitted since each set of assumptions S is closed under deduction if D is flat.

Assumption-based argumentation and abstract argumentation are closely related. For a given ABA framework D, a corresponding AF F_D can be constructed by identifying all arguments following Definition 3.3.11 and identify conflicts between them.

Translation 3.3.15. *The associated AF* $F_D = (A, R)$ *of an ABA* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$ *is given by* $A = \{S \vdash p \mid \exists R \subseteq \mathcal{R} : S \vdash_R p\}$ *and attacks* $(S \vdash p, S' \vdash p') \in R$ *iff* $p \in \overline{S'}$*.*

Example 3.3.16. *Consider an ABA framework* D *with assumptions* $\mathcal{A} = \{a, b, c\}$ *with* $\overline{a} = b$*,* $\overline{b} = p$*, and* $\overline{c} = q$*, and rules* $r_1 : p \leftarrow a$*,* $r_2 : p \leftarrow c$*, and* $r_3 : q \leftarrow b$*. Below we depict the attacks between the assumption-sets (left, we omit* $\emptyset$ *as it is not attacked as well as* $\{a, b\}$*,* $\{b, c\}$*, and* $\mathcal{A}$ *as they are in conflict with all sets) and the AF* F_D *(right) with proper arguments* x_i *(induced by rules* r_i*) and assumption-arguments* x_a*,* x_b*, and* x_c*.*





D has two stable assumption-sets: $S_1 = \{b\}$ *and* $S_2 = \{a, c\}$ *with* $Th_D(S_1) = \{b, q\}$ *and* $Th_D(S_2) = \{a, c, p\}$*. The stable extensions in* F_D *are* $\{x_3, x_b\}$ *and* $\{x_1, x_2, x_a, x_c\}$*.*

ABA semantics and AF semantics are closely related (see. e.g., [70, Theorem 4.3]).

Proposition 3.3.17. *Given an ABA* D *and a semantics* $\sigma \in \{ad, gr, co, pr, stb\}$*. If* $E \in \sigma(F)$ *then* $asms(E) \in \sigma(D)$*; if* $S \in \sigma(D)$ *then* $\{S' \vdash p \mid \exists S' \subseteq S : S' \vdash p\} \in \sigma(F)$*.*

We note that finite ABA frameworks might instantiate AFs with an infinite number of arguments. However, it suffices to consider only finite arguments in the instantiation to preserve the semantics [159].

Semi-stable semantics cannot be captured via standard AF instantiations [62]. While complete, grounded, preferred, and stable ABA and AF semantics are in one-to-one correspondence, admissible semantics potentially yields several argument-extensions that correspond to a single assumption-extension (the empty assumption-extension, for instance, corresponds to several argument-extensions if the framework contains facts).

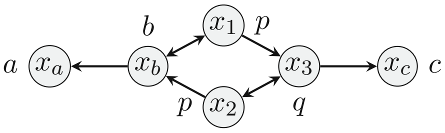
Representing ABA via CAFs. There is a natural adaption of the AF instantiation given in Translation 3.3.15 to CAFs by assigning each argument $S \vdash p$ its claim p.

Translation 3.3.18. *The associated CAF* $\mathcal{F}_D = (F_D, cl)$ *for an ABA* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$ *is obtained by constructing* F_D *from Translation 3.3.15 and assign each argument* $S \vdash p$ *its claim* $cl(S \vdash p) = p$*.*

By definition of the attack relation, each (flat or non-flat) ABA framework yields a well-formed CAF since the attacks depend on the claim of the attacking argument. Indeed, an argument x attacks argument y if $cl(x) = \overline{a}$ for some $a \in asms(y)$.

Example 3.3.19. *Consider the ABA framework D from Example 3.3.16. The CAF instantiation equips the AF F_D with the claims of the arguments (depicted next to the arguments). Each assumption $a \in \mathcal{A}$ appears as conclusion in $\mathcal{F}_D$ attached to its unique argument x_a obtained from the derivation $\{a\} \vdash_\emptyset a$. Observe that $\mathcal{F}_D$ is well-formed.*

CAF $\mathcal{F}_D$:

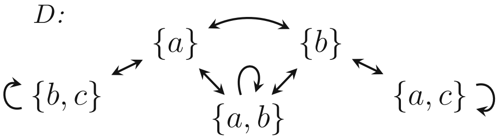


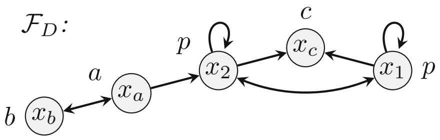
From the one-to-one correspondence for grounded, complete, preferred, and stable semantics in Proposition 3.3.17 and since $Th_D(S) = cl(S)$ for each σ-assumption-extension $S \in \sigma(D)$ we obtain that the translation preserves σ-conclusion-extensions of an ABA framework D; assumption-extensions can be obtained by projecting the claim-sets to $\mathcal{A}$.

Proposition 3.3.20. *For an ABA $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$, its associated CAF $\mathcal{F}_D$ and $\sigma \in \{gr, co, pr, stb\}$, it holds that $\sigma_{Th}(D) = \sigma_i(\mathcal{F}_D)$ and $\sigma(D) = \{C \cap \mathcal{A} \mid C \in \sigma_i(\mathcal{F}_D)\}$.*

Similar as for AFs, admissible ABA and AF semantics are not in one-to-one correspondence to each other. Moreover, the correspondence does not extend to semi-stable semantics.

Example 3.3.21. *Consider the ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ with $\mathcal{A} = \{a, b, c\}$ with $\overline{a} = b$, $\overline{b} = a$, $\overline{e} = p$, and rules $r_1 : p \leftarrow e$ and $r_2 : p \leftarrow e, b$. We depict attacks between assumption-sets and the associated CAF $\mathcal{F}_D$ below:*





D has no stable extensions; the $\{a\}$ and $\{b\}$ are both semi-stable. In the corresponding CAF $\mathcal{F}_D$, on the other hand, only $\{a\}$ is semi-stable since in the underlying AF, the argument x_a attacks both x_b and x_2 while x_b attacks only x_a.

In what follows, we will thus focus on complete, grounded, preferred, and stable semantics.

Representing ABA via CAFs: proper conclusions. Naturally, our main focus in this work lies on the *claims* and the conclusion-extensions of the constructed arguments in an ABA framework. In what follows, we will thus take a closer look at proper conclusion-extensions of ABA frameworks. We present a more flexible approach to instantiate a given ABA framework that considers proper arguments only. This means that we will modify Translation 3.3.18 by removing all assumption-arguments. We will show that this modification has no impact on the proper conclusion-extensions if the ABA framework has *separated contraries*, that is, no assumption is the contrary of another assumption. We will furthermore show that each ABA framework can be transformed into a framework with separated contraries. As we will see, the proposed conclusion-focused instantiation will help to establish an even closer connection between CAFs and ABA frameworks: as done for LPs, we will consider a translation that transforms a given well-formed CAF into an ABA framework while preserving Dung semantics.

Let us start by formally stating our flexible instantiation for ABA frameworks.

Translation 3.3.22. *For an ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$*, we define the associated CAF* $\mathcal{F}_{D\text{-}\mathcal{A}} = (A', R', cl)$ *by constructing* (A, R', cl) *via Translation 3.3.18 and restrict the arguments to non-assumptions, i.e.,* $A' = A \setminus \{S \vdash p \mid p \in \mathcal{A}\}$ *and* $R' = R \cap A' \times A'$*.*

In general, the above translation does not preserve the semantics. Consider for instance the ABA framework D from Example 3.3.19. The removal of the assumption-arguments x_a, x_b and x_c results in a change of the semantics: the argument x_3 is now attacked by the unattacked argument x_1 and thus cannot be accepted with respect to any admissible-based semantics. The reason is that we removed the argument associated to the contrary of a, i.e., the assumption-argument x_b, which plays a crucial role for defending x_3.

On the other hand, if an ABA framework has no such assumptions whose contrary is an assumption, then the translation indeed preserves the semantics, as we will see.

Definition 3.3.23. *An ABA* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ *has* separated contraries *iff* $\mathcal{A} \cap \overline{\mathcal{A}} = \emptyset$*.*

The following lemma is crucial: we show that the removal of an argument a with no outgoing attacks in an AF F corresponds to the removal of a from each extension E of F.

Lemma 3.3.24. *For an AF* $F = (A, R)$*, an argument* $a \in A$ *with* $a^+ = \emptyset$*, a set of arguments* $E \subseteq A \setminus \{a\}$*, and a semantics* $\sigma \in \{cf, ad, gr, co, pr, stb\}$*, it holds that* $E \in \sigma(F \setminus \{a\})$ *iff* $E \in \sigma(F)$ *or* $E \cup \{a\} \in \sigma(F)$*.*

Proof. Let us start with conflict-free sets. Consider a set E not containing a. First, $E \cup \{a\} \in cf(F)$ implies $E \in cf(F)$; moreover, E is conflict-free in $F \setminus \{a\}$ iff E is conflict-free in F. If E is admissible, then it attacks the same arguments $b \in A \setminus \{a\}$ in F and $F \setminus \{a\}$. Moreover, since a has no outgoing attacks the statement extends to the set $E \cup \{a\}$, i.e., $E^+_{F \setminus \{a\}} = E^+_F \setminus \{a\} = (E \cup \{a\})^+_F \setminus \{a\}$. Since a has no outgoing attacks, it follows that E defends the same arguments $b \in A \setminus \{a\}$ in F and $F \setminus \{a\}$. Thus the statement holds true for admissible, complete, grounded, and preferred semantics. For stable semantics, we furthermore observe that removing a only causes the removal of a from the range of a stable set E; moreover, a is not undecided if a stable set exists in $F \setminus \{a\}$, thus the statement follows. □

Let us next show that the translations preserve semantics when we restrict them to ABA frameworks with separated contraries.

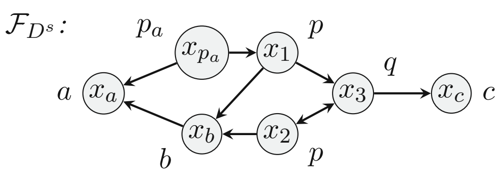
Proposition 3.3.25. *For an ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ with separated contraries, for $\sigma \in \{gr, co, pr, stb\}$, it holds that $\sigma_{pTh}(D) = \sigma_i(\mathcal{F}_{D\text{-}\mathcal{A}}) = \{S \setminus \mathcal{A} \mid S \in \sigma_i(\mathcal{F}_D)\}$.*

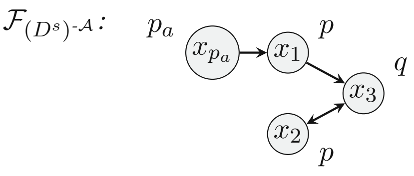
Proof. By Lemma 3.3.24, it holds that the removal of an argument $a \in A$ with no outgoing attacks in a given AF F is corresponds to the removal of a from each extension E of F. Coming back to our CAF instantiation $\mathcal{F}_D$, we obtain $E \in \sigma(F_D \setminus A')$ iff $E \in \sigma(F_D)$ or $E \setminus A' \in \sigma(F_D)$ for $A' = \{a \in A \mid cl(a) \in \mathcal{A}\}$. Since D is flat, the CAF $\mathcal{F}_D \setminus A'$ does not contain any arguments with claims in $\mathcal{A}$. The result thus carries over to claim-level: $S \in \sigma_i(\mathcal{F}_D \setminus A') = \sigma_i(\mathcal{F}_{D\text{-}\mathcal{A}})$ iff $S \in \sigma_i(\mathcal{F}_D)$ or $S \setminus \mathcal{A} \in \sigma_i(\mathcal{F}_D)$ for each set of claims $S \subseteq cl(A)$. By Proposition 3.3.20, we have $\sigma_{Th}(D) = \sigma_i(\mathcal{F}_D)$. The result follows when restricting $\sigma_i(\mathcal{F}_D)$ to the set of proper conclusions. ☐

Next we state that each ABA framework D can be transformed into an ABA framework D' with separated contraries while preserving assumption-extensions. Intuitively, we split all assumptions with two roles, i.e., being assumption and contrary of another assumption, into two literals: one taking over the assumptions-part, the other one the contrary-part.

Translation 3.3.26. *For an ABA $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$, we let $\mathcal{A}_c = \{a \in \mathcal{A} \mid \overline{a} \in \mathcal{A}\}$ denote the sets of all assumptions that have an assumption as contrary. We define the corresponding* separated ABA framework *$D^s = (\mathcal{L}', \mathcal{R}', \mathcal{A}, ^{-\prime})$ with $\mathcal{L}' = \mathcal{L} \cup \{p_a \mid a \in \mathcal{A}_c\}$, $\mathcal{R}' = \mathcal{R} \cup \{p_a \leftarrow \overline{a} \mid a \in \mathcal{A}_c\}$, and $\overline{a}' = p_a$ if $a \in \mathcal{A}_c$ and $\overline{a}' = \overline{a}$ otherwise.*

Example 3.3.27. *Let us consider again the ABA framework D from Example 3.3.19. D contains one assumption that has an assumption as contrary, i.e., $\mathcal{A}_c = \{a\}$. We construct $D^s = (\mathcal{L}', \mathcal{R}', \mathcal{A}, ^{-\prime})$ with $\mathcal{L}' = \mathcal{L} \cup \{p_a\}$ and $\overline{a}' = p_a$ moreover, we add $r_4 : p_a \leftarrow b$. Below, we depict the resulting CAFs (including resp. excluding assumption-arguments):*





The translation preserves the attack structure of D since each assumption-set $S \subseteq \mathcal{A}$ derives contraries of the same assumptions in both frameworks. We obtain that the assumption-extensions are preserved.

Proposition 3.3.28. *For an ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ and its corresponding ABA framework D^s with separated contraries obtained from Translation 3.3.26, it holds that $\sigma(D) = \sigma(D^s)$ for all semantics σ under consideration.*

The conclusion-extensions are extended by the additional arguments p_a. The advantage of Translation 3.3.26 is that each contrary is explicitly given as proper argument, which makes it possible to focus on the proper conclusion-extensions without any losses.

Going backwards: from CAF to ABA. In the next part of this section, we investigate which scenarios expressible via CAFs can be modeled with ABA frameworks. As a general observation, we note that without additional features such as e.g., preferences, a CAF can be modeled by an ABA framework only if it is well-formed.

In what follows, we treat assumptions as *background-information* which are *implicitly* given in the CAF. We show that for each well-formed CAF, it is possible to find an associated ABA framework in which the proper conclusion-extensions correspond to the claim-extensions of the CAF. For this, we identify each claim c in the CAF with the contrary of a hidden assumption a_c. Each argument x corresponds to a rule with head $cl(x)$ and body $\{a_{c_1}, \ldots, a_{c_n}\}$ where each c_i corresponds to a claim which attacks x.

Translation 3.3.29. *The associated ABA framework* $D_{\mathcal{F}} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$ *of a well-formed CAF* $\mathcal{F} = (A, R, cl)$ *is given by* $\mathcal{A} = \{a_c \mid c \in cl(A)\}$, $\mathcal{L} = \mathcal{A} \cup cl(A)$, *contrary function* $\overline{a_c} = c$ *for all* $c \in cl(A)$, *and* $\mathcal{R} = \{cl(x) \leftarrow \{a_{cl(y)} \mid y \in x^{-}\} \mid x \in A\}$.

The resulting ABA framework has separated contraries; moreover, the contrary of each assumption can be derived; also, each proper conclusion is the contrary of an assumption. This fragment of ABA has been already considered in the literature when studying the relation between ABA and LP [44].

Definition 3.3.30. *An ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$ *is an* LP-ABA framework *iff*

(i) $\mathcal{A} \cap \overline{\mathcal{A}} = \emptyset$ *(i.e., D has separated contraries);*

(ii) $\overline{a} \in Th_D(\mathcal{A})$ *for all* $a \in \mathcal{A}$*; and*

(iii) for each proper conclusion $p \in \mathcal{L}$ *of* D *, there is* $a \in \mathcal{A}$ *with* $p = \overline{a}$.

Moreover, the translation yields ABA frameworks with *atomic rules.*

Definition 3.3.31. *Given an ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$*, a rule* $r \in \mathcal{R}$ *is* atomic *iff* $body(r) \subseteq \mathcal{A}$. *The ABA framework D is called atomic iff each rule in $\mathcal{R}$ is atomic.*

The class of atomic LP-ABA frameworks is in one-to-one-correspondence with the class of copy-free well-formed CAFs, up to argument- respectively assumption-names.

Definition 3.3.32. *Two ABA frameworks* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^{-})$ *and* $D' = (\mathcal{L}', \mathcal{R}', \mathcal{A}', {}^{-'})$ *are equivalent up to assumption-names* ($D \cong_a D'$) *iff there is a mapping* $f : \mathcal{L} \to \mathcal{L}'$ *such that* $f(\mathcal{A}) = \mathcal{A}'$ *and* f *acts as identity function on* $\mathcal{L} \setminus \mathcal{A}$ *satisfying that* $\mathcal{L}' = f(\mathcal{L})$, $\mathcal{R}' = \{f(p) \leftarrow f(L) \mid p \leftarrow L \in \mathcal{R}\}$, *and* $\overline{a}' = \overline{f(a)}$ *for all* $a \in \mathcal{A}$.

Proposition 3.3.33. *For each copy-free well-formed CAF* $\mathcal{F}$, *it holds that* $\mathcal{F} \cong \mathcal{F}_{D^{\text{-}\mathcal{A}}_{\mathcal{F}}}$. *For each atomic LP-ABA framework* D, *it holds that* $D \cong_a D_{\mathcal{F}_{D\text{-}\mathcal{A}}}$.

Proof. Each argument x attacked by claims C in $\mathcal{F}$ corresponds to a rule $cl(x) \leftarrow C$ in $D_{\mathcal{F}}$ which yields an argument $y = C \vdash cl(x)$ with claim $cl(x)$ attacked by claims in C in $\mathcal{F}_{D_{\mathcal{F}}^{\mathcal{A}}}$. Since $\mathcal{F}$ is copy-free, each argument yields precisely one rule which in turn corresponds to precisely one argument.

Likewise, each rule r having assumptions $\mathcal{A}' \subseteq \mathcal{A}$ in its body corresponds to an argument x_r with claim $cl(x)$ which is attacked by assumptions in $\mathcal{A}'$. Since $\mathcal{A}$ does not contain any non-defeasible assumptions, and since D has separated contraries, each rule yields precisely one argument which in turn corresponds again to precisely one rule. ☐

Since each CAF can be transformed into a copy-free CAF without changing the semantics, we obtain that the translation preserves proper conclusion-extensions.

Corollary 3.3.34. *For each well-formed CAF $\mathcal{F}$ and semantics $\sigma \in \{gr, co, pr, stb\}$, it holds that $\sigma_i(\mathcal{F}) = \sigma_{pTh}(D_{\mathcal{F}})$.*

We furthermore obtain that each ABA framework can be equivalently expressed as atomic LP-ABA framework without changing proper conclusion-extensions.

3.4 Claims and Collective Attacks

Argumentation Frameworks with collective Attacks (SETAFs), as introduced by Nielsen and Parsons [140], generalize the binary attack-relation in AFs to collective attacks of arguments. In this section, we discover a surprising connection between CAFs and SETAFs: each SETAF corresponds to a well-formed CAF and vice versa. Hence they provide alternative abstract representations of conflicting information in knowledge bases.

Definition 3.4.1. *A SETAF is a pair $SF = (A, R)$ where A is a finite set of arguments and $R \subseteq (2^A \setminus \{\emptyset\}) \times A$ is the attack relation.*

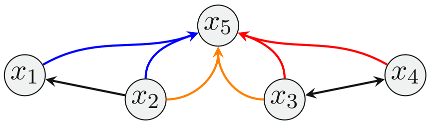
Given a SETAF $SF = (A, R)$, we say that $S \subseteq A$ *attacks* a if there is a set $S' \subseteq S$ with $(S', a) \in R$ (we also call attacks in SETAFs set-attacks). S is *conflicting* in SF if S attacks some $a \in S$; otherwise S is *conflict-free* in SF ($S \in cf(SF)$). We write $a^-_{SF} = \{T \subseteq A \mid (S, a) \in R\}$ for $a \in A$; moreover, $S^+_{SF} = \{a \mid \exists S' \subseteq S : (S', a) \in R\}$ and $S^{\oplus}_{SF} = S \cup S^+_{SF}$ for $S \subseteq A$ (we omit subscript if it does not cause confusion). We say a is *defended* by S if $a^-_{SF} \subseteq S^+_{SF}$. AF semantics generalize to SETAFs as follows.

Definition 3.4.2. *Let $SF = (A, R)$ be a SETAF. For a set $S \in cf(SF)$, we say*

- *$S \in ad(SF)$ if each $a \in S$ is defended by S in SF;*
- *$S \in na(SF)$, if there is no $T \in cf(SF)$ with $T \supset S$,*
- *$S \in co(SF)$, if $S \in ad(SF)$ and $a \in S$ for all $a \in A$ defended by S in SF;*

- $S \in gr(SF)$, *if* $S = \bigcap_{T \in co(SF)} T$;
- $S \in stb(SF)$, *if each* $a \in A \setminus S$ *is attacked by* S *in* SF;
- $S \in pr(SF)$, *if* $S \in ad(SF)$ *and there is no* $T \supset S$ *such that* $T \in ad(SF)$.
- $S \in stg(F)$, *if* $\nexists T \in cf(F)$ *with* $T^{\oplus}_{SF} \supset S^{\oplus}_{SF}$, *and*
- $S \in ss(F)$, *if* $S \in ad(F)$ *and* $\nexists T \in ad(F)$ *s.t.* $T^{\oplus}_{SF} \supset S^{\oplus}_{SF}$.

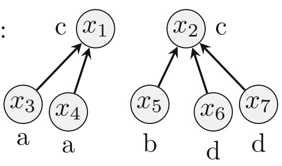
Example 3.4.3. *Let us consider the following SETAF SF depicted below:*



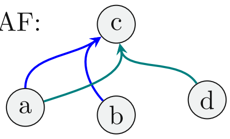
SF contains three collective attacks: each set $\{x_1, x_2\}$, $\{x_2, x_3\}$, *and* $\{x_3, x_4\}$ *attacks* x_5. *Moreover,* x_1 *and* x_2 *as well as* x_3 *and* x_4 *mutually attack each other. The admissible extensions of SF are* $\emptyset$, $\{x_2\}$, $\{x_3\}$, $x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, *and* $\{x_2, x_4, x_5\}$ *(because* x_2 *defends* x_5 *against the attack from* $\{x_1, x_2\}$ *and* x_4 *defends* x_5 *against the remaining collective attacks since it attacks* x_3*). The sets* $\{x_2, x_3\}$ *and* $\{x_2, x_4, x_5\}$ *are stable (observe that* $\{x_2, x_3\}$ *collectively attacks* x_5*).*

In what follows, we will discuss the close correspondence of well-formed CAFs and SETAFs. The crucial observation is that a set of claims S in a well-formed CAF that attacks all occurrences of a claim c can be interpreted as a collective attack from S on c. That is, if a CAF $\mathcal{F}$ contains two arguments x_1, x_2 with the same claim c, and x_1 is attacked by claim a while x_2 is attacked by claims b and d, then both the set $\{a, b\}$ and $\{a, d\}$ can be seen as collective attack on c. Note that the well-formedness property is necessary to speak of *claims* attacking arguments (since each claim attacks the same arguments). We will use this interpretation to relate claims with arguments in SETAFs; collective attacks are then determined by considering all possible (minimal) sets of claims which jointly defeat a given claim by attacking all of its occurrences. Consider the CAF $\mathcal{F}$ from before; the resulting SETAF will have two collective attacks $\{a, b\}$ and $\{a, d\}$ determined by the attackers of x_1 and x_2. We sketch both frameworks below:

CAF $\mathcal{F}$:



Resulting SETAF:



Technically, we construct the SETAF by determining the *minimal hitting sets* of the attackers of all occurrences of a given claim.

Definition 3.4.4. *Let $\mathcal{M}$ be a set of sets. We call $\mathcal{H}$ a* hitting set *of $\mathcal{M}$ if $\mathcal{H} \cap M \neq \emptyset$ for each $M \in \mathcal{M}$. By $HS_{min}(\mathcal{M})$ we denote the $\subseteq$-minimal hitting sets of $\mathcal{M}$.*

We will make use of the following result.

Lemma 3.4.5 ([37])**.** *Let $X = \{X_1, \dots, X_n\}$ be a set of sets with $X_i \not\subseteq X_j$ for $i \neq j$. It holds that $HS_{min}(HS_{min}(X)) = X$.*

Representing CAFs as SETAFs. Let us now formally state our translation from well-formed CAFs to SETAFs. We obtain our SETAF as sketched above. First, each claim c in the CAF is an argument in the SETAF. Regarding the attacker of c in the SETAF, we consider the set of attacking claims $cl(x_1^-), \dots, cl(x_n^-)$ of the occurrences $x1, \dots, x_n$ of c; the argument c is then attacked by the minimal hitting sets of $\{cl(x_1^-), \dots, cl(x_n^-)\}$.

Translation 3.4.6. *For a well-formed CAF $\mathcal{F} = (A, R, cl)$, we define the corresponding SETAF $SF_{\mathcal{F}} = (A_{SF_{\mathcal{F}}}, R_{SF_{\mathcal{F}}})$ by letting $A_{SF_{\mathcal{F}}} = cl(A)$ and*

$$R_{SF_{\mathcal{F}}} = \{(T, c) \mid c \in cl(A), T \in HS_{min}(\{cl(x_R^-) \mid x \in A, cl(x) = c\})\}.$$

We observe that the SETAF obtained by Translation 3.4.6 satisfies $\subseteq$-minimality of the attacking sets due to the attack construction via minimal hitting sets, i.e., for every two attacks $(S, c), (S', c) \in R_{SF_{\mathcal{F}}}$, it holds that $S' \not\subset S$. SETAFs satisfying this property are said to be in *normal form* [148].

We show that the translation preserves all classical Dung semantics.

Proposition 3.4.7. *For each well-formed CAF $\mathcal{F}$, its associated SETAF $SF_{\mathcal{F}}$, and semantics $\sigma \in \{cf, ad, co, gr, pr, stb\}$, it holds that $\sigma_i(\mathcal{F}) = \sigma(SF_{\mathcal{F}})$.*

Proof. The proof relies on the following correspondence, which is true by definition of the translation: (a) a set S of arguments in $SF_{\mathcal{F}}$ (claims in $\mathcal{F}$) attacks c in $SF_{\mathcal{F}}$ iff S attacks each occurrence of c in $\mathcal{F}$.

First, we show that (1) conflict-free sets coincide. A set S of claims in $\mathcal{F}$ is conflict-free iff it has a realization E in $\mathcal{F}$ which is conflict-free, that is, E is not attacked by any claims in S: for each claim $c \in S$, there is some occurrence of c in $\mathcal{F}$ which is unattacked by S. By the above observation (a), this is equivalent to S does not attack c in $SF_{\mathcal{F}}$. Thus S is conflict-free in $\mathcal{F}$ iff S is conflict-free in $SF_{\mathcal{F}}$.

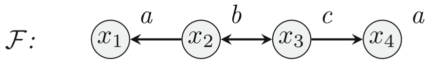
Next, we show that (2) a set of claims S defends an occurrence of a claim c in $\mathcal{F}$ iff S defends c in $SF_{\mathcal{F}}$. Given a claim c in $\mathcal{F}$ (argument c in $SF_{\mathcal{F}}$), and let $x_1, \dots, x_n$ denote all occurrences of c in $\mathcal{F}$. Now, S defends c in $\mathcal{F}$ iff there is some argument x_i which is defended against each attack. Due to well-formedness, this is the case iff S attacks each occurrence of each claim in $cl(x_i^-)$. This in turn means that for all $d \in cl(x_i^-)$ there is $S' \subseteq S$ such that S' attacks d in $SF_{\mathcal{F}}$. Since the set $HS_{min}(\{cl(x_1^-), \dots, cl(x_n^-)\})$ contains all sets that attack c in $SF_{\mathcal{F}}$ and since each such T contains some $d \in cl(x_i^-)$, this is equivalent to for all $T \in HS_{min}(\{cl(x_1^-), \dots, cl(x_n^-)\})$, there is $d \in T$, there is $S' \subseteq S$ such that S' attacks d in $SF_{\mathcal{F}}$, or, in other words, S defends c in $SF_{\mathcal{F}}$.

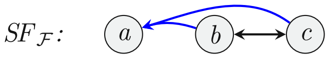
It follows that admissible semantics coincide: S is admissible in $\mathcal{F}$ iff it has an admissible realization iff S defends an occurrence of each $c \in S$ (by well-formedness of $\mathcal{F}$) iff S defends c in $SF_{\mathcal{F}}$ (by (2)). Since the same claims are defended it follows that complete semantics coincide. We thus obtain $\sigma_i(\mathcal{F}) = \sigma(SF_{\mathcal{F}})$ for all except stable semantics.

Regarding stable semantics, first recall that a set S attacks c in $SF_{\mathcal{F}}$ iff S attacks all occurrences of c in $\mathcal{F}$. Thus in case S is stable in $\mathcal{F}$ we have S is conflict-free in $SF_{\mathcal{F}}$ by (1) and for each $c \in cl(A) \setminus S$, each occurrence of c is attacked. By (a), this implies that each such c is attacked by S in $SF_{\mathcal{F}}$. It follows that $stb_i(\mathcal{F}) \subseteq stb(SF_{\mathcal{F}})$. In case S is stable in $SF_{\mathcal{F}}$, it holds that S is conflict-free in $\mathcal{F}$ and attacks each occurrence of each claim $c \in cl(A) \setminus S$. Now, consider a cf_i-realization E of S in $\mathcal{F}$. Since E attacks each occurrence of each claim $c \in cl(A) \setminus S$, it remains to deal with those arguments having claims in S but are not contained in E. It holds that each such argument x with $cl(x) = c \in S$ is either attacked by S in $\mathcal{F}$ or $E \cup \{x\}$ is conflict-free: indeed, in case $E \cup \{x\}$ is conflicting but x is not attacked by E it holds that x attacks E. Consequently, by well-formedness, this means that each argument with c attacks E since $c \in S$, thus E is not conflict-free, contradiction to the assumption. Thus x is attacked or we can extend E by adding x and consider $E' = E \cup \{x\}$ instead. Repeating this step for all arguments $x \in A$ with $cl(x) \in S$ thus yields a stable realization of S in $\mathcal{F}$. It follows that stable semantics coincide. □

The result shows that multiple occurrences of claims in CAFs can be equivalently treated as collective attacks, if the framework satisfies well-formedness. The translation however does not preserve naive, semi-stable, and stage semantics.

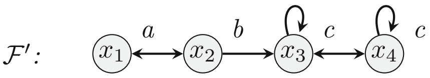
Example 3.4.8. *Let us consider the following CAF $\mathcal{F}$ and its associated SETAF $\mathcal{F}_{SF}$:*

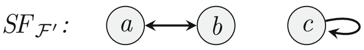




The naive extensions in F are $\{x_1, x_3\}$, $\{x_1, x_4\}$, and $\{x_2, x_4\}$, yielding the claim-sets $\{a, c\}$, $\{a\}$, and $\{b, c\}$. The naive extensions of $SF_{\mathcal{F}}$, however, are $\{a, c\}$ and $\{b, c\}$.

For semi-stable and stage semantics, consider the following CAF and associated SETAF:

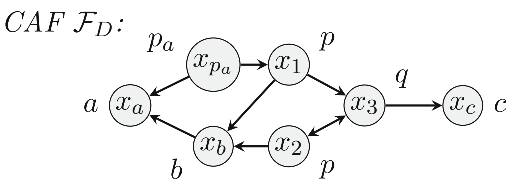


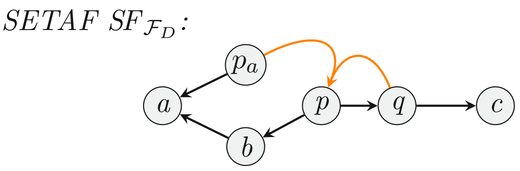


The unique semi-stable and stage extension in F' is $\{x_2\}$, yielding the claim-set $\{b\}$. In $SF_{\mathcal{F}'}$, however, both $\{a\}$ and $\{b\}$ are semi-stable and stage.

A natural question that arises is about the contents the arguments in the resulting SETAF are representing with respect to an initial instantiation. To this end, let us have one more look on our running example ABA framework from Section 3.3.2.

Example 3.4.9. *Let us recall the ABA framework* $D^s = (\mathcal{L}, \mathcal{R}, \mathcal{A}, {}^-)$ *with separated contraries from Example 3.3.27: the framework contains the assumptions* $\mathcal{A} = \{a, b, c\}$ *and rules* $r_1 : p \leftarrow a$, $r_2 : p \leftarrow c$, $r_3 : q \leftarrow b$, *and* $r_4 : p_a \leftarrow b$. *Contraries are given as follows:* $\overline{a} = p_a$, $\overline{b} = p$, *and* $\overline{c} = q$. *In* D, *we have two arguments* x_1, x_2 *with conclusion* p *(arising from rule* r_1 *and* r_2*) where* x_1 *is attacked by* p_a *and* x_2 *is attacked by* q. *Transforming the associated CAF into a SETAF thus yields a single argument* p *which is attacked by the set* $\{p_a, q\}$. *We depict the CAF and the resulting SETAF below:*





Representing SETAFs as CAFs. We will dedicate the following part of this section to the translation of SETAFs to well-formed CAFs. Given a SETAF, we obtain the corresponding CAF by relating arguments with claims as follows: For each argument $c \in A$ with attacking sets $T_1, \dots, T_n$, we introduce an argument for each minimal hitting set of $\{T_1, \dots, T_n\}$. The translation is defined as follows:

Translation 3.4.10. *For a SETAF* $SF = (A, R)$, *we define the corresponding CAF* $\mathcal{F}_{SF} = (A_{\mathcal{F}_{SF}}, R_{\mathcal{F}_{SF}}, cl_{\mathcal{F}_{SF}})$ *with* $A_{\mathcal{F}_{SF}} = \{x_{c,H} \mid c \in A,\ H \in HS_{min}(c^-_R)\}$, $cl_{SF}(x_{c,H}) = c$ *for all* $x_{c,H} \in A_{\mathcal{F}_{SF}}$, *and* $R_{\mathcal{F}_{SF}} = \{(x_{c,H}, y_{c',H'}) \mid c \in H'\}$.

Similar as for Translation 3.4.6, we observe that the translation results in a well-formed CAF that satisfies a certain minimality property: for each claim $c \in cl(A)$, it holds that $x^- \subseteq y^-$ implies $x = y$ for all occurrences x, y of c in $\mathcal{F}_{SF}$. Such arguments are a special type of redundant arguments. From Theorem 3.1.15 we know that such arguments can be safely removed without changing the semantics of the CAF.

We will show that Translation 3.4.6 and Translation 3.4.10 are each other's inverse (up to isomorphism between argument-names in CAFs) when restricted to the class of all SETAFs in minimal form and to the class of all normalized CAFs, respectively. Recall that, by Proposition 3.4.7, $\sigma_i(\mathcal{F}) = \sigma(SF_{\mathcal{F}})$ for each well-formed CAF and for $\sigma \in \{cf, ad, co, gr, pr, stb\}$, consequently we get that $\sigma(SF) = \sigma(SF_{\mathcal{F}_{SF}}) = \sigma_i(\mathcal{F}_{SF})$.

Proposition 3.4.11. *For each normalized well-formed CAF* $\mathcal{F}$, *it holds that* $\mathcal{F} \cong \mathcal{F}_{SF_{\mathcal{F}}}$. *For each SETAF* SF *in normal form, it holds that* $SF = SF_{\mathcal{F}_{SF}}$.

Proof. Given a normalized well-formed CAF $\mathcal{F}$, a claim c, and its occurrences $x_1, \dots, x_n$ in $\mathcal{F}$. The attackers of c in $SF_{\mathcal{F}}$ are $c^-_{SF_{\mathcal{F}}} = HS_{min}(\{cl(x_1^-), \dots, cl(x_n^-)\})$. Note that the set $\{cl(x_1^-), \ldots, cl(x_n^-)\}$ of attacking claims of c satisfies $cl(x_i^-) \not\subseteq cl(x_j^-)$ for $i \neq j$, i.e., its elements are pairwise incomparable due to $\mathcal{F}$ being in normal form. Conversely, we identify the occurrences and their attacker of claim c in the CAF $\mathcal{F}_{SF_\mathcal{F}}$ via the minimal hitting sets $HS_{min}(c^-_{SF_\mathcal{F}})$: for each set $H \in HS_{min}(\{c^-_{SF_\mathcal{F}}\})$, we introduce a new argument $x_{c,H}$ with claim c that is attacked by all claims in H. By Lemma 3.4.5, we obtain $HS_{min}(c^-_{SF_\mathcal{F}}) = \{cl(x_1^-), \ldots, cl(x_n^-)\}$ and thus $\mathcal{F} \cong \mathcal{F}_{SF_\mathcal{F}}$. To show that for each SETAF SF in normal form, it holds that $SF = SF_{\mathcal{F}_{SF}}$, we proceed analogously. First, we construct CAF $\mathcal{F}_{SF}$ using Translation 3.4.10 and introduce, for each argument $c \in A$, an occurrence of c in $\mathcal{F}$ for each set $H \in HS_{min}(c^-_{SF})$. Since the attackers $c^-_{SF} = \{T_1, \ldots, T_n\}$ of c are pairwise incomparable, Lemma 3.4.5 applies and we obtain $HS_{min}(\{x^-_{\mathcal{F}_{SF}} \mid cl(x) = c\}) = \{T_1, \ldots, T_n\}$. □

Corollary 3.4.12. *For each SETAF SF and semantics* $\sigma \in \{cf, ad, co, gr, pr, stb\}$, *it holds that* $\sigma(SF) = \sigma_i(\mathcal{F}_{SF})$.

We have shown that SETAFs and well-formed CAFs can be transformed into each other.

Theorem 3.4.13. *Let* $\sigma \in \{cf, ad, co, gr, pr, stb\}$. *For any well-formed CAF* $\mathcal{F}$, *there is a SETAF SF such that* $\sigma_i(\mathcal{F}) = \sigma(SF)$, *and vice versa.*

3.5 Discussion

In this chapter, we presented several instantiations of non-monotonic reasoning formalisms and examined the role of claims in these formalisms. We discussed logic programming, structured argumentation, in particular ASPIC+ and ABA, and SETAFs in connection of CAFs. We have seen that in all of these formalisms, claims play an important role; be it for attack construction in the instantiation of LPs or ASPIC+ instances, or crucial for defining semantics in ABA, as well as for splitting collective attacks into single attacks on arguments with the same claim.

We showed that LPs, ABA frameworks, and SETAFs are closely connected to well-formed CAFs: for each instance of the aforementioned formalisms, there is a well-formed CAF which yields the same extensions under classical Dung semantics, and vice versa. We identified several translations between well-formed CAFs, and LPs, ABA frameworks, as well as SETAFs, respectively: Translations 3.3.29, 3.3.18, and 3.3.22 connect well-formed CAFs and ABA frameworks; Translations 3.4.6 and 3.4.10 establish a connection between well-formed CAFs and SETAFs; and Translations 3.2.13 and 3.2.9 connect well-formed CAFs and normal LPs. We showed that all translations preserve complete, preferred, grounded, and stable semantics. We identified fragments of the aforementioned formalisms on which the translations are each other's inverse (up to isomorphism): copy-free well-formed CAFs are in one-to-one correspondence with atomic normal LPs as well as atomic LP-ABA frameworks; moreover, normalized well-formed CAFs and SETAFs in normal form are in one-to-one correspondence. See Figure 3.1 for an overview of the relations presented in this work.

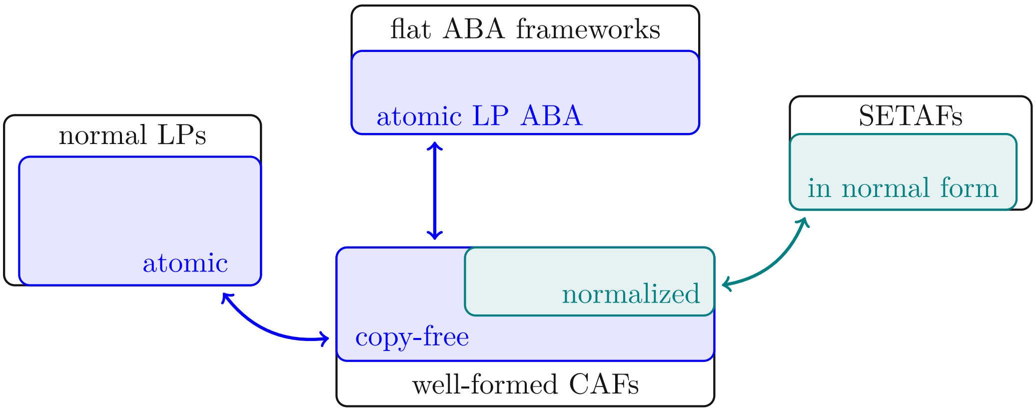


Figure 3.1: Overview of the Dung-semantics-preserving translations between CAFs, ABA frameworks, SETAFs, and LPs presented in this work. Fragments of the formalisms which are in one-to-one correspondence are indicated in blue and teal, respectively.

We furthermore showed that ASPIC+ instances potentially yield CAFs that are not well-formed. While the underlying mechanism of attack construction is well-formed, the incorporation of preferences labels some attacks unsuccessful, thus violating well-formedness. We showed that each CAF can be modeled by an induced sub-graph of an ASPIC+ instance.

3.5.1 Related Work (or: Translations are everywhere!)

The relation between different formalisms is frequently discussed in the literature; exploring the connection between different formalisms gives a better understanding of the underlying concepts and mechanisms of the involved formalisms.

There exists further semantics-preserving translations between several classes of the aforementioned formalisms. The connection between LP and ABA has been frequently discussed in the literature. The fragment of ABA that corresponds to normal logic programs, also known as *LP-ABA* (cf. Definition 3.3.30), appears, for instance, in [44, 75, 71, 59]. Moreover, the correspondence between SETAFs in normal form and *redundancy-free* atomic LPs has been discussed in [88]. An LP P is redundancy-free iff there are no rules $r_1, r_2 \in P$ with $body(r_1) \subseteq body(r_2)$. It has been shown that the class of SETAFs in normal form and the class of redundancy-free atomic LPs are in one-to-one correspondence with each other. Notable to mention in this context is also the close connection of all of the aforementioned formalisms with *abstract dialectical argumentation frameworks (ADFs)* [50]. ADFs generalize AFs by assigning each argument an acceptance condition (a logical formula) which encodes the relation between the arguments. Strass [170] and Alcântara et. al [2] provide semantics-preserving translations between LPs and ADFs. Here, The latter translation maps logic programs to *attacking (support-free) ADFs*—a subclass of ADFs, where the acceptance condition is in disjunctive

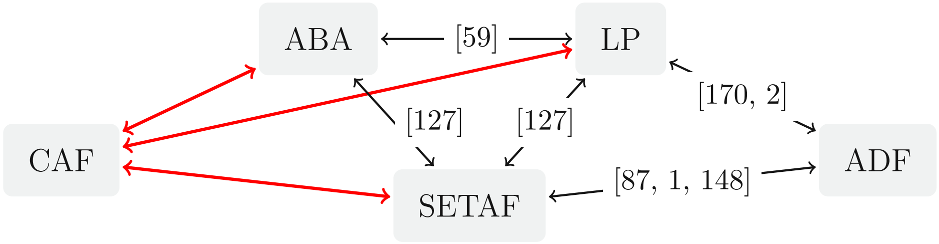


Figure 3.2: Overview of existing and novel (highlighted in red) transformations.

normal form. This subclass of ADFs has also been investigated by Dvořák et. al [87] and Alcântara and Sá [1] in its relation to SETAFs. Finally, the other direction, i.e. translating a SETAF into an attacking ADF, has been covered by Polberg [148]. The close connection of ADFs to LPs, ABA frameworks, and SETAFs, respectively, (together with our novel CAF translations) can be used to obtain translations between CAFs and ADFs as well. A closer investigation of the connection between the two formalisms without detours would be an interesting avenue for future research on this matter.

Hence if we collect all available results, we obtain the following insight: (classes of) ABA frameworks, LPs, ADFs, SETAFs, and CAFs can all be viewed, to some extent, as different sides of the same (pentagonal) coin. We summarize this insight in Figure 3.2.

3.5.2 Beyond well-formedness: Preferences and Uncertainties

As we have seen in Section 3.3.1, the incorporation of preferences yields CAFs which are no longer well-formed. In this regard, we mention Bernreiter et al. [38] who study *CAFs with preferences* and discuss several attack modifications studied in the literature, e.g., the removal or reversal of attacks [125]. They show that these preference incorporation techniques violate well-formedness; moreover, they identify several classes of CAFs lying between well-formed and general CAFs. In contrast to their work, we show that *each* CAF corresponds to an induced sub-graph of an ASPIC+ instance.

Apart from preferences in argumentation, we further mention two other well-established formalisms that might yield CAFs which are not well-formed: *probabilistic frameworks* [130] allow for probabilities assigned to arguments or attacks and *incomplete AFs* [30, 31] in which arguments or attacks can be uncertain. Both probabilistic and incomplete AFs are used to model situations in which attacks (or arguments) possibly exist e.g., when merging several frameworks that represent the subjective world-view of different agents. Even if the underlying attack relation of the framework is well-formed, the incorporation or probabilities of uncertainties of attacks leads to the removal of attacks which in turn might violate well-formedness. We consider a closer study on probabilistic argumentation and incomplete AFs in context with claims a promising direction for future research, in particular in relation with structural aggregation of several frameworks [32].

CHAPTER 4

Claim-sensitive Semantics

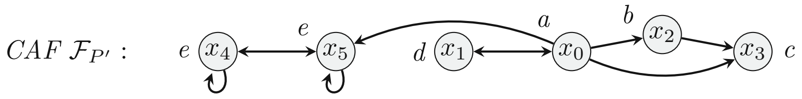
Claims are an integral part of formal argumentation and other knowledge representation formalisms. They play a significant role when connecting logic programming and argumentation; moreover, they are crucial for the construction of the attack relation in structured argumentation formalisms; finally, they yield an alternative view on collective attacks when interpreting them as joint effort to defeat a given claim. Overall, the representation as CAF helps to streamline many instantiation procedures. We have, however, experienced some shortcomings when we deviate from classical Dung semantics: Examples 3.2.12, 3.3.21, and 3.4.8 demonstrate that the translations between CAFs and LPs, ABA frameworks, and SETAFs, respectively, do not preserve semi-stable semantics (L-stable semantics, respectively) which belong to the group of *range-based semantics*, i.e., semantics that take rejected entities (arguments, claims, atoms, respectively) into account. We also experienced certain mismatches for naive semantics for SETAFs, indicating issues already when it comes to *maximization*.

To identify possible sources of irregularities let us recall our LP example from Section 3.2:

Example 3.2.12 (cont.). *Consider again our LP P' from Example 3.2.12 with rules*

$r_0 : a \leftarrow \mathtt{not}\ d.$ $\quad r_1 : d \leftarrow \mathtt{not}\ a.$ $\quad r_2 : b \leftarrow d.$
$r_3 : c \leftarrow d, \mathtt{not}\ b.$ $\quad r_4 : e \leftarrow \mathtt{not}\ e.$ $\quad r_5 : d \leftarrow \mathtt{not}\ a, \mathtt{not}\ e.$

and the CAF $\mathcal{F}_{P'}$ corresponding to P' following Translation 3.2.9:



The L-stable models of P' are $\{a\}$ and $\{d, b\}$; the CAF $\mathcal{F}_{P'}$, however, has only one semi-stable extension, namely $\{x_0\}$; extracting the claim yields the claim-set $\{a\}$. Consequently, L-stable and semi-stable semantics do not necessarily yield the same outcome.

As discussed in the previous chapter, AF semantics are powerful enough to handle many semantics in non-monotonic reasoning semantics very well: for classical Dung semantics it often suffices to evaluate the underlying AF and extract the claim at the very end of the procedure (cf. Definition 2.2.5 of *inherited semantics* for CAFs). The example above, however, indicates certain mismatches between inherited CAF semantics and the intended evaluation of the underlying knowledge base (in this case, the LP P) when it comes to other than classical Dung semantics. A crucial observation is that semantics for LPs operate on conclusion (claim) level while abstract argumentation semantics as well as inherited CAF semantics are evaluated on argument level. We are thus interested in developing adequate semantics for CAFs which mimic the behavior of semantics performing maximization on conclusion-level of the original problem.

We observe two sources that may lead to a different outcome of the evaluation methods:

- First, maximization is considered on different levels. In LPs, we maximize over sets of atoms while in the associated CAFs we maximize over arguments. This, however, is a mismatch since atoms in the LP correspond to *claims* in the CAF.
- The second issue is more subtle: while we successfully identify the claims of acceptable arguments with atoms that are set to true, we do not have a similar correspondence for atoms that are set to false. Coming back to our running example, we observe that the arguments with claim e play a different role for the claim-sets $\{a\}$ and $\{b, d\}$ (the realization $\{x_0\}$ of $\{a\}$ attacks one of them while the realization $\{x_1, x_2\}$ of $\{b, d\}$ does not) although the atom e is undecided with respect to both L-stable models of P. The underlying issue is that evaluation methods for CAFs that have been considered so far do not take the *defeat of claims*, i.e., the successful attack of *all* occurrences of a given claim, into account.

To resolve these issues, we will make use of a powerful advantage of CAFs: they are flexible enough to capture semantics that make use of the conclusions in the evaluation. This advantage, however, has not been fully exploited so far. Inspired by the observations above, we propose semantics that shift certain evaluation steps to the level of claims. With these adjustments, we are able to capture semantics of conclusion-oriented formalisms. Let us demonstrate the idea:

Example 3.2.12 (cont.). *Let us consider again our CAF $\mathcal{F}_{P'}$ and its complete argument-sets $\emptyset$, $\{x_0\}$, and $\{x_1, x_2\}$. We propose a new evaluation method for semi-stable semantics by maximizing accepted and defeated claims: The set $\{x_0\}$ defeats the claims b, c, d; the claim e is not defeated because x_0 does not attack all occurrences of e. The set of accepted and defeated claims with respect to the extension $\{x_0\}$ (the* claim-range *of $\{x_0\}$) is thus given by $\{a, b, c, d\}$. The set $\{x_1, x_2\}$ defeats the claims a, c, thus $\{x_1, x_2\}$ has claim-range $\{a, b, c, d\}$ which coincides with the claim-range of $\{x_0\}$.*

Both sets are maximal with respect to accepted and defeated claims (the claim-range of $\emptyset$ is empty; the set $\{x_1, x_3\}$ only defeats claim a). The evaluation method yields indeed the same outcome as L-stable model semantics for P.

In this chapter, we introduce novel semantics for CAFs that put claims into a stronger position. We choose a *hybrid approach*: we consider defeat and maximization on *claim-level* while the acceptance of claims depends on their realizations on *argument-level*. With our novel *hybrid semantics* (*h-semantics*) for CAFs, we propose an adaption of range-based semantics for CAFs that covers maximization on atom-level in LPs and thus gives rise to the missing argumentation-based counterpart of L-stable model semantics. Likewise, our novel semantics correspond to semantics for SETAFs and (a fragment of) ABA frameworks. We introduce new variants of naive, preferred, stable, semi-stable, and stage semantics in Section 4.1, settle their relations in Section 4.2, and show in Section 4.3 that h-semantics indeed fill many gaps in the landscape of claim semantics.

4.1 Introducing Hybrid Semantics

In this section, we establish claim-based semantics that perform maximization on sets of acceptable claims as well as on the range on claim-level. For this, we establish a defeat notion for claims: intuitively, a claim is defeated if each occurrence of the claim is attacked. Our investigations give rise to novel versions of preferred and naive semantics (when considering maximization of claim-sets) which are discussed in Section 4.1.1; variants of stable semantics (using our novel notion of claim-defeat) which are introduced in Section 4.1.2; and semi-stable and stage semantics (when maximizing over sets of accepted and defeated claims) which combine both aspects and are discussed in Section 4.1.3.

4.1.1 Maximization of Claim-Sets

Let us start with two prominent semantics involving maximization: preferred and naive semantics return $\subseteq$-maximal admissible resp. conflict-free sets. We introduce variants of preferred and naive semantics by shifting maximization from argument- to claim-level.

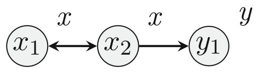
Definition 4.1.1. *Given a CAF $\mathcal{F}$ and a set of claims $S \subseteq cl(A_{\mathcal{F}})$. Then*

- *S is* h-preferred *($S \in pr_h(\mathcal{F})$) iff S is $\subseteq$-maximal in $ad_i(\mathcal{F})$;*
- *S is* h-naive *($S \in na_h(\mathcal{F})$) iff S is $\subseteq$-maximal in $cf_i(\mathcal{F})$.*

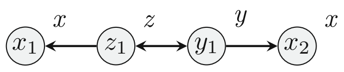
*For a set $S \in pr_h(\mathcal{F})$ ($S \in na_h(\mathcal{F})$), we call a set $E \in ad(F)$ ($E \in cf(F)$) with $cl(E) = S$ a pr_h-*realization *(na_h-*realization*, resp.) of S in $\mathcal{F}$.*

Example 4.1.2. *Let us consider the following two CAFs $\mathcal{F}$ and $\mathcal{F}'$:*

$\mathcal{F}$:



$\mathcal{F}'$:



The CAF $\mathcal{F}$ already appears in Example 2.2.7; it is not well-formed. Its i-preferred and i-naive claim-sets are $\{x\}$ and $\{x, y\}$ since $\{x_1, y_1\}$ and $\{x_2\}$ are naive and preferred in F.

To compute the h-naive and h-preferred claim-sets of $\mathcal{F}$, we first compute the admissible and naive claim-sets of $\mathcal{F}$, which yields the conflict-free claim-sets $\{x\}$, $\{y\}$, and $\{x, y\}$; and the admissible claim-sets $\{x, y\}$ and $\{x\}$. Thus $pr_h(\mathcal{F}) = na_h(\mathcal{F}) = \{\{x, y\}\}$.

The CAF $\mathcal{F}'$, yield the same claim-sets under inherited and hybrid preferred semantics, namely the sets $\{x, y\}$ and $\{x, z\}$. For naive semantics, the variants differ: inherited semantics yield the sets $\{x\}$, $\{x, y\}$ and $\{x, z\}$ while hybrid semantics return $\{x, y\}$ and $\{x, z\}$. Observe that $\mathcal{F}'$ is well-formed.

Next we show that each h-preferred (h-naive) claim-set is also i-preferred (i-naive).

Proposition 4.1.3. *For each CAF $\mathcal{F}$, $\sigma_h(\mathcal{F}) \subseteq \sigma_i(\mathcal{F})$ for $\sigma \in \{pr, na\}$.*

Proof. Consider a set $S \in \sigma_h(\mathcal{F})$ and let E denote an admissible (conflict-free) realization of S in $\mathcal{F}$ that is $\subseteq$-maximal among all admissible (conflict-free) realizations of S. We observe that E is a $\subseteq$-maximal admissible (conflict-free) set in F: otherwise, there is an admissible (conflict-free) set $D \subseteq A_{\mathcal{F}}$ with $E \subset D$. By choice of E, D contains an argument a with claim $cl(a) \notin S$. Thus we have found an admissible (conflict-free) set of claims $cl(D)$ that properly extends S, contradiction to $\subseteq$-maximality of S. ☐

The other direction does not hold: We have already seen that i-preferred as well as i-naive claim-sets are not necessarily I-maximal (cf. Example 4.1.2); h-preferred and h-naive semantics, on the other hand, yield I-maximal sets per definition.

The above proposition reveals an alternative view on h-preferred and h-naive semantics: they can be equivalently defined by maximizing over i-preferred or i-naive sets, respectively.

Proposition 4.1.4. *For a CAF $\mathcal{F}$ and a set of claims $S \subseteq cl(A_{\mathcal{F}})$, it holds that*

- *$S \in pr_h(\mathcal{F})$ iff S is $\subseteq$-maximal in $pr_i(\mathcal{F})$;*
- *$S \in na_h(\mathcal{F})$ iff S is $\subseteq$-maximal in $na_i(\mathcal{F})$.*

Proof. In Proposition 4.1.3, we have already seen that each h-σ claim-set is contained in $\sigma_i(\mathcal{F})$. We moreover observe that each set that is $\subseteq$-maximal in $\sigma_i(\mathcal{F})$ is also $\subseteq$-maximal in $ad_i(\mathcal{F})$ ($cf_i(\mathcal{F})$, resp.) by monotonicity of the claim-function; moreover, each $\subseteq$-maximal i-preferred (i-naive) claim-set is has an admissible (conflict-free) realization. ☐

For well-formed CAFs, both variants of preferred semantics coincide, as we show next.

Proposition 4.1.5. *For each well-formed CAF $\mathcal{F}$, it holds that $pr_i(\mathcal{F}) = pr_h(\mathcal{F})$.*

Proof. We show that $pr_i(\mathcal{F})$ is I-maximal (i.e., $S \subset T$ implies $S = T$ for all $S, T \in pr_i(\mathcal{F})$). By Proposition 4.1.4, it follows that h-preferred and i-preferred semantics coincide.

Let $E, D \in pr(F)$, $E \neq D$. We show that $cl(E) \not\subseteq cl(D)$. First assume, there exists an $a \in E$ attacking some $b \in D$ in F. It follows that $cl(a) \notin cl(D)$, otherwise the argument $c \in D$ with $cl(c) = cl(a)$ also attacks b due to well-formedness; since D is conflict-free, this cannot be the case. Suppose now that no $a \in E$ attacks some $b \in D$. We need at least one attack (a, b) from E to D, otherwise $E \cup D \in pr(F)$. But then E needs to attack b since E is admissible, so we are done. □

For naive semantics, we cannot hope for an analogous result: the two variants might yield different claim-sets as the CAF $\mathcal{F}'$ from Example 4.1.2 demonstrates. The example also shows that i-naive semantics violates I-maximality (even for well-formed CAFs).

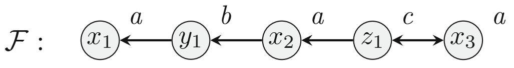
4.1.2 Introducing Claim-Attacks: Stable Semantics

Our next step is to establish the notion of *defeat of claims*. As sketched in the introduction of this chapter, inherited semantics lack a notion of claim-defeat that indicates the difference between *defeated* and *undecided* claims. Recall that in the CAF $\mathcal{F}$ associated to the LP in Example 3.2.12, the partial attack from set $\{x_0\}$ on the claim e (only one occurrence of e has been attacked) has led to accepting only the set $\{a\}$ as semi-stable claim-set, although the claim-range of $\{a\}$ and $\{b, d\}$ coincide. Our goal is to establish a definition of claim-defeat that renders e in this situation as undecided. Hereby, we call a claim defeated *if all occurrences are attacked*. Our choice is justified as such a behavior can be observed by LPs and other formalisms that evaluate on conclusion-level.

Let us furthermore point out that defeating a claim be achieved by a set of arguments and not by a set of claims. In Example 3.2.12, another argument would be necessary that helps x_0 to attack all occurrences of e.

Definition 4.1.6. *For a CAF $\mathcal{F}$, we say that a set of arguments $E \subseteq A_{\mathcal{F}}$* defeats *a claim $c \in cl(A_{\mathcal{F}})$ iff for all $x \in A_{\mathcal{F}}$ with $cl(x) = c$, there is $y \in E$ such that $(y, x) \in R$. We write $E^*_{\mathcal{F}} = \{c \in cl(A_{\mathcal{F}}) \mid E \text{ defeats } c \text{ in } \mathcal{F}\}$ to denote the set of claims defeated by E in $\mathcal{F}$.*

Example 4.1.7. *Consider the CAF $\mathcal{F}$ given as follows:*



The set of arguments $\{y_1, z_1\}$ defeats claim a ($\{y_1, z_1\}^*_{\mathcal{F}} = \{a\}$) because each occurrence of a is attacked: y_1 attacks x_1, and z_1 attacks x_2 and x_3. Moreover, the argument x_2 defeats claim b as it attacks y_1 which is the unique argument carrying this claim.

We are ready to define the *claim-range* as a claim-based counterpart to the range in AFs. Again, the claim-range depends on a set of arguments. Intuitively, the claim-range of a set of arguments E contains all claims that are *accepted* by E, i.e., all claims contained in E, as well as all claims that are *rejected* by E, i.e., all claims that are defeated by E.

Definition 4.1.8. *Given a CAF $\mathcal{F}$ and a set $E \subseteq A_\mathcal{F}$. By $E^\circledast_\mathcal{F} = cl(E) \cup E^*_\mathcal{F}$ we denote the* claim-range *of E in $\mathcal{F}$. If $E^\circledast_\mathcal{F} = cl(A_\mathcal{F})$ we say that E has* full claim-range *in $\mathcal{F}$.*

Example 4.1.9. *Let us consider again the CAF $\mathcal{F}$ from Example 4.1.7. The claim-range of $\{y_1, z_1\}$ is given by $\{a, b, c\}$ (i.e., $\{y_1, z_1\}^\circledast_\mathcal{F} = \{a, b, c\}$). Thus the set has full claim-range, i.e., it holds that $cl(A_\mathcal{F}) = \{y_1, z_1\}^\circledast_\mathcal{F}$. For $\{x_2\}$ we obtain $\{x_2\}^\circledast_\mathcal{F} = \{a, b\}$.*

From Example 4.1.7, we learn that the claim-range with respect to a given set of *claims* is in general not unique: the realization $\{x_1, x_2, x_3\}$ of claim a has full claim-range, while the realization $\{x_1\}$ has claim-range $\{a\}$, and the realization $\{x_2\}$ has claim-range $\{a, b\}$.

For well-formed CAFs, however, each claim-set admits a unique claim-range: recall that claims attack the same arguments in each well-formed CAF $\mathcal{F}$, i.e., $E^+_F = D^+_F$ for every realization of a given claim-set S. It follows that each realization defeats the same claims.

Proposition 4.1.10. *For a well-formed CAF $\mathcal{F}$ and a set of claims $S \subseteq cl(A_\mathcal{F})$, it holds that $E^*_\mathcal{F} = D^*_\mathcal{F}$ and $E^\circledast_\mathcal{F} = D^\circledast_\mathcal{F}$ for every two realizations E, D of S in $\mathcal{F}$.*

For well-formed CAFs, it thus makes sense to speak about the claim-range of a claim-set S.

Definition 4.1.11. *For a well-formed CAF $\mathcal{F}$ and a set of claims $S \in cl(A_\mathcal{F})$, we write $S^\circledast_\mathcal{F}$ ($= E^\circledast_\mathcal{F}$ for all realizations E of S in $\mathcal{F}$) to denote the unique claim-range of S.*

Intuitively, we consider a set to be hybrid stable if it has full claim-range: a set of claims S is h-stable if it has a realization E with full claim-range, i.e., $E^\circledast_\mathcal{F} = cl(A_\mathcal{F})$. Following AF semantics, we require that E is conflict-free. While a stable set of arguments is also admissible we observe that this is in general not the case for CAFs:

Example 4.1.12. *In the CAF $\mathcal{F}$ from Example 4.1.7, the set $\{a, b\}$ is h-stable, following our intuitive definition: the realization $E = \{y_1, x_3\}$ is conflict-free and defeats claim c, thus E has full claim-range. However, E is not admissible in F since y_1 is not defended.*

We thus consider also a variant of stable semantics that requires admissibility.

Definition 4.1.13. *Given a CAF $\mathcal{F}$ and a set $S \subseteq cl(A_\mathcal{F})$. We say that*

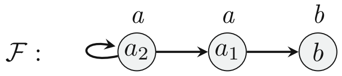
- *S is a* h-*cf*-stable claim-set *($S \in cf\text{-}stb_h(\mathcal{F})$) iff there exists a cf_i-realization E of S in $\mathcal{F}$ such that $E^\circledast_\mathcal{F} = cl(A_\mathcal{F})$;*
- *S is a* h-*ad*-stable claim-set *($S \in ad\text{-}stb_h(\mathcal{F})$) iff there exists an ad_i-realization E of S in $\mathcal{F}$ such that $E^\circledast_\mathcal{F} = cl(A_\mathcal{F})$.*

*A set E $cf\text{-}stb_h$-*realizes *a claim-set S iff $cl(E) = S$, $E \in cf(F)$, and $E^\circledast_\mathcal{F} = cl(A_\mathcal{F})$; likewise, E $ad\text{-}stb_h$-*realizes *a claim-set S iff $cl(E) = S$, $E \in ad(F)$, and $E^\circledast_\mathcal{F} = cl(A_\mathcal{F})$.*

Although the two variants of stable semantics differ in many aspects as we will see, there are several occasions in which it makes sense to address both of them. In such situations we simply write h-stable semantics instead of h-*cf*-stable and h-*ad*-stable semantics.

The proposed variants relax inherited stable semantics. Indeed, a set of arguments E can have full claim-range without attacking all arguments that are not contained in E. It suffices that *some* argument with claim c is contained in E in order to accept c.

Example 4.1.14. *Let us consider the following CAF $\mathcal{F}$:*



The framework has no stable extension since there are no arguments that attack the self-attacker a_2. Moreover, the only admissible set is $\emptyset$, thus there is no h-ad-stable claim-set either. We, however, obtain a h-cf-stable claim-set by considering the set $\{a_1\}$: the argument defeats claim b and carries claim a, thus $\{a_1\}^{\circledast}_{\mathcal{F}} = \{a, b\} = cl(A_{\mathcal{F}})$. We obtain that $cf\text{-}stb_h(\mathcal{F}) = \{\{a\}\}$. Observe that $\mathcal{F}$ is not well-formed.

Proposition 4.1.15. *For any $\mathcal{F}$, $stb_i(\mathcal{F}) \subseteq ad\text{-}stb_h(\mathcal{F}) \subseteq cf\text{-}stb_h(\mathcal{F})$.*

Proof. To show that $stb_i(\mathcal{F}) \subseteq ad\text{-}stb_h(\mathcal{F})$, we observe that each stable extension E of the underlying AF F is admissible and attacks all remaining arguments. Thus, each claim is either accepted by E (i.e., E contains an occurrence of the claim in question) or defeated by E. We obtain $E^{\circledast}_{\mathcal{F}} = cl(A_{\mathcal{F}})$ for each stable extension of F. Moreover, we observe that each set of arguments E that realizes a h-*ad*-stable claim-set is also conflict-free. Consequently, we obtain that $ad\text{-}stb_h(\mathcal{F}) \subseteq cf\text{-}stb_h(\mathcal{F})$. □

The CAF $\mathcal{F}$ from Example 4.1.14 satisfies $ad\text{-}stb_h(\mathcal{F}) \neq cf\text{-}stb_h(\mathcal{F})$. A small modification of $\mathcal{F}$ shows that $ad\text{-}stb_h(\mathcal{F}) \neq stb_i(\mathcal{F})$: if we delete the attack from a_2 to a_1 we obtain a single h-*ad*-stable claim-set $\{a\}$ (witnessed by the *ad*-realization $\{a_1\}$ in F) but $stb_i(\mathcal{F}_1) = \emptyset$. Observe that both considered CAFs are not well-formed. We will show next that for well-formed CAFs, all considered variants of stable semantics coincide.

Proposition 4.1.16. *$stb_i(\mathcal{F}) = ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$ for each well-formed CAF $\mathcal{F}$.*

Proof. We show that $cf\text{-}stb_h(\mathcal{F}) \subseteq stb_i(\mathcal{F})$: Consider a h-*cf*-stable claim-set S and a $cf\text{-}stb_h$-realization E of S in $\mathcal{F}$ that is $\subseteq$-maximal among all conflict-free realizations of S. We show that E is stable in the AF F. We show that E attacks all arguments that are not contained in E, i.e., $E^+_F = A_{\mathcal{F}} \setminus E$. Let $x \in A_{\mathcal{F}} \setminus E$ and let $cl(x) = c$. In case $c \notin S$, we have that all occurrences of c—including x—are attacked. Consider now the case $c \in S$, i.e., there is an argument $y \in E$ such that $cl(y) = c$. By maximality of E, we observe that $E \cup \{x\}$ is not conflict-free; thus either (a) $(x, x) \in R$ or there is $z \in E$ such that either (b) $(z, x) \in R$ or (c) $(x, z) \in R$. In case (a) then also $(y, x) \in R$ by well-formedness; in case (b) we are done; in case (c) we have $(y, z) \in R$ by well-formedness and therefore E is not conflict-free, contradiction.

We obtain that $cf\text{-}stb_h(\mathcal{F}) \subseteq stb_i(\mathcal{F})$. By Proposition 4.1.15, $stb_i(\mathcal{F}) \subseteq ad\text{-}stb_h(\mathcal{F}) \subseteq cf\text{-}stb_h(\mathcal{F})$, thus the statement follows. ☐

Finally, we show that both variants of stable semantics allow for alternative characterizations in terms of i-complete and i-preferred semantics (for admissible-based h-stable semantics) and in terms of i-naive semantics (for conflict-free-based stable semantics).

Proposition 4.1.17. *Given a CAF $\mathcal{F}$ and a set of claims $S \subseteq cl(A_{\mathcal{F}})$. The following statements are equivalent:*

1. *$S \in ad\text{-}stb_h(\mathcal{F})$;*
2. *there is a co_i-realization E of S in $\mathcal{F}$ with $E^{\circledast}_{\mathcal{F}} = cl(A_{\mathcal{F}})$;*
3. *there is a pr_i-realization E of S in $\mathcal{F}$ with $E^{\circledast}_{\mathcal{F}} = cl(A_{\mathcal{F}})$.*

Moreover, the following two statements are equivalent:

4. *$S \in cf\text{-}stb_h(\mathcal{F})$;*
5. *there is a na_i-realization E of S in $\mathcal{F}$ with $E^{\circledast}_{\mathcal{F}} = cl(A_{\mathcal{F}})$.*

Proof. To prove (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3), we first observe that (3) $\Rightarrow$ (2) $\Rightarrow$ (1) follows from the inclusions $pr(F) \subseteq co(F) \subseteq ad(F)$. To show (1) $\Rightarrow$ (3), consider a set $S \in ad\text{-}stb_h(\mathcal{F})$ and let E denote an ad_i-realization of S in $\mathcal{F}$ with $S \cup E^{*}_{\mathcal{F}} = cl(A_{\mathcal{F}})$. Then there is some $D \in pr(F)$ with $D \supseteq E$. We show that D is a pr_i-realization of S in $\mathcal{F}$, that is, $cl(D) = S$: Towards a contradiction, assume that there is some $c \in cl(A_{\mathcal{F}}) \setminus S$ such that $c \in cl(D)$, that is, there is some $x \in D$ with $cl(x) = c$. By $S \cup E^{*}_{\mathcal{F}} = cl(A_{\mathcal{F}})$ we have $c \in E^{*}_{\mathcal{F}}$ thus there is some $y \in E \subseteq D$ that attacks x in F, contradiction to D being conflict-free. It follows that $cl(D) = S$; moreover, D attacks each claim in $cl(A_{\mathcal{F}}) \setminus S$ by monotonicity of $\cdot^{*}$, thus the statement follows.

To prove (4) $\Leftrightarrow$ (5), it suffices to show (4) $\Rightarrow$ (5); the other direction is immediate since $cf(F) \subseteq na(F)$. Now, let $S \in ad\text{-}stb_h(\mathcal{F})$ and let E denote a cf_i-realization of S in $\mathcal{F}$ with $S \cup E^{*}_{\mathcal{F}} = cl(A_{\mathcal{F}})$. Similar as above, we consider a naive extension D in F with $E \subseteq D$ and show that $cl(D) = S$: In case there is some claim $c \in cl(A_{\mathcal{F}}) \setminus S$ that is contained in $cl(D)$, there is some $y \in E \subseteq D$ that attacks an argument $x \in D$ with claim $cl(x) = c$, contradiction to D being conflict-free. We obtain that D is a na_i-realization of S in $\mathcal{F}$ that defeats all claims in $cl(A_{\mathcal{F}}) \setminus S$. ☐

4.1.3 Bringing the Two Together: Semi-Stable and Stage Semantics

Semi-stable and stage semantics make use of both methods that we have established in the preceding sections: they are designed to minimize undecidedness (starting from admissible or conflict-free sets, respectively). In terms of claims, semi-stable and stage semantics return $\subseteq$-maximal sets of claims that are either accepted or defeated with respect to a given set of arguments.

Semi-stable and stage semantics weaken stable semantics by dropping the requirement that the claim-range has to contain all claims that are present in the framework.

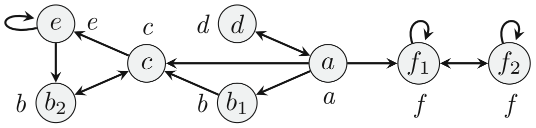
Definition 4.1.18. *Given a CAF $\mathcal{F}$ and a set of claims $S \subseteq cl(A_{\mathcal{F}})$. We say that*

- *S is a* h-stage claim-set *($S \in stg_h(\mathcal{F})$) iff there exists a cf_i-realization E of S in $\mathcal{F}$ such that there is no $D \in cf(F)$ with $E^{\circledast}_{\mathcal{F}} \subset D^{\circledast}_{\mathcal{F}}$;*
- *S is a* h-semi-stable claim-set *($S \in ss_h(\mathcal{F})$) iff there exists an ad_i-realization E of S in $\mathcal{F}$ such that there is no $D \in ad(F)$ with $E^{\circledast}_{\mathcal{F}} \subset D^{\circledast}_{\mathcal{F}}$.*

*A set E stg_h-*realizes *a claim-set S iff $cl(E) = S$, $E \in cf(F)$, and $E^{\circledast}_{\mathcal{F}}$ is $\subseteq$-maximal; likewise, E ss_h-*realizes *a claim-set S iff $cl(E) = S$, $E \in ad(F)$, and $E^{\circledast}_{\mathcal{F}}$ is $\subseteq$-maximal.*

In contrast to the semantics we considered so far, we observe that the proposed variant of semi-stable semantics neither constitutes a strengthening nor a weakening of its inherited counterpart. The following example shows that even for well-formed CAFs, h-semi-stable and i-semi-stable semantics potentially yield different claim-sets.

Example 4.1.19. *Consider the following well-formed CAF $\mathcal{F}$:*



The admissible claim-sets of $\mathcal{F}$ are given by $S_1 = \{d\}$, $S_2 = \{b, d\}$ and $S_3 = \{a\}$. Let us now consider the claims they defeat: S_1 defeats claim a, S_2 defeats the claims c and a; and S_3 defeats claims c and d. Computing the claim-range of the sets yields the range $\{a, d\}$ for S_1; the range $\{a, b, c, d\}$ for S_2, and $\{a, c, d\}$ for S_3 (recall that for well-formed CAFs, each realization of a claim-set has the same range). We obtain that $ss_h(\mathcal{F}) = \{\{b, d\}\}$. Observe that $\{a\}$ is the only i-semi-stable claim-set.

Regarding stage semantics, we observe that $\{c\}$ and $\{b, d\}$ are h-stage while $\{c\}$ and $\{a\}$ are i-stage in $\mathcal{F}$. Hence we see that both semi-stable and stage semantics yield different sets in both variants.

Finally, we consider alternative characterizations of both range-based semantics.

Proposition 4.1.20. *Given a CAF $\mathcal{F}$ and a set of claims $S \subseteq cl(A_{\mathcal{F}})$. The following statements are equivalent:*

1. $S \in ss_h(\mathcal{F})$;
2. *there is a co_i-realization E of S in $\mathcal{F}$ with $\subseteq$-maximal claim-range $E^{\circledast}_{\mathcal{F}}$ among complete extensions;*
3. *there is a pr_i-realization E of S in $\mathcal{F}$ with $\subseteq$-maximal claim-range $E^{\circledast}_{\mathcal{F}}$ among preferred extensions.*

Moreover, the following two statements are equivalent:

4. $S \in stg_h(\mathcal{F})$;
5. *there is a na_i-realization E of S in $\mathcal{F}$ with $E^{\circledast}_{\mathcal{F}} = cl(A_{\mathcal{F}})$.*

Proof. The proof proceeds analogous to the proof of Proposition 4.1.17. To prove (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3), we first observe that (3) $\Rightarrow$ (2) $\Rightarrow$ (1) follows from the inclusions $pr(F) \subseteq co(F) \subseteq ad(F)$. To show (1) $\Rightarrow$ (3), consider a set $S \in ss_h(\mathcal{F})$ and let E denote a ss_h-realization of S in $\mathcal{F}$, that is, $E^{\circledast}_{\mathcal{F}}$ is $\subseteq$-maximal among admissible extensions. Then there is some $D \in pr(F)$ with $D \supseteq E$. As in the proof of Proposition 4.1.17, we obtain that D is a pr_i-realization of S in $\mathcal{F}$; moreover, D defeats each claim that is defeated by E by monotonicity of $\cdot^*$, and thus $E^{\circledast}_{\mathcal{F}} = D^{\circledast}_{\mathcal{F}}$ holds. Consequently, $D^{\circledast}_{\mathcal{F}}$ is $\subseteq$-maximal among preferred extensions: Assume otherwise, then there is a preferred extension T in F with $T^{\circledast}_{\mathcal{F}} \supset D^{\circledast}_{\mathcal{F}} = E^{\circledast}_{\mathcal{F}}$, contradiction to $\subseteq$-maximality of $E^{\circledast}_{\mathcal{F}}$ in F among admissible extensions. We have shown $D^{\circledast}_{\mathcal{F}}$ is $\subseteq$-maximal among preferred extensions, thus the statement follows.

Likewise, we show (4) $\Rightarrow$ (5) to prove the equivalence (4) $\Leftrightarrow$ (5); the other direction is immediate since $cf(F) \subseteq na(F)$. Let $S \in stg_h(\mathcal{F})$ and let E denote a stg_h-realization of S in $\mathcal{F}$. As in the proof of Proposition 4.1.17, there exists a naive extension D in F with $E \subseteq D$ and $cl(D) = S$; similar as above, we obtain that $D^{\circledast}_{\mathcal{F}}$ is $\subseteq$-maximal among naive extensions. Thus the statement follows. □

4.1.4 Summary

In the preceding subsections, we introduced novel variants of claim-based argumentation semantics by lifting certain evaluation-steps onto claim-level. Performing maximization on claim-level gave rise to alternative variants of preferred and naive semantics. We discussed claim-defeat which led to two novel hybrid variants of stable semantics; finally, bringing the two together gave rise to hybrid semi-stable and stage semantics.

Interestingly, it turned out that h-preferred and i-preferred as well as all stable variants collapse when we consider them on well-formed CAFs. This means that if arguments with the same claim have the same outgoing attacks, it holds that argument-level and claim-level maximization of *admissible* sets yield the same outcome. Also, if stable extensions in well-formed CAFs defeat all claims it follows that all arguments are attacked as well. Hence for stable semantics in well-formed CAFs, claim-defeat and argument-attacks are interchangeable concepts. However, as we have seen, the notions do not coincide, even if the CAF is well-formed: range-based semantics potentially yield a different outcome as Example 4.1.19 demonstrates. This means as soon as we relax the condition and move to $\subseteq$-maximality instead of universal quantification over the set of all arguments/claims not contained in the extension we observe fundamental differences between claim-defeat and argument-attack. Likewise, claim-set and argument-set maximization on arbitrary sets does not necessarily yield the same outcome in well-formed CAFs. As we have seen, i-naive and h-naive extensions potentially differ (cf. Example 4.1.2). It turns out that admissibility plays an important role for the concurrence of i- and h-preferred semantics.

Let us end this section with a brief discussion about our focus on claim-based variants of maximization and defeat and why we did not provide a claim-based variant of defense (explaining the lack of hybrid variants of admissible, complete, and grounded semantics). Generally speaking, the reason is that claim-defense coincides with their traditional argument-based counter-part. Let us take a closer look on the notion. Intuitively, defense obeys the following logic: an entity (e.g., an argument, a claim) is *defended* iff each attacking unit is counter-attacked. Now, with our notion of claim-defeat at hand, this abstract view gives rise to the following notion of claim-defense:

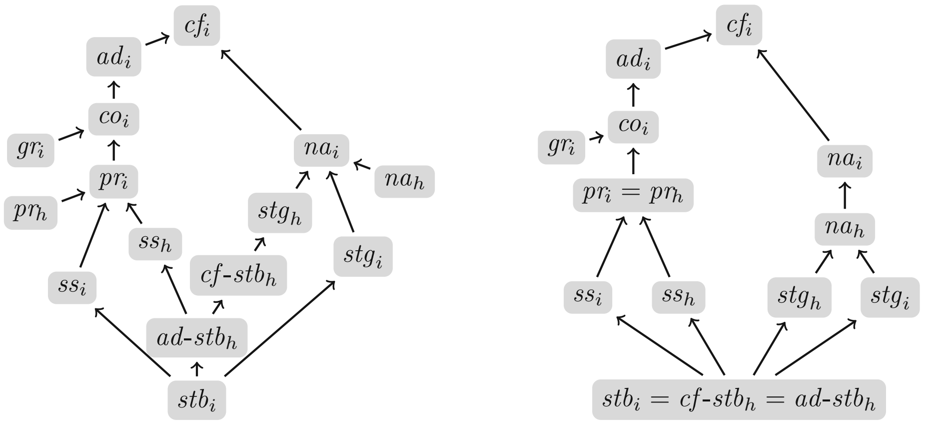
a set of arguments E claim-defends *a claim c in a given CAF $\mathcal{F}$ iff*
E attacks each set of arguments D that claim-defeats c.

That is, E must attack some argument $b \in D$ for each attacking set D of c. This means that there must be some argument x with claim c that is defended by E (in the underlying AF); otherwise, we can find a set of arguments that claim-defeats c but is not attacked by E. With these combinatorial considerations, claim-defense can be reformulated as follows: a set of arguments E claim-defends a claim c in $\mathcal{F}$ iff there exists an argument x with claim c that is defended by E in F. Thus claim-defense coincides with classical defense on argument-level.

Having settled all CAF semantics of interest, let us fix some basic notations and conventions used from now on in this work.

Notation 4.1.21. *For an AF semantics σ we denote by σ_i the inherited variant of σ and by σ_h the hybrid variant of σ. We write ρ to denote any CAF semantics.*

Notation 4.1.22. *We sometimes drop 'inherited' or 'hybrid' (prefix 'i-' or 'h-', resp.) when speaking about a semantics for which only one version exists or if both variants coincide. For example, we refer to 'i-grounded semantics' by 'grounded semantics' since it has no hybrid variant; and in the context of well-formed CAFs, we simply say 'stable semantics' instead of 'inherited', 'h-cf-'. or 'h-ad-stable semantics' because all variants coincide.*



(a) Relations between CAF semantics. (b) Relations for well-formed CAFs.

Figure 4.1: Relations between semantics for general CAFs (a) and well-formed CAFs (b). An arrow from σ to τ indicates that $\sigma(\mathcal{F}) \subseteq \tau(\mathcal{F})$ for each (well-formed) CAF $\mathcal{F}$.

4.2 Relations between Semantics

In this section, we will settle the relation between all hybrid and inherited semantics. We first state a general observation which clarifies the relation between inherited and hybrid semantics in case every argument possesses a unique claim. In that case, both variants coincide with the standard AF semantics.

Lemma 4.2.1. *For any $\sigma \in \{pr, na, stb, ss, stg\}$ and CAF $\mathcal{F}$ with $cl(x) = x$ for all $x \in A_{\mathcal{F}}$, we have $\sigma_h(\mathcal{F}) = \sigma_i(\mathcal{F}) = \sigma(F)$.*

It follows that negative results (via counter-examples) showing that two AF semantics are not in a subset-relation immediately apply to (well-formed) CAFs.

Theorem 4.2.2. *The relations between the semantics depicted in Figure 4.1 hold.*

As already discussed in Section 2.1 the relations between inherited semantics follow from the corresponding relations for AFs. In Section 4.1 the relations between semantics that are based on the same Dung semantics have been settled: For arbitrary CAFs we have

$$stb_i(\mathcal{F}) \subseteq ad\text{-}stb_h(\mathcal{F}) \subseteq cf\text{-}stb_h(\mathcal{F}) \quad \text{and} \quad pr_h(\mathcal{F}) \subseteq pr_i \text{ and } na_h(\mathcal{F}) \subseteq na_i$$

by Proposition 4.1.15 and 4.1.3. For well-formed CAFs, all stable variants coincide (by Proposition 4.1.16), also, i-preferred and h-preferred semantics yield the same outcome (by Proposition 4.1.5). Finally, semi-stable and stage semantics are incomparable, even in the well-formed case (cf. Example 4.1.19).

Next we discuss the remaining $\subseteq$-relations. First, we notice that each h-*ad*-stable claim-set is h-semi-stable, since each such set has full (and thus $\subseteq$-maximal) claim-range; likewise, each h-*cf*-stable set is h-stage.

Proposition 4.2.3. *$ad\text{-}stb_h(\mathcal{F}) \subseteq ss_h(\mathcal{F})$ and $cf\text{-}stb_h(\mathcal{F}) \subseteq stg_h(\mathcal{F})$ for each CAF $\mathcal{F}$.*

Furthermore, recall that h-semi-stable and h-stage semantics can be equivalently defined via preferred and naive semantics, respectively (cf. Proposition 4.1.20). We thus obtain that each h-semi-stable (h-stage) claim-set is h-preferred (h-naive, respectively).

Proposition 4.2.4. *$ss_h(\mathcal{F}) \subseteq pr_i(\mathcal{F})$ and $stg_h(\mathcal{F}) \subseteq na_i(\mathcal{F})$ for each CAF $\mathcal{F}$.*

This concludes the proofs for all $\subseteq$-relations for admissible-based semantics as shown in Figure 4.1 for both well-formed and general CAFs.

Although h-naive semantics does not coincide with i-naive semantics in the well-formed case, we observe that h-naive semantics joins in the $\subseteq$-chain of conflict-free-based semantics: for well-formed CAFs, h-naive semantics extend both i-stage and h-stage semantics.

Lemma 4.2.5. *$stg_h(\mathcal{F}) \subseteq na_h(\mathcal{F})$, $stg_i(\mathcal{F}) \subseteq na_h(\mathcal{F})$ for each well-formed CAF $\mathcal{F}$.*

Proof. First, consider a set $S \in stg_h(\mathcal{F})$. Towards a contradiction, assume $S \notin na_h(\mathcal{F})$. That is, there is some $T \in cf_i(\mathcal{F})$ with $T \supset S$. Since $\mathcal{F}$ is well-formed, each realization of S and T attacks the same claims. By monotonicity of the range-function, we obtain that $D^{\circledast}_{\mathcal{F}} \supset E^{\circledast}_{\mathcal{F}}$ for each realization D of T and E of S; contradiction to $S \in stg_h(\mathcal{F})$.

Now, consider a set $S \in stg_i(\mathcal{F})$, i.e., there is a set $E \subseteq A_{\mathcal{F}}$ with $cl(E) = S$ such that $E \cup E^+_F$ is maximal w.r.t. subset-relation. Now, assume that $S \notin na_h(\mathcal{F})$, i.e. there exists a set $T \in cf_i(\mathcal{F})$ such that $T \supset S$. Consider a cf_i-realization D of T in $\mathcal{F}$. Now, since E is stage in F, there is some $x \in E \cup E^+_F$ such that $x \notin D \cup D^+_F$. By well-formedness, $D^+_F \supseteq E^+_F$, thus we have $x \in E$ and $x \notin D$. We can assume that x and D are conflicting; otherwise consider $D' = D \cup \{x\}$ instead. Since x and D are conflicting and since $x \notin D^+_F$, there exists $y \in D$ such that $(x, y) \in R$. Since $T \subset S$, there is $z \in D$ such that $cl(x) = cl(z)$. By well-formedness, $(z, y) \in R$, contradiction to D being conflict-free. □

Having settled all positive relations between semantics, we discuss counter-examples for the remaining cases. We obtain counter-examples for the following cases by Lemma 4.2.1.

Proposition 4.2.6. *Let `Sem` be the set of all semantics under our consideration. There is a well-formed CAF $\mathcal{F}$ such that $\alpha(\mathcal{F}) \nsubseteq \beta(\mathcal{F})$ for*

1. $\alpha = cf_i$, $\beta \in \texttt{Sem} \setminus \{cf_i\}$;
2. $\alpha = ad_i$, $\beta \in \texttt{Sem} \setminus \{cf_i, ad_i\}$;
3. $\alpha = co_i$, $\beta \in \texttt{Sem} \setminus \{cf_i, ad_i, co_i\}$;
4. $\alpha = gr_i$, $\beta \in \texttt{Sem} \setminus \{cf_i, ad_i, co_i, gr_i\}$;

5. $\alpha \in \{pr_h, pr_i\}$, $\beta \in \texttt{Sem} \setminus \{cf_i, ad_i, co_i, pr_h, pr_i\}$;

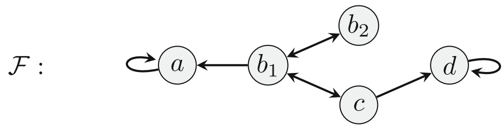
6. $\alpha \in \{na_h, na_i\}$, $\beta \in \texttt{Sem} \setminus \{cf_i, na_h, na_i\}$;

7. $\alpha \in \{ss_h, ss_i\}$, $\beta \in \{stg_h, stg_i, na_h, na_i, cf\text{-}stb_h, ad\text{-}stb_h, stb_i\}$ *and*

8. $\alpha \in \{stg_h, stg_i\}$, $\beta \in \{ad_i, ss_h, ss_i, pr_h, pr_i, cf\text{-}stb_h, ad\text{-}stb_h, stb_i\}$.

It remains to provide a counter-example for the absence of $\subseteq$-relations between ss_i, ss_h and pr_h (stg_i, stg_h and na_h respectively) for general CAFs.

Example 4.2.7. *Consider the following (non-well-formed) CAF $\mathcal{F}$:*

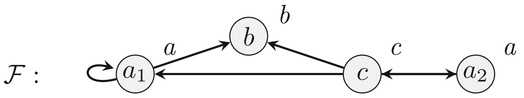


Let us first note that in $\mathcal{F}$, the set of conflict-free and admissible sets coincides, thus it holds that $pr_h(\mathcal{F}) = na_h(\mathcal{F})$, $ss_h(\mathcal{F}) = stg_h(\mathcal{F})$, and $ss_i(\mathcal{F}) = stg_i(\mathcal{F})$. The sets $E_1 = \{b_1\}$ and $E_2 = \{b_2, c\}$ are $\subseteq$-maximal conflict-free sets in F and have $\subseteq$-maximal (claim-)range. Hence $\{b\}$ and $\{b, c\}$ are i- and h-semi-stable and stage in $\mathcal{F}$. On the other hand, $\{b, c\}$ is the unique h-naive and h-preferred claim-extension of $\mathcal{F}$.

The crucial observation in the above example is that h-naive and h-preferred semantics are I-maximal while the others are not; i.e., it might be the case that semi-stable and stage variants yield claim-sets S, T that are in $\subseteq$-relation to each other ($S \subset T$). Among other principles, we will discuss this property in depth in Chapter 5.

Finally, let us discuss the connection between h-stable and h-semi-stable and h-stage semantics. Recall that for inherited semantics, $stb_i(\mathcal{F}) = ss_i(\mathcal{F}) = stg_i(\mathcal{F})$ in case $stb_i(\mathcal{F}) \neq \emptyset$. We observe that this does not extend to h-stable semantics.

Example 4.2.8. *Let us consider the following CAF $\mathcal{F}$:*



In $\mathcal{F}$, we have $ad\text{-}stb_h(\mathcal{F}) = ss_h(\mathcal{F}) = \{\{c\}\}$ and $cf\text{-}stb_h(\mathcal{F}) = stg_h(\mathcal{F}) = \{\{c\}, \{a, d\}\}$.

However, we can obtain the following weaker version.

Lemma 4.2.9. *For any CAF $\mathcal{F}$, (a) $cf\text{-}stb_h(\mathcal{F}) \neq \emptyset$ implies $cf\text{-}stb_h(\mathcal{F}) = stg_h(\mathcal{F})$ and (b) $ad\text{-}stb_h(\mathcal{F}) \neq \emptyset$ implies $ad\text{-}stb_h(\mathcal{F}) = ss_h(\mathcal{F})$.*

Proof. In case $cf\text{-}stb_h(\mathcal{F})$ is non-empty, it holds that for each $S \in stb_h(\mathcal{F})$, there is a cf_i-realization E of S in $\mathcal{F}$ such that $E^{\circledast}_{\mathcal{F}} = cl(A_{\mathcal{F}})$. We obtain $stb_h(\mathcal{F}) = stg_h(\mathcal{F})$. Similar arguments hold for the respective admissible-based semantics. □

4.3 Non-monotonic Reasoning Formalisms Revisited: Extending Relations

In this section, we revisit our non-monotonic reasoning formalisms from Chapter 3 and show that our h-semantics capture semantics which do not have a corresponding AF counter-part, meaning that inherited semantics failed to model their behavior in CAFs. For this, we first will discuss the normal form of CAFs and show that h-semantics are preserved under normalization. Next we will consider logic programs and show that L-stable semantics correspond to h-semi-stable semantics and vice versa. Moreover, we prove the correspondence between semi-stable ABA semantics and h-semi-stable CAF semantics. Finally, we revisit collective attacks and show that our h-semantics indeed provide counter-parts for naive, semi-stable, and stage SETAF semantics. Overall, our results demonstrate that hybrid semantics are often the natural choice.

Background & Notation. *In this section, we make use of concepts introduced in Section 3. We refer the reader to Section 3.2, 3.3.2, and 3.4 for an overview about logic programs (LP), assumption-based argumentation (ABA), and AFs with collective attacks (SETAFs), respectively. The CAF normal form is discussed in Section 3.1.*

4.3.1 Normalized CAFs revisited

Let us consider normalized CAFs (cf. Definition 3.1.14). Recall that we run into problems regarding i-semi-stable, i-stage, and i-naive semantics (cf. Example 3.1.13). We show that our hybrid variants that we have developed in this chapter preserve normalization.

Proposition 4.3.1. *Let $\mathcal{F}$ be a CAF, $x \in A_{\mathcal{F}}$ redundant in $\mathcal{F}$ w.r.t. $y \in A_{\mathcal{F}}$, and $\mathcal{F}' = \mathcal{F} \setminus \{x\}$. For $\rho \in \{na_h, pr_h, cf\text{-}stb_h, ad\text{-}stb_h, ss_h, stg_h\}$, it holds that $\rho(\mathcal{F}) = \rho(\mathcal{F}')$.*

Proof. For h-naive and h-preferred semantics, we obtain the result from the correspondence of conflict-free and admissible semantics, respectively (cf. Lemma 3.1.7 and 3.1.8).

To prove the correspondence for the remaining semantics, it suffices to show that the claim-range of conflict-free and admissible sets remains the same. To be more precise, we show that an admissible claim-set S has a realization E that attacks claims C in $\mathcal{F}$ iff S has a realization E' that attacks claims C in $\mathcal{F}'$. First, consider a conflict-free (admissible) claim-set S in $\mathcal{F}$ and consider an arbitrary conflict-free (admissible) realization E of S in $\mathcal{F}$. In case $x \in E$, consider the set $E' = (E \setminus \{x\}) \cup \{y\}$. By Lemma 3.1.6, E' attacks the same arguments in $\mathcal{F}'$ as E does in $\mathcal{F}$, moreover, E' is conflict-free (admissible) in $\mathcal{F}'$ as shown in Lemma 3.1.7 and 3.1.8, respectively. In case $x \notin E$ it holds that E is admissible in $\mathcal{F}'$. If $cl(x) \in cl(E)$ we have that the claim-range of E coincides in $\mathcal{F}$ and $\mathcal{F}'$. Now assume $cl(x) \notin cl(E)$. In case $cl(x) \in E^*_{\mathcal{F}}$ it follows that $cl(x) \in E^*_{\mathcal{F}'}$ (all arguments with claim $cl(x)$ are attacked in $\mathcal{F}'$). In case $cl(x) \notin E^*_{\mathcal{F}}$ we have $y \notin E^+_{\mathcal{F}}$ (since $y_R^- \subseteq x_R^-$) and thus $cl(x) \notin E^*_{\mathcal{F}'}$ as well. The other direction follows since each conflict-free (admissible) set of arguments E in $\mathcal{F}'$ is conflict-free (admissible) in $\mathcal{F}$ by Lemma 3.1.7 and 3.1.8. We thus obtain $\rho(\mathcal{F}) = \rho(\mathcal{F}')$ for $\rho \in \{\textit{cf-stb}_h, \textit{ad-stb}_h, ss_h, stg_h\}$. $\square$

We obtain the following result which extends Theorem 3.1.15.

Theorem 4.3.2. *Each CAF $\mathcal{F}$ can be transformed into a normalized CAF $\mathcal{F}'$ such that $\rho(\mathcal{F}) = \rho(\mathcal{F}')$ for $\rho \in \{na_h, pr_h, \textit{cf-stb}_h, \textit{ad-stb}_h, ss_h, stg_h\}$.*

4.3.2 Logic Programs and Hybrid Semantics

We show that L-stable model semantics correspond to h-semi-stable semantics under Translations 3.2.13 and 3.2.9. To do so, we first prove a result which we obtained 'for free' for the remaining LP semantics when proving that the translations from LPs to CAFs and vice versa are each other's inverse: each LP can be transformed into an atomic LP without changing semantics. Since the translations between CAFs and LPs do not preserve semi-stable (L-stable) semantics, we will give a direct proof showing that such a transformation to an atomic LP is indeed possible while preserving L-stable semantics by showing that so-called *unreachable* atoms are always false and thus can be disregarded.

Given two rules r and s with $head(s) \in body(r)$, we apply *rule-chaining* to obtain the rule r' by replacing the atom $head(s)$ with $body(s)$, i.e., r' is a rule with $head(r') = head(r)$ and $body(r') = (body(r) \setminus head(s)) \cup body(s)$.

Definition 4.3.3. *Let P be a logic program. An atom a in P is called* reachable *in P iff it is possible to construct an atomic rule r from rules in P by successive rule-chaining with $head(r) = a$. Atom a is called* unreachable *in P iff a is not reachable in P.*

Proposition 4.3.4. *Let P be a logic program. It holds that all atoms in T of a 3-valued model $I = (T, F)$ of P are reachable in P.*

Proof. Let $I = (T, F)$ denote a 3-valued model of P and let U denote the set of unreachable atoms in P. We show that there is a 3-valued model $I' = (T', F)$ of P with $T' \subseteq T$ and $T' \cap U = \emptyset$. Since $I' \leq I$, it follows that $I' = I$ and thus T contains no unreachable atoms. We construct I via fixed point iteration:

$$I^0 = (T^0, F) = (T \setminus U, F)$$
$$I^{n+1} = (T^{n+1}, F) = (\{a \in T^n \mid \nexists r \in P/I : (a = head(r) \wedge pos(r) \subseteq T^n)\}$$

Starting with the set of unreachable atoms in P, we remove in each step atoms from T which require atoms outside of T to satisfy condition (a); one could say, we shrink T until we reach a state in which all atoms in T are reachable within T. The procedure has a fixed point (worst case we remove all atoms from T) and is thus guaranteed to terminate. We denote this fixed point by $I' = (T', F)$.

We show that I' is a 3-valued model of P. First observe that I' satisfies condition (b) since (b) is satisfied by I and since the fixed point iteration did not change atoms that are set to false in I. Moreover, I' satisfies condition (a):

($\Rightarrow$): Consider an atom $a \in T'$. That is, a is reachable in P with atoms from T. By construction, there is a rule r in the reduct P/I with $head(r) = a$ and $pos(r) \subseteq T'$, consequently the condition is satisfied.

($\Leftarrow$): Consider an atom $a \in \mathcal{L}(P)$ such that there is a rule $r \in P/I$ with $a = head(r)$ and $pos(r) \subseteq T'$. Since $pos(r) \subseteq T' \subseteq T$ it holds that $a \in T$ (by assumption I is a 3-valued model of P); consequently, $a \in T'$ as required. Thus I' satisfies (a) and (b), moreover, we have $I' \leq I$ by construction. Hence $I' = I$ and T contains no unreachable atoms. □

Next we show that unreachable atoms are always false.

Proposition 4.3.5. *Let P be a logic program and let a denote an atom which is unreachable in P. For all 3-valued models $I = (T, F)$ of P, it holds that $a \in F$.*

Proof. Consider an unreachable atom $a \in \mathcal{L}(P)$ and a 3-valued Herbrand interpretation $I = (T, F)$ with $a \notin F$. By Proposition 4.3.4, it holds that $a \notin T$. Then $I' = (T, F \cup \{a\})$ is a Herbrand interpretation satisfying conditions (a) and (b) in the reduct P/I for all atoms $a \in \mathcal{L}(P)$, moreover, it holds that $I' < I$. Thus I is not a 3-valued model of P. □

We give a fixed-point procedure to generate all rules obtainable from a program.

Definition 4.3.6. *Let P be a logic program. Set $P^0 = P$ and let*

$$\begin{aligned} P^{i+1} = &\{head(s) \leftarrow (body(s) \setminus \{head(r)\}) \cup body(r) \mid r, s \in P^i, head(r) \in body(s)\} \\ &\cup \{r \in P^i \mid r \textit{ is atomic in } P^i\}. \end{aligned}$$

$P^\infty = P^i = P^{i+1}$ for some large enough $i \in \mathbb{N}$ denotes the fixed point of this procedure.

We prove a result that is considered folklore: rule-chaining is a syntactic operation that does not change the semantics of a program.

Proposition 4.3.7. *Let P be a logic program. $I = (T, F)$ denote a 3-valued model of P iff I is a 3-valued model of P^∞.*

Proof. First, we note that the addition of a rule s' which is obtained by replacing the atom $head(r) \in body(s)$ with $body(r)$ for given rules $r, s \in P$ does not affect the semantics.

(1) $I = (T, F)$ is a 3-valued model of P iff I is a 3-valued model of $P' = P \cup \{head(s) \leftarrow (body(s) \setminus \{head(r)\}) \cup body(r)\}$ for rules $r, s \in P$.

Proof of (1). Consider rules $r, s \in P$ with $p = head(r)$ and $p \in body(s)$. Let s' denote the rule $head(s) \leftarrow (body(s) \setminus \{p\}) \cup body(r)$ and let $head(s) = head(s') = a$. First, we observe that $P/I \subseteq P'/I$ (since P' properly extends P by rule s') for any model I of P.

First, consider a 3-valued model $I = (T, F)$ of P. Note that conditions (a) and (b) are satisfied in P'/I for each atom $b \neq a$. It thus suffices to check the conditions for atom a. In case $a \in T$, there is a rule $t \in P/I$ with $head(t) = a$ and $pos(t) \subseteq T$. Since P'/I is a superset of P/I, it holds that $t \in P'/I$. Now assume $a \in F$ and let us assume that (a modified version of) s' is contained in P'/I (otherwise, we are done as $P'/I = P/I$ in this case). Let s'' denote the modified version. It holds that $s'' \in P/I$. Since $a \in F$ we have $pos(s'') \cap F \neq \emptyset$. In case there is some $b \in pos(s'') \cap F$ different from p (i.e., $b \neq p = head(r)$), we are done: in this case, $b \in pos(s')$. Now assume that $p \in pos(s'') \cap F$ is the unique atom contained in the intersection. But then $pos(r) \cap F \neq \emptyset$ since $p \in F$. Consequently, we obtain that $pos(s') \cap F \neq \emptyset$.

For the other direction, let us assume that I is a 3-valued model of P'/I. Again, conditions (a) and (b) are satisfied in P/I for each atom $b \neq a$. Let us now consider the atom a. In case $a \in T$, there is a rule $t \in P'/I$ with $head(t) = a$ and $pos(t) \subseteq T$. In case $t \neq s''$ for s'' being the modified version of s' in the reduct P'/I we are done because then it holds that $t \in P/I$ as well. In case $t = s''$ for s'' being the modified version of s' in the reduct P'/I, it holds that (the modified version of) s serves as witness for $a \in T$ in P/I: indeed, we have $head(s) = a$ and $pos(s) \subseteq pos(s') \subseteq T$. Now assume $a \in F$. That is, for each rule $t \in P'/I$ with $head(t) = a$ we have $pos(t) \cap F \neq \emptyset$. From $P/I \subseteq P'/I$ we obtain that condition (b) is satisfied in P/I as well. ◊

Next, we show that replacing an atom $p \in body(s)$ with the body of *each* rule r_i with $head(r_i) = p$ (yielding a new rule s_i for each such rule r_i) allows for deletion of the rule s.

(2) Given $s \in P$ with $p \in body(s)$, and let $R = \{r_1, \dots, r_m\} \subseteq P$ denote the set of rules with rule head p. For each $i \leq m$, we let s_i denote the rule obtained from replacing p in $body(s)$ with $body(r_i)$, i.e., s_i is of the form $head(s) \leftarrow (body(s) \setminus \{p\}) \cup body(r_i)$. It holds that $I = (T, F)$ is a 3-valued model of P iff I is a 3-valued model of $P' = (P \setminus \{s\}) \cup \{s_1, \dots, s_m\}$.

Proof of (2). From (1) we know that the addition of rules $s_1, \dots, s_m$ to P does not affect the semantics. Let $P^* = P \cup \{s_1, \dots, s_m\}$. Then I is a 3-valued model of P iff I is a 3-valued model of P^*. The programs P' and P^* differ in exactly one rule, namely rule s. Let $head(s) = a$. We show that the deletion of s preserves 3-valued models. Similar as in (1), it suffices to discuss conditions (a) and (b) for atom a.

First, assume $I = (T, F)$ is a 3-valued model of P (and thus of P^*). Observe that $P'/I \subseteq P^*/I$ (in case $T \cap neg(s) = \emptyset$ we have $P'/I = P^*/I$). Let $a \in T$. Then there is a rule $t \in P^*/I$ with $head(t) = a$ and $pos(t) \subseteq T$. Again, we are done in case $t \neq s$ because then $t \in P'/I$ holds. Now assume $t = s$. Then $pos(s) \subseteq T$ and (a modified version of) s is contained in the reduct P^*. That is, $neg(s) \cap T \neq \emptyset$. From $pos(s) \subseteq T$

we obtain $p \in T$. Thus there is a rule $r_i' \in P^*/I$ with $head(r_i') = p$ and $pos(r_i') \subseteq T$ where r_i' is a modified version of rule $r_i \in P^*$ with head p. Thus there is a rule $s_i' \in P'/I$ with $head(s_i') = a$ and $pos(s_i') \subseteq T$ which corresponds to the rule $s_i \in P'$ obtained by replacing $p \in body(s)$ by $body(r_i)$. Consequently, condition (a) is satisfied. In case $a \in F$ it holds that condition (b) is satisfied in P'/I because $P'/I \subseteq P^*/I$.

For the other direction, assume $I = (T, F)$ is a model of P'. Similar as above, in case $a \in T$ we obtain that condition (a) is satisfied in P^*/I because $P'/I \subseteq P^*/I$. Now assume $a \in F$. That is, each rule t with $head(t) = a$ satisfies $pos(t) \cap F \neq \emptyset$. We show that the modified version s' of s in P^*/I satisfies the condition as well. Each s_i' (where s_i' being the modified version of s_i in the reduct P'/I) satisfies condition (b). In case there is $b \in pos(s_i')$ with $b \notin pos(r_i)$ for some $i \leq m$ we are done. In this case, $b \in pos(s')$. Otherwise, it holds that for all rules $r_i' \in P'/I$ with $head(r_i') = p$ there is some $c \in pos(r_i') \cap F$. As $r_i' \in P'/I$ iff $r_i' \in P^*/I$ we obtain $p \in F$. Consequently, $pos(s') \cap F \neq \emptyset$ and we obtain that condition (b) is satisfied. ◊

Given P^i we obtain P^{i+1} as follows: for each rule $s \in P^i$, for each $p \in pos(s)$, we replace s with the set of rules obtained by replacing p with the body of all rules in P^i with head p. In case s is atomic we add it to P^{i+1}. As shown in (2), replacing rules does not change the 3-valued models of a program. □

Reachability can be alternatively defined via P^∞: An atom a is reachable if there exists an atomic rule $r \in P^\infty$ with $head(r) = a$. Rules in P^∞ which are not atomic can be deleted without changing the semantics in case each atom in P^∞ is reachable. Intuitively, such rules do not carry any additional information which has not been incorporated yet. Recall that unreachable atoms are set to false. We obtain the following result.

Proposition 4.3.8. *For each logic program P with unreachable atoms $U \subseteq \mathcal{L}(P)$, there exists an atomic program P' such that $I' = (T, F)$ is a 3-valued model of P' iff $I = (T, F \cup U)$ is a 3-valued model of P.*

As discussed in Section 3.2, L-stable semantics cannot be captured via established AF semantics that operate exclusively on argument-level. Having formally defined our claim-sensitive version of semi-stable semantics, we have successfully identified a semantics for CAFs that matches L-stable model semantics, as the following result demonstrates.

Proposition 4.3.9. *Let P be a logic program, $\mathcal{F}_P$ the associated CAF, and $I = (T, F)$ be a 3-valued interpretation. Then I is L-stable in P iff $T \in ss_h(\mathcal{F}_P)$.*

Proof. By Proposition 4.3.5, it suffices to consider LPs without unreachable atoms: indeed, if atom a is unreachable, then we have that $a \in F$ for each model $I = (T, F)$. Removing unreachable atoms therefore does not change $\subseteq$-maximality of $T \cup F$. Now, consider a logic program P without unreachable atoms. Notice that the corresponding CAF $\mathcal{F}_P$ contains (at least) one argument for each atom in P. By Proposition 3.2.11, we have $T \in co(\mathcal{F}_P)$ iff $I = (T, F)$ is p-stable in P. We obtain the correspondence of L-stable semantics with h-semi-stable semantics by observing that defeated claims (in $\mathcal{F}_P$) correspond to (reachable) atoms that are set so false (in P).

By Proposition 4.3.7, we obtain that moving from P to P^∞ does not change the semantics of P, i.e., I is a 3-valued model of P iff I is a 3-valued model of P^∞. It thus suffices to show $F = T^*_{\mathcal{F}_P}$ for all p-stable models $I = (T, F)$ of P^∞. By assumption each atom is reachable we observe that each rule in P^∞ is atomic. As each atomic rule induces exactly one argument, there is a one-to-one correspondence between the arguments constructed from P and the rules in P^∞.

Let $I = (T, F)$ denote a 3-valued model of P.

First, we show that all arguments in the corresponding CAF $\mathcal{F}_P$ with claims in F are attacked by T. Consider some $p \in F$ and let r denote a rule of P^∞ with $head(r) = p$. The rule r is of the form $p \leftarrow \texttt{not}\ b_1, \dots, \texttt{not}\ b_m$. Since $p \in F$ and since $pos(r) = \emptyset$ it holds that $T \cap neg(r) \neq \emptyset$. By definition of an argument in $\mathcal{F}_P$, each $b \in neg(r)$ is a vulnerability of A, i.e., $b \in \textsc{Vul}(A)$. By definition of the attack relation, it holds that each argument with claim b attacks A.

For the other direction, consider some claim p that is attacked by T in $\mathcal{F}_P$. That is, for each argument A with claim p, it holds that $\textsc{Vul}(A) \cap T \neq \emptyset$. Thus for each rule r with $head(r) = p$, it holds that $T \cap neg(r) \neq \emptyset$. Consequently, P^∞ does not contain rules with head p. It follows that $p \in F$. □

For the other direction, we utilize Theorem 4.3.2 to transform a given well-formed CAF $\mathcal{F}$ into a well-formed CAF $\mathcal{F}'$ without copies. By Proposition 3.2.15, Translation 3.2.13 and 3.2.9 are each others inverse on the fragment of copy-free well-formed CAFs (up to isomorphism). Hence a claim-set S is h-semi-stable in $\mathcal{F}$ iff it is h-semi-stable in $\mathcal{F}'_{P_{\mathcal{F}'}}$ iff S is L-stable in $P_{\mathcal{F}'}$. Hence Translation 3.2.13 preserves h-semi-stable semantics.

Proposition 4.3.10. *For each well-formed CAF $\mathcal{F}$, it holds that $S \in ss_h(\mathcal{F})$ iff S is L-stable in $P_{\mathcal{F}}$.*

4.3.3 Assumption-based Argumentation and Hybrid Semantics

In this section, we discuss how h-semi-stable semantics and semi-stable semantics for ABA relate to each other. We show that Translation 3.3.18 and 3.3.22 preserve semi-stable semantics for each ABA framework $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, {}^-)$ satisfying $head(r) \in \overline{\mathcal{A}}$ in case $body(r) \neq \emptyset$ for all $r \in \mathcal{R}$.

Proposition 4.3.11. *For each ABA framework $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, {}^-)$ satisfying $head(r) \in \overline{\mathcal{A}}$ for all $r \in \mathcal{R}$ it holds that $ss_{Th}(D) = ss_h(\mathcal{F}_D)$.*

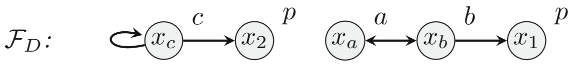
Proof. Consider a set $S \in ss_{Th}(D)$, let $\mathcal{S}$ denote its corresponding complete assumption-set, and E the complete set of arguments in $\mathcal{F}_D$ corresponding to S. Recall that complete sets of arguments and assumptions are in one-to-one correspondence to each other.

Let us start with some general observations: First, it holds that E and $\mathcal{S}$ defeat the same assumptions. Indeed, $\mathcal{S}$ attacks assumption $a \in \mathcal{A}$ iff $\mathcal{S}$ derives $\overline{a}$ implying that $\overline{a} \in S$, i.e., the unique argument with claim a is defeated in $\mathcal{F}_D$ by E. Thus it holds that $\mathcal{S} \cup \mathcal{S}_D^+ \subseteq E^{\circledast}_{\mathcal{F}_D}$. Second, it holds that E defeats conclusion $\overline{a}$ (i.e., the contrary of some assumption a) in $\mathcal{F}_D$ iff $\mathcal{S}$ attacks all assumption-sets which derive $\overline{a}$ in D. Indeed, E defeats claim $\overline{a}$ iff E defends assumption a which is the case iff a contained in $\mathcal{S}$, meaning that $\mathcal{S}$ indeed attacks each assumption-set which derives $\overline{a}$.

To sum up, we have shown that E defeats a conclusion p which is either assumption or contrary if and only if $\mathcal{S}$ attacks all assumption-sets deriving p. By our assumption $head(r) \in \overline{\mathcal{A}}$ for all $r \in \mathcal{R}$, the ABA framework D does not contain other conclusions. Thus we obtain that $ss_{Th}(D) = ss_h(\mathcal{F}_D)$. ☐

We present an example showing that the restriction to ABA frameworks which derive only facts or contraries of assumptions is necessary.

Example 4.3.12. *Consider the ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{})$ *with* $\mathcal{A} = \{a, b, c\}$ *with* $\overline{a} = b$, $\overline{b} = a$, $\overline{c} = c$ *and rules* $r_1 : p \leftarrow a$, $r_2 : p \leftarrow c$. *The attacks between the assumption-sets (we depict only singletons) and the resulting CAF* $\mathcal{F}_D$ *are depicted below:*



In D, *both* $\{a\}$ *and* $\{b\}$ *are semi-stable. The conclusion-extensions are thus* $\{a, p\}$ *and* $\{b\}$. *In* $\mathcal{F}_D$, *however, only the set* $\{a, p\}$ *is h-semi-stable since* $\{x_a, x_1\}$ *has maximal range (*x_b *does not defeat claim* p*). This error does not stem from the fact that assumptions and contraries are not separated in* D *(applying Translation 3.3.26 yields the same mismatch).*

Next we show that h-semi-stable semantics are also preserved by Translation 3.3.22. For this, we show that it is indeed possible to remove assumption-arguments from $\mathcal{F}_D$ without changing h-semi-stable semantics if the contraries in the framework D are separated.

Proposition 4.3.13. *For an ABA framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{})$ *with separated contraries and* $head(r) \in \overline{\mathcal{A}}$ *for all* $r \in \mathcal{R}$, *we have* $ss_{pTh}(D) = ss_h(\mathcal{F}_{D\text{-}\mathcal{A}}) = \{S \setminus \mathcal{A} \mid S \in ss_h(\mathcal{F}_D)\}$.

Proof. Consider an ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{})$ with separated contraries satisfying $body(r) \neq \emptyset$ for all $r \in \mathcal{R}$ and an assumption $a \in \mathcal{A}$. We show that $ss_h(\mathcal{F}_D \setminus \{a\}) = \{S \setminus \{a\} \mid S \in ss_h(\mathcal{F}_D)\}$ (iterative application of this claim yields the desired result).

We proceed by case distinction: (a) $\overline{a} \notin cl(\mathcal{F}_D)$ and (b) $\overline{a} \in cl(\mathcal{F}_D)$.

First assume $\overline{a}$ does not exist in $\mathcal{F}_D$. In this case, a is unattacked in $\mathcal{F}_D$ and thus a is contained in each (complete, preferred, grounded) claim-set of $\mathcal{F}_D$, i.e., a is contained in the range of each preferred claim-set. Removing a from $\mathcal{F}_D$ thus does not change the (in-)comparability of the range of preferred claim-sets (since a is removed from each claim-set). Consequently, it holds that the removal of a does not change h-semi-stable semantics. We obtain $ss_h(\mathcal{F}_D \setminus \{a\}) = \{S \setminus \{a\} \mid S \in ss_h(\mathcal{F}_D)\}$.

Now assume $\overline{a} \in cl(\mathcal{F}_D)$. We show that $a \in S^{\circledast}_{\mathcal{F}_D}$ iff $\overline{a} \in S^{\circledast}_{\mathcal{F}_D}$ for each $S \in ss_h(\mathcal{F}_D)$:

First assume $a \in S^{\circledast}_{\mathcal{F}_D}$. In case S defeats a in $\mathcal{F}_D$ it holds that $\overline{a} \in S$ (arguments having claim $\overline{a}$ are the only attacker of a). In case $a \in S$ it must be the case that each occurrence of $\overline{a}$ is attacked (since S must be admissible realizable), thus $\overline{a} \in S^{\circledast}_{\mathcal{F}_D}$. Now, in case $\overline{a} \in S$, we have that a is attacked by S in $\mathcal{F}_D$ and thus $a \in S^{\circledast}_{\mathcal{F}_D}$. In case $\overline{a}$ is defeated by S, we have that each attacker of a is attacked by S, moreover, the argument x_a corresponding to a has no outgoing attacks and thus it holds that $S \cup \{a\}$ are admissible realizable in $\mathcal{F}_D$. We conclude that $a \in S$ (by maximality of the range).

We have shown that $a \in S^{\circledast}_{\mathcal{F}_D}$ iff $\overline{a} \in S^{\circledast}_{\mathcal{F}_D}$. Intuitively, this means that each assumption a has a witness ($\overline{a}$) which remains in the CAF after removing a. Now, given two claim-sets S and T which are h-semi-stable in $\mathcal{F}_D$. It is easy to show that that the fact that a and $\overline{a}$ only come in pairs implies that the range of S and T are not in $\subseteq$-relation after removing a: Otherwise (w.l.o.g. assume $S^{\circledast}_{\mathcal{F}_D \setminus \{a\}} \subset T^{\circledast}_{\mathcal{F}_D \setminus \{a\}}$) it must be the case that a is the separating element which is contained in the range of S in $\mathcal{F}_D$ but not in the range of T in $\mathcal{F}_D$. But then it holds also that $\overline{a} \in S^{\circledast}_{\mathcal{F}_D}$. Since the range of S is properly contained in the range of T in $\mathcal{F}_D \setminus \{a\}$ it must be the case that $\overline{a}$ is contained in the range of T in $\mathcal{F}_D$ as well. This in turn implies that a is contained in the range of T in $\mathcal{F}_D$, contradiction to the assumption. The other direction follows since the addition of an argument which does not attack any other argument does not change the incomparability of the range of two claim-sets. We obtain $ss_h(\mathcal{F}_D \setminus \{a\}) = \{S \setminus \{a\} \mid S \in ss_h(\mathcal{F}_D)\}$. □

Similar as for LPs, we utilize Theorem 4.3.2 to obtain the following result.

Proposition 4.3.14. *For each well-formed CAF $\mathcal{F}$ it holds that $ss_h(\mathcal{F}) = ss_{pTh}(D_{\mathcal{F}})$.*

4.3.4 SETAFs and Hybrid Semantics

We end this overview with a brief discussion of SETAFs in connection with hybrid semantics. We will show that h-semi-stable, h-stage, and h-naive semantics correspond to semi-stable, stage, and naive SETAF semantics, respectively.

Proposition 4.3.15. *For each well-formed CAF $\mathcal{F}$, its associated SETAF $SF_{\mathcal{F}}$, and semantics $\sigma \in \{na, ss, stg\}$, it holds that $\sigma_h(\mathcal{F}) = \sigma(SF_{\mathcal{F}})$.*

Proof. By Proposition 3.4.7, admissible and conflict-free sets of $\mathcal{F}$ and $SF_{\mathcal{F}}$ coincide. Moreover, a set $S \in cl(A) = A_{SF_{\mathcal{F}}}$ of arguments/claims attacks c in $SF_{\mathcal{F}}$ iff S attacks each occurrence of c in $\mathcal{F}$. Hence the claim-range of a set $S \in ad_i(\mathcal{F}) = ad(SF_{\mathcal{F}})$ ($S \in cf_i(\mathcal{F}) = cf(SF_{\mathcal{F}})$, respectively) in $\mathcal{F}$ corresponds to the range S in $SF_{\mathcal{F}}$. Thus h-semi-stable (h-stage) semantics correspond to semi-stable (stage) SETAF semantics. For naive semantics, the correspondence follows from $cf_i(\mathcal{F}) = cf(SF_{\mathcal{F}})$. □

Corollary 4.3.16. *For each well-formed CAF* $\mathcal{F}$ *and semantics* $\sigma \in \{na, ss, stg\}$*, it holds that* $\sigma(SF) = \sigma_h(\mathcal{F}_{SF})$*.*

4.4 Conclusion

In this chapter, we introduced a novel class of CAF semantics by shifting maximization and defeat to claim-level. Our hybrid approach gave rise to novel variants of *naive*, *preferred*, *stable*, *stage*, and *semi-stable semantics*. We settled the relation between the semantics in Sections 4.1 and 4.2. We showed that for well-formed CAFs, stable and preferred variants coincide, while naive, stage, and semi-stable variants differ. The latter highlights the fundamental difference between claim-set maximization on claim- and on argument-level in particular for range-based semantics.

We observe several advantages of hybrid semantics which will be discussed below.

First, hybrid semantics constitute an argumentation-based formalization of semantics from conclusion-focused knowledge representation formalisms, as we have discussed in Section 4.3. In particular, hybrid semi-stable semantics play an important role in this matter: we have shown that hybrid semi-stable semantics correspond to semi-stable semantics for SETAFs and assumption-based argumentation frameworks and capture L-stable semantics for logic programs. Let us point out that the latter is—under standard instantiation methods—impossible for Dung AFs without claims. In this way, we significantly deepen the close connection of logic programming semantics and argumentation semantics. For SETAFs, we furthermore show that h-naive and h-stage CAF semantics serve as counter-parts for naive and stage SETAF semantics.

Second, hybrid semantics provide an alternative view on claim justification in the spirit of abstract argumentation semantics. Adapting our hybrid semantics to ASPIC, for instance, leads to notable simplifications when identifying acceptable conclusions and even to novel evaluation aspects that take defeated conclusions into account. As we have seen, instantiated ASPIC frameworks are not necessarily well-formed. Finding a $\subseteq$-maximal set of jointly acceptable conclusions is thus not necessarily the same as finding jointly acceptable arguments, even with respect to preferred semantics (cf. Example 4.1.2). Here, our hybrid semantics fill in the gap: h-preferred semantics are guaranteed to return $\subseteq$-maximal sets of claims by design. Moreover, having a notion of defeated claims yields a novel perspective: it provides means to take defeated claims into account. This can be useful one aims to refute a particular statement.

Our novel semantics incorporate evaluation methods which are common to conclusion-focused knowledge representation formalisms. Moreover, they yield a novel perspective to argumentation semantics by putting the focus on claim acceptance (via claim-set maximization and claim-defeat). With this, we hope to broaden the argumentation semantics landscape and to increase the flexibility of the abstract model to capture even more potential use cases.

CHAPTER 5

Principles and Expressiveness

Argumentation semantics differ in their characteristics. Within the last decades, numerous semantics have been proposed in the literature [16]. Due to the sheer number of semantics it is often not easy to choose the 'correct' semantics, i.e., the semantics that fits best in a particular scenario. A central topic of research in argumentation is the development of tools that help to make this decision. The *principle-based methodology* [11, 175] is well-suited for a systematic analysis of semantics: such a classification yields theoretical insights into the nature of the semantics on the one hand and guides the search for suitable semantics appropriate in different scenarios on the other hand. Moreover, knowing the *expressiveness* of a semantics is central to decide whether a semantics is capable to appropriately model a particular setting. The characterization of the so-called *signature* [84] of a semantics, i.e., the set of all possible extension-sets a framework can possess under the given semantics, is key to understand its expressive power. Apart of the theoretical insights, knowing which extensions can jointly be modeled within a single framework under a given semantics is for instance crucial in dynamic scenarios in which argumentation frameworks undergo certain changes [27].

This chapter is concerned with the characteristics of CAF semantics. We present a systematic analysis of claim semantics with respect to general and to well-formed CAFs by investigating and comparing fundamental properties. We identify, adapt, and develop fundamental principles for claim semantics in Section 5.1. In Section 5.2, we study the expressive power of the semantics by characterizing their signatures.

Background & Notation. *In this chapter, we make use of concepts and results from Chapter 4 where we have developed* hybrid semantics (h-semantics) *for CAFs. Let us recall our naming conventions. We sometimes drop 'inherited' or 'hybrid' (prefix 'i-' or 'h-', respectively) when speaking about a semantics for which only one version exists or if both variants coincide. For instance, we say 'complete' instead of 'i-complete' semantics because it has no hybrid variant; and for well-formed CAFs, we say 'stable' instead of 'inherited', 'h-cf-', or 'h-ad-stable' semantics because all variants coincide.*

5.1 Principles

Inspired by similar studies on AF semantics, we conduct a *principle-based analysis* of CAF semantics in this section. The goal of our studies is to identify differences between inherited and hybrid semantics on the one hand and to analyze the different behavior of the semantics when restricted to well-formed CAFs when compared to the general case on the other hand. We have already experienced in Section 4.1 that differences between inherited and hybrid semantics vanish when restricting them to well-formed CAFs (cf. Proposition 4.1.5 and 4.1.16). Our principle-based analysis aims to work out such specific differences in greater detail. We consider principles restricted to a CAF-class $\mathfrak{C}$.

Definition 5.1.1. *By* $\mathfrak{C}_u$ *and* $\mathfrak{C}_{wf}$*, we denote the class of CAFs resp. well-formed CAFs.*

In this section, we identify not only principles that are genuine for CAF semantics, but consider also principles that extend well-known principles for AF semantics to claim-focused reasoning. In this aspect, let us recall that AFs can be seen as a special case of CAFs by taking the identity function as claim-function. By Lemma 4.2.1, negative results carry over to CAFs for those principles that are a faithful generalization of AF principles. To compare our principles with the corresponding AF case, it will be useful to consider the CAF class that contains each AF as equivalent CAF representation.

Definition 5.1.2. *We define the CAF class* $\mathfrak{C}_{id} = \{(F, id) \mid F \text{ is an } AF\}$*.*

We subdivide our principles in three different groups: in Section 5.1.1, we consider principles that address properties of the underlying structure of the framework with respect to specific semantics; in Section 5.1.2, we consider basic properties following similar investigations for AFs; finally, we study set-theoretical principles in Section 5.1.3 and set the grounds for a rigorous expressiveness analysis of our semantics under consideration.

5.1.1 Meta-principles

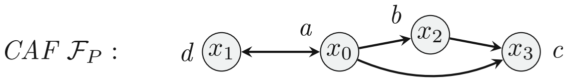
Let us start our principle-based analysis with a fundamental principle of claim-focused reasoning: the *realizability principle* states that a claim-set requires a set of arguments that supports it in order to be acceptable in a given framework.

Principle 5.1.3 (Realizability)**.** *A semantics* ρ *satisfies the* realizability principle *in* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$*, for every claim-set* S*,* $S \in \rho(\mathcal{F})$ *only if there is a set of arguments* $E \subseteq A_{\mathcal{F}}$ *that realizes* S *in* $\mathcal{F}$*.*

The realizability principle is at the core of argumentative claim justification: a claim cannot be accepted if there is no argument for it. By definition, each semantics under consideration satisfies this fundamental principle.

The next principle we consider is common to many claim-focused reasoning formalisms: the *argument-name independence principle* states that the specific names of the arguments do not play a role when evaluating a given framework with respect to the claims.

Example 5.1.4. *Let us recall our logic program P from Example 3.2.5 and the associated CAF $\mathcal{F}_P$ obtained from Translation 3.2.9. We depict $\mathcal{F}_P$ below.*



The CAF $\mathcal{F}_P$ has four arguments x_0, x_1, x_2, x_3. In the claim-focused evaluation of the CAF, however, the particular names of the arguments are irrelevant. More precisely, it would have been equally possible to name them x, y, z, u. Evaluating the resulting CAF with respect to complete semantics yields in both cases the claim-sets $\emptyset$, $\{a\}$, and $\{d, b\}$.

Principle 5.1.5 (Argument-names independence)**.** *A semantics ρ satisfies the* argument-names independence principle *in $\mathfrak{C}$ iff for every two CAFs $\mathcal{F}$ and $\mathcal{G}$ in $\mathfrak{C}$ which are isomorphic to each other, it holds that $\rho(\mathcal{F}) = \rho(\mathcal{G})$.*

It is easy to see that all considered CAF semantics satisfy this principle.

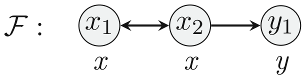
Remark 5.1.6 (Relation to AFs)**.** *The adaption of argument-name independence to AFs by restricting it to the class $\mathfrak{C}_{id}$ yields a principle that allows to compare only identical AFs (due to Definition 2.2.8 of CAF isomorphisms) and is thus trivially satisfied by all possible semantics. An alternative adaption of the principle is to consider classical isomorphisms disregarding the labels (graph-theoretically speaking, an arc-preserving bijection). In this case, the principle is not satisfied by any non-trivial AF semantics since the names of the arguments trivially matter when evaluating AFs.*

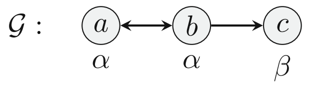
Next we discuss a principle that seems closely related at first sight: the *language independence principle* [11, 175] or *abstraction* [7, 45], states that a semantics is independent of the specific names of the elements that occur in a framework.

In contrast to argument-name independence, which states that two isomorphic frameworks yield *identical* claim-extensions independently of the considered argument-names, the language independence principle states that the evaluation process does not depend on the names of the abstract objects (i.e., arguments and claims) in the frameworks. For an appropriate adaption to CAFs, we consider generalized isomorphisms that preserve the claim-structure but not their specific names (speaking in graph-theoretical terms, we consider an arc-preserving vertex bijection which preserves equivalence classes of labels).

Definition 5.1.7. *A bijective function $f : A_\mathcal{F} \to A_\mathcal{G}$ between two CAFs $\mathcal{F}$ and $\mathcal{G}$ is a* generalized isomorphism *if f is attack-preserving i.e., $(x, y) \in R_F$ iff $(f(x), f(y)) \in R_G$ for all $a, b \in A_\mathcal{F}$, and preserves the claim-structure, i.e., $cl(x) = cl(y)$ iff $cl(f(x)) = cl(f(y))$ for all $x, y \in A_\mathcal{F}$. We say that $\mathcal{F}$ and $\mathcal{G}$ are generalized isomorphic to each other iff there is a generalized isomorphism $f : A_\mathcal{F} \to A_\mathcal{G}$. We call the function $f_c : cl(A_\mathcal{F}) \to cl(A_\mathcal{G})$ with $f_c(cl(x)) = cl(f(x))$* f-induced claim-isomorphism*.*

Example 5.1.8. *Let us consider our CAF $\mathcal{F}$ from Example 2.2.7 and another CAF $\mathcal{G}$ also having three arguments. Both $\mathcal{F}$ and $\mathcal{G}$ are depicted below:*





The CAFs $\mathcal{F}$ and $\mathcal{G}$ are not isomorphic to each other since their claims differ. They are, however, generalized isomorphic to each other: indeed, the function f with $x_1 \mapsto a$, $x_2 \mapsto b$, and $y_1 \mapsto c$ satisfies $(x, y) \in R_{\mathcal{F}}$ iff $(f(x), f(y)) \in R_{\mathcal{G}}$ and preserves the claim-structure by associating claim x in $\mathcal{F}$ with claim α in $\mathcal{G}$ and claim y with claim β. The induced claim-isomorphism f_c behaves accordingly and maps x to α and y to β.

Principle 5.1.9 (Language independence)**.** *A semantics ρ satisfies the* language independence principle *in $\mathfrak{C}$ iff for every two CAFs $\mathcal{F}$ and $\mathcal{G}$ in $\mathfrak{C}$ which are generalized isomorphic to each other (via isomorphism f), it holds that $\rho(\mathcal{F}) = \{f_c(S) \mid S \in \rho(\mathcal{G})\}$ for the f-induced claim-isomorphism $f_c : cl(A_{\mathcal{F}}) \to cl(A_{\mathcal{G}})$.*

Language independence is a faithful adaption of the corresponding AF principle: each generalized isomorphism between $\mathcal{F}, \mathcal{G} \in \mathfrak{C}_{id}$ corresponds to an isomorphism between F and G. We note that all considered semantics satisfy this principle.

Next we consider another principle that is genuine for CAFs. The *unique realizability principle* gives insights into the correspondence of claim-sets and their realizations.

Principle 5.1.10 (Unique realizability)**.** *A semantics ρ satisfies the* unique realizability principle *in $\mathfrak{C}$ iff for every CAF $\mathcal{F} \in \mathfrak{C}$, for every $S \in \rho(\mathcal{F})$ there is a unique set of arguments $E \subseteq A_{\mathcal{F}}$ that ρ-realizes S in $\mathcal{F}$.*

This principle is satisfied by most of the semantics for well-formed CAFs.

Proposition 5.1.11 (cf. Proposition 3.1.2)**.** *Grounded, complete, i-preferred, i-semi-stable, i-naive, i-stage, and i-stable semantics satisfy unique realizability in $\mathfrak{C}_{wf}$.*

Interestingly, hybrid semantics are not uniquely realized as they do not require $\subseteq$-maximality of their admissible (or conflict-free) realizations. Consider the following trivial example with only two arguments both having the same claim c.

Example 5.1.12. *Consider the well-formed CAF $\mathcal{F} = (\{x, y\}, \emptyset, cl)$ with $cl(x) = cl(y) = c$. In $\mathcal{F}$, all hybrid semantics σ_h return the same claim-set $\{c\}$. However, $\{c\}$ has three possible σ_h-realizations: $\{x\}$, $\{y\}$, and $\{x, y\}$. Thus, $\{c\}$ is not uniquely realized in $\mathcal{F}$.*

Note that the alternative definitions of h-semantics that consider complete, preferred, or naive extensions instead of admissible or conflict-free extensions on argument-level (cf.

Propositions 4.1.4, 4.1.17, and 4.1.20) satisfy unique realizability since the property is transferred from the underlying inherited semantics.

This observation is crucial for the following weaker version of unique realizability: *maximal realizability* requires that each extension has a unique $\subseteq$-maximal realization.

Principle 5.1.13 (Maximal realizability)**.** *A semantics ρ satisfies the* maximal realizability principle *in $\mathfrak{C}$ iff for every CAF $\mathcal{F} \in \mathfrak{C}$, for every $S \in \rho(\mathcal{F})$, the set $E^{\max} = \bigcup_{E\ \rho\text{-real. } S} E$ is a ρ-realization of S in $\mathcal{F}$.*

Proposition 5.1.14. *All considered semantics satisfy maximal realizability in in $\mathfrak{C}_{wf}$.*

Proof. Starting with inherited conflict-free and admissible semantics, we first observe that two cf_i-realizations E, D of a claim-set S are conflict-free since they attack the same arguments, thus $E \cup D$ cf_i-realizes S as well. Moreover, if E and D are ad_i-realizations of S, it holds that both defend the same arguments, thus $E \cup D$ ad_i-realizes S. We thus obtain that i-conflict-free and admissible semantics satisfy maximal realizability. The inherited semantics in question satisfy the principle since they build on either i-conflict-free or i-admissible semantics (and since they already satisfy unique realizability).

For h-preferred and both variants of stable semantics, the statement follows since they coincide with their respective inherited counter-parts. For the remaining semantics, it suffices to consider the i-preferred (for h-semi-stable semantics) respectively the i-naive (for h-naive and h-stage semantics) realization of the claim-set in question: Consider a well-formed CAF $\mathcal{F}$ and let S denote a h-semi-stable claim-set of $\mathcal{F}$. By our results from Section 4.1, S has a pr_i-realization E in $\mathcal{F}$. This realization contains all ss_h-realizations of S in $\mathcal{F}$, i.e., $E = E^{\max}$. The proof for h-naive and h-stage semantics is analogous. ☐

Apart from grounded semantics, all considered semantics violate unique and maximal realizability in the general case. It suffices to extend Example 5.1.12 in a minimal way:

Example 5.1.15. *Consider the CAF $\mathcal{F} = (\{x, y\}, \{(x, y), (y, x)\}, cl)$ with $cl(x) = cl(y) = c$. In $\mathcal{F}$, all semantics return the claim-set $\{c\}$. However, the extension $\{c\}$ has two possible realizations $\{x\}$ and $\{y\}$ which shows that $\{c\}$ is neither uniquely realizable nor possesses a maximal realization.*

Table 5.1 and 5.2 summarize our results from this section. Table 5.1 presents all considered principles for general CAFs while Table 5.2 contains all principles for well-formed CAFs. The realizability principle as well as the argument-name and language independence principle are satisfied by all considered semantics, which confirms that these principles formalize fundamental properties of claim-focused reasoning. On the other hand, we observe that the desirable unique and maximal realizability principles are not satisfied by any (except the single-status grounded) semantics in the general case. For well-formed CAFs, the picture is more diverse, in particular due to the difference between inherited and hybrid semantics regarding unique realizability. Maximal realizability on the other hand is satisfied by all considered semantics.

	Realizability	Arg-name Ind.	Language Ind.	Unique Realizability	Maximal Realizability
cf_i	✓	✓	✓	✗	✗
ad_i	✓	✓	✓	✗	✗
gr_i	✓	✓	✓	✓	✓
co_i	✓	✓	✓	✗	✗
pr_i	✓	✓	✓	✗	✗
pr_h	✓	✓	✓	✗	✗
stb_i	✓	✓	✓	✗	✗
$cf\text{-}stb_h$	✓	✓	✓	✗	✗
$ad\text{-}stb_h$	✓	✓	✓	✗	✗
ss_i	✓	✓	✓	✗	✗
ss_h	✓	✓	✓	✗	✗
na_i	✓	✓	✓	✗	✗
na_h	✓	✓	✓	✗	✗
stg_i	✓	✓	✓	✗	✗
stg_h	✓	✓	✓	✗	✗

Table 5.1: Meta-principles w.r.t. general CAFs.

	Realizability	Arg-name Ind.	Language Ind.	Unique Realizability	Maximal Realizability
cf_i	✓	✓	✓	✗	✓
ad_i	✓	✓	✓	✗	✓
gr_i	✓	✓	✓	✓	✓
co_i	✓	✓	✓	✓	✓
pr_i	✓	✓	✓	✓	✓
pr_h	✓	✓	✓	✗	✓
stb_i	✓	✓	✓	✓	✓
$cf\text{-}stb_h$	✓	✓	✓	✗	✓
$ad\text{-}stb_h$	✓	✓	✓	✗	✓
ss_i	✓	✓	✓	✓	✓
ss_h	✓	✓	✓	✗	✓
na_i	✓	✓	✓	✓	✓
na_h	✓	✓	✓	✗	✓
stg_i	✓	✓	✓	✓	✓
stg_h	✓	✓	✓	✗	✓

Table 5.2: Meta-principles w.r.t. well-formed CAFs.

5.1.2 Basic Principles

In this section, we investigate fundamental properties of argumentation semantics in the context of claim-focused reasoning. To begin with, we study claim semantics on argument-level by analyzing the corresponding realizations.

Principle 5.1.16 (Conflict-freeness)**.** *A semantics ρ satisfies* conflict-freeness *in $\mathfrak{C}$ iff for each CAF $\mathcal{F} \in \mathfrak{C}$, for each $S \in \rho(\mathcal{F})$, there is a conflict-free realization E of S in $\mathcal{F}$.*

Principle 5.1.17 (Defense)**.** *A semantics ρ satisfies* defense *in $\mathfrak{C}$ iff for each CAF $\mathcal{F} \in \mathfrak{C}$, for each $S \in \rho(\mathcal{F})$, there is a realization E of S in $\mathcal{F}$ that defends itself.*

Principle 5.1.18 (Admissibility)**.** *A semantics ρ satisfies* admissibility *in $\mathfrak{C}$ iff for each CAF $\mathcal{F} \in \mathfrak{C}$, for each $S \in \rho(\mathcal{F})$, there is an admissible realization E of S in $\mathcal{F}$.*

The principles conflict-freeness, admissibility, and defense faithfully generalize the corresponding AF principles. These principles all formalize properties of extensions. We have generalized these principles using the following schema.

Schema 5.1.19. *Let $\mathcal{P}$ be an AF principle of the form*

"Semantics σ satisfies $\mathcal{P}$ iff for all AFs F, for all $E \in \sigma(F)$, E satisfies property p."

We adapt the principle to claim semantics as follows:

"Semantics ρ satisfies $\mathcal{P}_i$ iff for all $\mathcal{F} \in \mathfrak{C}$, for all $S \in \rho(\mathcal{F})$, there exists a realization E of S in $\mathcal{F}$ such that E satisfies property p".

We call principles obtained in this way *inherited principles*. By definition, each inherited semantics σ_i satisfies principle $\mathcal{P}_i$ iff the corresponding AF semantics σ satisfies $\mathcal{P}$. likewise, the hybrid semantics inherit satisfaction from their realizations. From the relations between the semantics established in Section 4.2, we obtain the following results.

Lemma 5.1.20. *Let $\mathcal{P}_i$ be the generalization of principle $\mathcal{P}$ obtained by Schema 5.1.19. A semantics σ_i satisfies $\mathcal{P}_i$ in $\mathfrak{C}_u$ iff σ satisfies $\mathcal{P}$; moreover, if preferred semantics satisfies $\mathcal{P}$ then $\rho \in \{pr_h, ss_h, ad\text{-}stb_h\}$ satisfies $\mathcal{P}_i$ in $\mathfrak{C}_u$; and if naive semantics satisfies $\mathcal{P}$ then $\rho \in \{na_h, stg_h, cf\text{-}stb_h\}$ satisfies $\mathcal{P}_i$ in $\mathfrak{C}_u$.*

For our considered principles, we obtain the following results.

Proposition 5.1.21. *All semantics under consideration satisfy conflict-freeness in $\mathfrak{C}_u$. Admissible, complete, grounded, h-ad-stable, i-stable and both variants of semi-stable and preferred semantics satisfy defense and admissibility in $\mathfrak{C}_u$.*

We note that conflict-freeness and defense are properties of argument-extensions rather than claim-sets. In this context, we recall that claim-defense corresponds to defense on argument-level (cf. Section 4.1.4). Hence the shift to the extensions when analyzing these properties is reasonable. There are, however, principles for which the generalization is less straightforward. In the following, we consider the naivety principle and (CF-)reinstatement and discuss different generalizations.

The *naivety principle* has been introduced in [175] for AFs. In the context of claims, this principle can be extended in two ways: First, by requiring the existence of a realization that is maximal with respect to set-inclusion, and second, by requiring that the claim-set itself is $\subseteq$-maximal. Notice that these two natural choices reflect the different approaches that underlie inherited and hybrid semantics, respectively.

Principle 5.1.22 (i-Naivety)**.** *A semantics* ρ *satisfies the* inherited naivety principle *in* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$*, for every* $S \in \rho(\mathcal{F})$*, there is a conflict-free realization* E *of* S *in* $\mathcal{F}$ *which is* $\subseteq$*-maximal in* $cf(F)$*.*

Principle 5.1.23 (h-Naivety)**.** *A semantics* ρ *satisfies the* hybrid naivety principle *in* $\mathfrak{C}$ *iff for each CAF* $\mathcal{F} \in \mathfrak{C}$*, for each* $S \in \rho(\mathcal{F})$*, it holds that* S *is* $\subseteq$*-maximal in* $cf_i(\mathcal{F})$*.*

Both principles faithfully generalize the AF naivety principle. By Lemma 4.2.1, we thus obtain counter-examples for admissible, complete, grounded, preferred, and semi-stable semantics. By Lemma 5.1.20, we obtain the following result.

Proposition 5.1.24. *All variants of naive, stage, and stable semantics satisfy inherited naivety in* $\mathfrak{C}_u$*.*

H-naivety, on the other hand, is not satisfied by any of the considered semantics in the general case, except for h-naive semantics. As shown in Example 4.1.2, i-naive semantics violate this principle even in the well-formed case. Using results from Section 4.2, we obtain the following result.

Proposition 5.1.25. *H-naive semantics satisfies hybrid naivety in* $\mathfrak{C}_u$*. Moreover, all variants of stage and stable semantics satisfy hybrid naivety in* $\mathfrak{C}_{wf}$*.*

The *reinstatement principle* first studied in [11] states that an extension should contain all arguments it defends. One possible way is to extend it using the same schema as for conflict-freeness, admissibility, and defense.

Principle 5.1.26 (i-Reinstatement)**.** *A semantics* ρ *satisfies* inherited reinstatement *in* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$*, for every* $S \in \rho(\mathcal{F})$*, there is a realization* E *of* S *in* $\mathcal{F}$ *that contains all arguments it defends.*

As sketched above, this principle is satisfied by all semantics that admit complete realizations. By Lemma 4.2.1, we obtain counter-examples for the remaining cases.

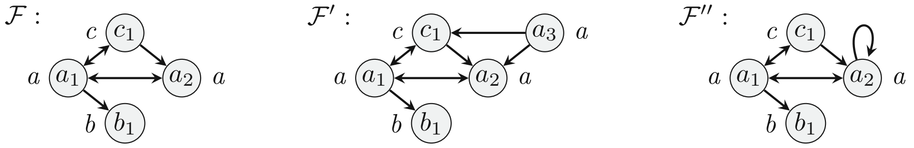
Proposition 5.1.27. *Complete, grounded, h-ad-stable, i-stable, and both variants of preferred and semi-stable semantics satisfy inherited reinstatement in* $\mathfrak{C}_u$.

We consider another generalization of reinstatement that weakens the conditions. The intuition is that it suffices to require that some *some* argument with claim c is defended in order to reinstate the claim.

Principle 5.1.28 (Reinstatement)**.** *A semantics* ρ *satisfies* reinstatement *in* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$, *for every* $S \in \rho(\mathcal{F})$, *if there is a realization* E *of* S *in* $\mathcal{F}$ *that defends an argument* $a \in A_{\mathcal{F}}$ *then* $cl(a) \in S$.

The principle is a faithful generalization of the reinstatement principle for AFs. Indeed, the principle corresponds to its AF counter-part if if $cl(a) = a$. Interestingly, the principle is not satisfied by any claim semantics under consideration in the general case.

Example 5.1.29. *Let us consider the following three CAFs* $\mathcal{F}$, $\mathcal{F}'$, *and* $\mathcal{F}''$*:*



First, we consider the CAF $\mathcal{F}$ *and observe that he claim-set* $S = \{a\}$, *witnessed by realization* $\{a_1\}$, *is a* ρ*-extension of* $\mathcal{F}$ *for all except grounded and h-naive semantics. The realization* $E = \{a_2\}$ *of* S *defends the argument* b_1 *against the attack from* a_1, *nevertheless,* $cl(b_1) = b$ *is not contained in* S.

For grounded semantics, we adapt $\mathcal{F}$ *by adding another argument* a_3 *with claim* a *that attacks* c_1 *and* a_2*—the resulting CAF is called* $\mathcal{F}'$ *and is depicted above. This argument defends* a_1, *thus* $\{a\}$, *witnessed by* $\{a_1, a_3\}$, *is grounded in the modified CAF. Since* $E = \{a_2\}$ *defends* b_1 *in* $\mathcal{F}'$ *we obtain the desired counter-example.*

The third CAF $\mathcal{F}''$ *shows that h-naive semantics fail to satisfy i-reinstatement for general CAFs: The realization* $E = \{a_2\}$ *of* $S = \{a\}$ *defends* b_1 *although* b *is not contained in* S.

The underlying issue is that the semantics are in general not uniquely realized. Hence we can realize some claim-set S under semantics ρ via extension E_1 and defend claim c with some other realization E_2 of S. For well-formed semantics, each realization of a claim-set S attacks—and thus defends—the same arguments. We obtain that all complete-based semantics satisfy this principle in $\mathfrak{C}_{wf}$.

Proposition 5.1.30. *Complete, grounded, preferred, stable, and both variants of semi-stable semantics satisfy reinstatement in* $\mathfrak{C}_{wf}$.

Let us consider a strengthening of reinstatement which has been considered in the literature for the AF case already. CF-reinstatement [11] additionally requires that the extension is not in conflict with the argument it defends.

Principle 5.1.31 (CF-Reinstatement)**.** *A semantics ρ satisfies* CF-reinstatement *in $\mathfrak{C}$ iff for every CAF $\mathcal{F} \in \mathfrak{C}$, for every $S \in \rho(\mathcal{F})$, if there is a realization E of S in $\mathcal{F}$ that defends an argument $a \in A_\mathcal{F}$ and $E \cup \{a\}$ is conflict-free then $cl(a) \in S$.*

Interestingly, h-naive semantics satisfies CF-reinstatement even in $\mathfrak{C}_u$ as we show next. This observation gives h-naive semantics an exclusive status as it is the only semantics under consideration that retains this fundamental property for general CAFs.

Proposition 5.1.32. *H-naive semantics satisfy CF-reinstatement in $\mathfrak{C}_u$.*

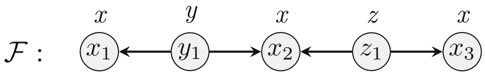
Proof. Consider a CAF $\mathcal{F}$, a h-naive extension S of $\mathcal{F}$, and a realization E of S in $\mathcal{F}$ that defends argument $a \in A_\mathcal{F}$ and satisfies $E \cup \{a\} \in cf(F)$. It holds that $S \subseteq cl(E \cup \{a\})$. It follows that $cl(a)$ is contained in S, otherwise, S is not $\subseteq$-maximal in $cf_i(\mathcal{F})$. □

Since both variants of stage semantics are contained in h-naive semantics for well-formed CAFs, we obtain that both semantics satisfy CF-reinstatement in $\mathfrak{C}_{wf}$. Moreover, each complete-based semantics satisfies CF-reinstatement.

Proposition 5.1.33. *Complete, grounded, preferred, h-naive, stable, and both variants of semi-stable and stage semantics satisfy CF-reinstatement in $\mathfrak{C}_{wf}$.*

Counter-examples for the remaining cases are by Example 5.1.29 (observe that $E \cup \{b_1\}$ is conflict-free in $\mathcal{F}$ and $\mathcal{F}'$). For i-naive semantics, consider the following example:

Example 5.1.34. *Consider the well-formed CAF $\mathcal{F}$ given as follows:*



The i-naive extensions of $\mathcal{F}$ are $\{x\}$, $\{x, y\}$, $\{x, z\}$, and $\{y, z\}$. For $S = \{x\}$, we can find a conflict-free realization E of x, namely $E = \{x_3\}$, that defends y_1 (the argument has no attacker) and $E \cup \{y_1\}$ is conflict-free. Nevertheless, $cl(y_1) = y$ is not contained in S. Therefore, i-naive semantics violates CF-reinstatement, even for well-formed CAFs.

Finally, let us consider a principle which states that a claim is credulously accepted if it is not defeated by any claim-extension. We consider only claims that are *cf-realizable*, that is, there is some argument with this claim that is not self-attacking.

Principle 5.1.35. *A semantics ρ satisfies* justified rejection *in $\mathfrak{C}$ iff for every CAF $\mathcal{F} \in \mathfrak{C}$, for every cf-realizable claim $c \in cl(A)$, if there is no $S \in \rho(\mathcal{F})$ with $c \in S$ then there is some ρ-realization E of a claim-set $S' \in \rho(\mathcal{F})$ that defeats c in $\mathcal{F}$.*

Proposition 5.1.36. *Conflict-free, h-stage and all variants of naive and stable semantics satisfy justified rejection in* $\mathfrak{C}_u$*.*

Proof. Conflict-free, i-naive, and h-naive semantics satisfy this principle because, by definition, if a claim c has a non-self-attacking occurrence, then there is an extension that contains c; thus the premise is never satisfied. Also, all stable variants satisfy justified rejection: if an extension does not contain a given claim c then c is defeated by it. Finally, also h-stage semantics satisfy justified rejection: consider some conflict-free set E in the underlying AF that contains claim c. Then either $cl(E)$ extends to a set with $\subseteq$-maximal range (thus the premise is not satisfied) or there is some other set D that defeats c. ☐

Proposition 5.1.37. *I-stage semantics satisfies justified rejection in* $\mathfrak{C}_{wf}$*.*

Proof. Consider a CAF $\mathcal{F}$, a claim $c \in cl(A)$ and let E_c denote the set of all cf-realizable arguments with claim c in $\mathcal{F}$. By well-formedness, each remaining argument with claim c is self-attacking and attacked by E_c. Hence $x \in E_c^{\oplus}$ for all $x \in A_{\mathcal{F}}$ with $cl(x) = c$. Thus there is some stage set E with $E_c^{\oplus} \subseteq E^{\oplus}$; i.e., $x \in E^{\oplus}$ for all $x \in A_{\mathcal{F}}$ with $cl(x) = c$. ☐

In general, i-stage and i-semi-stable semantics do not satisfy this principle; moreover, admissible-based semantics violate justified rejection in $\mathfrak{C}_{wf}$, as we show next.

Example 5.1.38. *Let us consider the following CAFs* $\mathcal{F}$ *and* $\mathcal{F}'$*:*

In $\mathcal{F}$*, the set* $\{z\}$ *is the unique stage and semi-stable extension in the underlying AF. However, the extension does not defeat claim* c*.*

In $\mathcal{F}'$*,* $\{z\}$ *is the only admissible set, thus it is the unique candidate for all admissible-based realizations. Nevertheless,* z *does not defeat* y*. Observe that* $\mathcal{F}'$ *is well-formed.*

Table 5.3 and Table 5.4 summarize our results for general and well-formed CAFs, respectively. We observe that conflict-freeness, admissibility, i-naivety, and justified rejection behave similar in general and well-formed CAFs. The only exception are h-*cf*-stable semantics which violate defense and admissibility in $\mathfrak{C}_u$ but satisfy both principles in $\mathfrak{C}_{wf}$ (recall that h-*cf*-stable semantics coincide with the other stable variants in this case). Comparing our results with the respective AF principles, we moreover obtain that—apart of h-*cf*-stable semantics—the aforementioned principles behave as expected.

For h-naivety, reinstatement, and CF-reinstatement, the picture looks different: reinstatement is not satisfied by any semantics in $\mathfrak{C}_u$ while h-naivety and CF-reinstatement are both only satisfied by h-naive semantics. As both properties are considered characteristic for naive semantics, our results indicate that h-naive semantics generalize naive semantics to CAFs in a reasonable way. This theory is supported by the fact that i-naive semantics does not satisfy any of the aforementioned principles, even in the well-formed case.

	Confl.-free	Defense/ Adm.	i-Na.	h-Na.	i-Reinst.	Reinst.	CF-Reinst.	Just. Reject.
cf_i	✓	✗	✗	✗	✗	✗	✗	✓
ad_i	✓	✓	✗	✗	✗	✗	✗	✗
gr_i	✓	✓	✗	✗	✓	✗	✗	✗
co_i	✓	✓	✗	✗	✓	✗	✗	✗
pr_i	✓	✓	✗	✗	✓	✗	✗	✗
pr_h	✓	✓	✗	✗	✓	✗	✗	✗
stb_i	✓	✓	✓	✗	✓	✗	✗	✓
$cf\text{-}stb_h$	✓	✗	✓	✗	✗	✗	✗	✓
$ad\text{-}stb_h$	✓	✓	✓	✗	✓	✗	✗	✓
ss_i	✓	✓	✗	✗	✓	✗	✗	✗
ss_h	✓	✓	✗	✗	✓	✗	✗	✗
na_i	✓	✗	✓	✗	✗	✗	✗	✓
na_h	✓	✗	✓	✓	✗	✗	✓	✓
stg_i	✓	✗	✓	✗	✗	✗	✗	✗
stg_h	✓	✗	✓	✗	✗	✗	✗	✓

Table 5.3: Basic principles w.r.t. general CAFs.

	Confl.-free	Defense/ Adm.	i-Na.	h-Na.	i-Reinst.	Reinst.	CF-Reinst.	Just. Reject.
cf_i	✓	✗	✗	✗	✗	✗	✗	✓
ad_i	✓	✓	✗	✗	✗	✗	✗	✗
gr_i	✓	✓	✗	✗	✓	✓	✓	✗
co_i	✓	✓	✗	✗	✓	✓	✓	✗
pr_i	✓	✓	✗	✗	✓	✓	✓	✗
stb_i	✓	✓	✓	✓	✓	✓	✓	✓
ss_i	✓	✓	✗	✗	✓	✓	✓	✗
ss_h	✓	✓	✗	✗	✓	✓	✓	✗
na_i	✓	✗	✓	✗	✗	✗	✗	✓
na_h	✓	✗	✓	✓	✗	✗	✓	✓
stg_i	✓	✗	✓	✓	✗	✗	✓	✓
stg_h	✓	✗	✓	✓	✗	✗	✓	✓

Table 5.4: Basic principles w.r.t. well-formed CAFs.

5.1.3 Set-theoretical Principles

In this section, our object of interest is the structure of so-called *extension-sets*, i.e., sets of sets of claims or, to be more precise, the set of all claim-extensions that are acceptable with respect to a given semantics. We recall classical set-theoretical principles and introduce novel principles in order to identify subtle differences between extension-sets for CAF semantics. Our set-theoretical principles give rise to certain closure-criteria of the extension-sets and are crucial to provide expressiveness-results for CAF semantics.

Let us first consider the well-known I-maximality principle [11].

Principle 5.1.39 (I-maximality)**.** *A semantics* ρ *satisfies* I-maximality *in class* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$*, for every* $S, T \in \rho(\mathcal{F})$*, if* $S \subseteq T$ *then* $S = T$*.*

Let us first discuss the general case. By definition, h-preferred and h-naive semantics satisfy I-maximality; moreover, grounded semantics yields a unique extension and thus satisfies this principle as well.

Proposition 5.1.40. *Grounded, h-naive, and h-preferred semantics satisfy I-maximality.*

The principle is not satisfied by any of the remaining semantics under consideration for general CAFs. The CAF from Example 2.2.7 possesses the claim-extensions $\{x\}$, $\{x, y\}$ which are accepted under all except grounded, h-naive, and h-preferred semantics.

We obtain more positive results on well-formed CAFs: using our $\subseteq$-inclusion results from Section 4.2, we obtain that preferred, stable, as well as all variants of semi-stable and stage semantics satisfy I-maximality in $\mathfrak{C}_{wf}$.

Proposition 5.1.41. *Grounded, h-naive, and all variants of preferred, semi-stable, stage, and stable semantics satisfy I-maximality in* $\mathfrak{C}_{wf}$*.*

We obtain counter-examples for the remaining semantics utilizing Lemma 4.2.1.

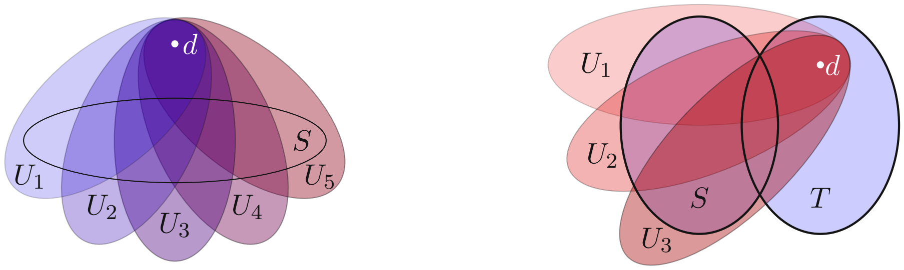
Next we consider the downward closure principle [83].

Principle 5.1.42 (Downward closure)**.** *A semantics* σ *is* downward closed *in* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$*, for every* $S \in \sigma(\mathcal{F})$*, if* $T \subseteq S$ *then* $T \in \sigma(\mathcal{F})$*.*

The unique semantics satisfying downward closure is conflict-free semantics.

Proposition 5.1.43. *Conflict-free semantics satisfy downward-closure in* $\mathfrak{C}_u$*.*

In what follows, we will recall principles from [84], which, roughly speaking, explain why particular sets (of arguments or, in our case, of claims) are not jointly acceptable with respect to a particular semantics. Moreover, we introduce novel principles in the same spirit of the aforementioned properties. In order to study such type of principles, the following notion will be useful.



(a) Tightness-criteria visualized. (b) Conflict-sensitivity-criteria visualized.

Figure 5.1: Graphical representation of the required conditions of tightness (5.1a) and conflict-sensitivity (5.1b): In Figure 5.1a, the set S is covered by the upper union $\bigcup_{i\leq 5} U_i$ of d. If tightness is satisfied by semantics ρ, then $S \cup \{d\}$ is contained in $\rho(\mathcal{F})$ for each $\mathcal{F}$. Figure 5.1b depicts the upper union $\bigcup_{i\leq 3} U_i$ of an element $d \in T$ which contains S. If S is contained in the upper union of each element of T, then $S \cup T$ is a claim-extension with respect to a semantics ρ that satisfies conflict-sensitivity.

Definition 5.1.44. *For $\mathbb{S} \subseteq 2^{\mathcal{C}}$ and $S \subseteq \bigcup_{T\in\mathbb{S}} T$, we define the* upper union *of S in $\mathbb{S}$ as*

$$\mathtt{up}_{\mathbb{S}}(S) = \bigcup_{S\subseteq T\in\mathbb{S}} T.$$

If we consider an I-maximal extension-set $\mathbb{S}$, we observe that the upper union becomes the identity function on $\mathbb{S}$. The upper union contains in this case only the input-set.

Proposition 5.1.45. *Given a semantics ρ that satisfies I-maximality and a CAF $\mathcal{F}$, it holds that $S = \mathtt{up}_{\rho(\mathcal{F})}(S)$ for each $S \in \rho(\mathcal{F})$.*

Let us next recall *tightness* and the *conflict-sensitivity* as introduced in [84].

Principle 5.1.46 (Tightness)**.** *A semantics ρ satisfies* tightness *in class $\mathfrak{C}$ iff for every CAF $\mathcal{F} \in \mathfrak{C}$, for every $S \in \rho(\mathcal{F})$ and for every claim $d \in cl(A)$, if $S \in \mathtt{up}_{\rho(\mathcal{F})}(\{d\})$ then $S \cup \{d\} \in \rho(\mathcal{F})$.*

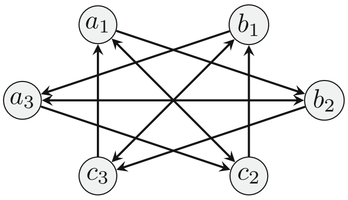
Principle 5.1.47 (Conflict-Sensitivity)**.** *A semantics ρ satisfies* conflict-sensitivity *in class $\mathfrak{C}$ iff for every CAF $\mathcal{F} \in \mathfrak{C}$, for every $S, T \in \rho(\mathcal{F})$, if $S \in \mathtt{up}_{\rho(\mathcal{F})}(\{d\})$ for all $d \in T$ then $S \cup T \in \rho(\mathcal{F})$.*

Figure 5.1 visualizes both properties. If tightness is satisfied by a semantics ρ, then $S \subseteq \bigcup_{i\leq 5} U_i = \mathtt{up}_{\rho(\mathcal{F})}(\{d\})$ (as shown in Figure 5.1a) implies $S \cup \{d\} \in \rho(\mathcal{F})$ for all CAFs $\mathcal{F}$. Conflict-sensitivity is satisfied by a semantics ρ, if $S \subseteq \bigcup_{i\leq 3} U_i = \mathtt{up}_{\rho(\mathcal{F})}(\{d\})$ as depicted in Figure 5.1b for all $d \in T$ implies $S \cup T \in \rho(\mathcal{F})$ for each CAF $\mathcal{F}$.

Remark 5.1.48. *In [84], conflict-sensitivity and tightness have been introduced via so-called* pairs*: a couple c, d forms a pair if there is an extension that contains both a* *and b. A semantics satisfies conflict-sensitivity iff for every two extensions S, T, if every couple c, d forms a pair then the union of S and T is an extension itself. A semantics satisfies tightness if for every extension S, for every claim d, if each couple c, d is a pair for every $c \in S$, then $S \cup \{d\}$ is an extension. Our formulation is indeed equivalent to the original formulation: S is contained in the upper union of a claim d iff c, d form a pair for all $c \in S$; conflict-sensitivity generalizes this concept to each claim $d \in T$.*

Grounded semantics satisfy conflict-sensitivity and tightness since they are single-status semantics. However, both properties turn out to be too strong when it comes to claim semantics, even for well-formed CAFs.

Example 5.1.49. *We consider the extension-set $\mathbb{S} = \{\{a, b\}, \{b, c\}, \{a, c\}\}$ which is neither tight nor conflict-sensitive. We generate the following well-formed CAF $\mathcal{F}$:*



For each claim c in set $S_i \in \mathbb{S}$, we introduce an argument c_i. Each set S is attacked by claims not appearing in S, e.g., the set $\{a, b\}$ is attacked by claim c. In this way, we ensure that $\mathcal{F}$ is well-formed. It can be checked that $\rho(\mathcal{F}) = \mathbb{S}$ for h-naive and for (all variants of) preferred, stable, semi-stable and stage semantics, moreover, $\mathbb{S} \cup \{\emptyset\}$ corresponds to $ad_i(\mathcal{F})$ and $co_i(\mathcal{F})$, while $\mathbb{S} \cup \{\{a\}, \{b\}, \{c\}\} = na_i(\mathcal{F})$ and $\mathbb{S} \cup \{\emptyset, \{a\}, \{b\}, \{c\}\} = cf_i(\mathcal{F})$.

We consider a novel principle that generalizes tightness and conflict-sensitivity.

Principle 5.1.50 (Cautious closure)**.** *A semantics ρ is* cautiously closed *iff for every CAF $\mathcal{F}$, for every $S, T \in \rho(\mathcal{F})$, if $S \subseteq \mathtt{up}_{\rho(\mathcal{F})}(T)$ then $S \cup T \in \rho(\mathcal{F})$.*

Figure 5.2 provides a graphical representation of this generalized criteria. Next we show that each semantics that satisfies conflict-sensitivity also satisfies cautious closure. Hence each AF semantics that satisfies conflict-sensitivity (e.g., admissible, grounded, preferred, stable, semi-stable, and stage semantics) satisfies the generalized principle as well.

Proposition 5.1.51. *Conflict-sensitivity implies cautious closure.*

Proof. Given a CAF $\mathcal{F}$ and two sets $S, T \in \rho(\mathcal{F})$. Moreover, let $S \subseteq \mathtt{up}_{\rho(\mathcal{F})}(T)$. This means in particular that S is contained in the upper union of each single claim $d \in T$, i.e., $S \in \mathtt{up}_{\rho(\mathcal{F})}(\{d\})$ for all $d \in T$. If $\rho(\mathcal{F})$ is conflict-sensitive, we obtain $S \cup T \in \rho(\mathcal{F})$. ☐

Since I-maximal extension-sets $\mathbb{S}$ satisfy $S = \mathtt{up}_{\mathbb{S}}(S)$ for each $S \in \mathbb{S}$, we obtain that each semantics that satisfies I-maximality satisfies cautious closure as well.

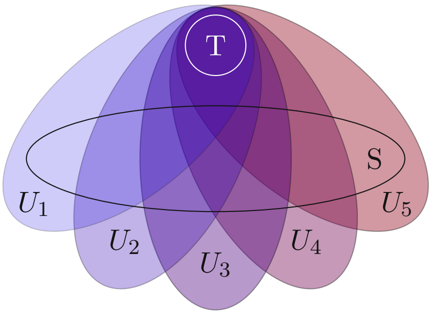


Figure 5.2: Graphical representation of the required conditions of cautious closure: the set S is covered by the upper union $\bigcup_{i \leq 5} U_i$ of T. If cautious closure is satisfied by semantics ρ, then this implies that $S \cup T$ is contained in $\rho(\mathcal{F})$. We have replaced the single claim d in Figure 5.1a by a set of claims T.

Proposition 5.1.52. *I-maximality implies cautious closure.*

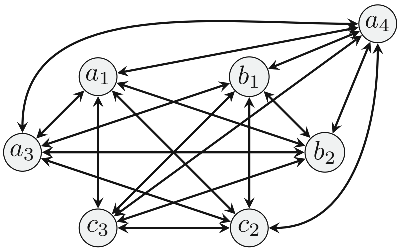
We obtain that grounded (by Proposition 5.1.51), h-preferred and h-naive semantics satisfy cautious closure even in the general case.

Proposition 5.1.53. *Grounded, h-preferred and h-naive semantics satisfy cautious closure in $\mathfrak{C}_u$.*

For the remaining semantics, we consider the following counter-example:

Example 5.1.54. *We consider the extension-set $\mathbb{S} = \{\{a,b\},\{b,c\},\{a,c\},\{a\}\}$. The set $\mathbb{S}$ is not cautiously closed: indeed, the upper union of $\{a\}$ is given by $\{a,b,c\}$ and thus contains $\{b,c\}$. Nevertheless, $\{a,b,c\}$ is not contained in $\mathbb{S}$.*

We generate the following CAF $\mathcal{F}$ by introducing an argument c_i for each claim c, for each claim-set $S_i \in \mathbb{S}$. Moreover, $cl(c_i) = c$. The attack-relation is defined as follows: two arguments c_i, d_j attack each other iff $i \neq j$. We obtain the following CAF:



The construction ensures that each claim-set has its unique realization that attacks all remaining arguments. In $\mathcal{F}$, all attacks are symmetric and thus admissible-based and conflict-free semantics coincide. We obtain that all considered semantics ρ apart from *grounded, h-naive, and h-preferred semantics satisfy* $\mathbb{S} \subseteq \rho(\mathcal{F})$, *moreover, the set* $\{a, b, c\}$ *is not accepted with respect to any of the considered semantics. It follows that cautious closure is violated by all semantics under consideration (apart from grounded, h-naive, and h-preferred semantics).*

Cautious closure is satisfied by several semantics if one considers the restriction to well-formed CAFs. First, by Proposition 5.1.52, we obtain that preferred, stable, h-naive, and both variants of semi-stable and stage semantics satisfy cautious closure.

Proposition 5.1.55. *Grounded, preferred, stable, h-naive, and both variants of semi-stable and stage semantics satisfy cautious closure in* $\mathfrak{C}_{wf}$.

Next we show that admissible semantics satisfy this principle in $\mathfrak{C}_{wf}$.

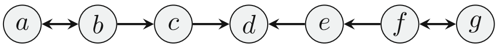
Proposition 5.1.56. *Admissible semantics satisfy cautious closure in* $\mathfrak{C}_{wf}$.

Proof. Given a well-formed CAF $\mathcal{F}$ and let $S, T \in ad_i(\mathcal{F})$ with $S \subseteq \mathtt{up}_{\rho(\mathcal{F})}(T)$. We show that $S \cup T \in ad_i(\mathcal{F})$.

Consider ad-realizations $E, D \subseteq A_{\mathcal{F}}$ of S and T, respectively. By Dung's fundamental lemma, the union $E \cup D$ defends itself in F. Now assume there is a conflict in $E \cup D$, i.e., there are arguments $x, y \in E \cup D$ such that $(x, y) \in R$. W.l.o.g. let $x \in E$ and $y \in D$ (as both E, D are admissible it is not the case that both arguments x, y are contained in either E or D). Since $S \subseteq \mathtt{up}_{\rho(\mathcal{F})}(T)$ there is some admissible superset $T' \supseteq T$ such that $T \cup \{cl(x)\} \subseteq T'$. Let D' denote an ad-realization of T' and let $x' \in D'$ denote the occurrence of $cl(x)$ in D', that is, $cl(x') = cl(x)$. Then $(x', y) \in R$ by well-formedness. Since D defends itself, there is an argument $z \in D$ that attacks x'. Let $z' \in D'$ denote the occurrence of claim $cl(z')$ in D', that is, $cl(z') = cl(z)$. By well-formedness, we have that $(z', x') \in R$, contradiction to $D' \in ad(F)$. ☐

Complete, conflict-free and i-naive semantics do not satisfy cautious closure. Example 5.1.49 serves as a counter-example for conflict-free and i-naive semantics; for complete semantics, we consider the following counter-example.

Example 5.1.57. *Consider the following CAF with* $cl = id$:



$\{b\}$ *and* $\{f\}$ *are complete, but* $\{b, f\}$ *is not complete since it defends the argument* d.

We consider a relaxation of cautious closure.

Principle 5.1.58 (Weak cautious closure)**.** *A semantics* ρ *is* weakly cautiously closed *iff for every CAF* $\mathcal{F}$, *for every* $S, T \in \rho(\mathcal{F})$, *if* $\mathtt{up}_{\rho(\mathcal{F})}(T)$ *then there is* $U \in \rho(\mathcal{F})$ *with* $S \cup T \subseteq U$.

Since each semantics that satisfies cautious closure also satisfies weak cautious closure, we obtain the following result.

Proposition 5.1.59. *Grounded, h-preferred and h-naive semantics satisfy weak cautious closure in* $\mathfrak{C}_u$.

Example 5.1.54 shows that the remaining semantics do not satisfy weak cautious closure. For well-formed CAFs, we first observe that Example 5.1.49 provides a counter-example for i-naive and conflict-free semantics. The remaining semantics satisfy this principle.

Proposition 5.1.60. *Grounded, admissible, complete, preferred, stable, h-naive, and both variants of semi-stable and stage semantics satisfy weak cautious closure in* $\mathfrak{C}_{wf}$.

Proof. To show that $co_i(\mathcal{F})$ is weakly cautiously closed for each well-formed CAF $\mathcal{F}$, consider two claim-sets $S, T \in co_i(\mathcal{F})$ with $\mathtt{up}_{\rho(\mathcal{F})}(T)$. Clearly, S and T are admissible in $\mathcal{F}$. By Proposition 5.1.56, we obtain $S \cup T \in ad_i(\mathcal{F})$, thus there is some complete claim-set $U \in co_i(\mathcal{F})$ with $S \cup T \subseteq U$. For the remaining semantics, we obtain the statement since the principle generalizes cautious closure. ☐

Let us next consider a principle that characterizes a crucial property of complete semantics. if two extensions S, T are contained in some other extension U, i.e., $S \cup T \subseteq U$, then there is a unique $\subseteq$-minimal extension that contains $S \cup T$. For this, it will be useful to define so-called *completion-sets* of a given set of claims.

Definition 5.1.61. *Given a CAF* $\mathcal{F}$*, a semantics* ρ *and a set of claims* $S \subseteq cl(A_{\mathcal{F}})$*, we let* $\mathtt{C}_{\rho(\mathcal{F})}(S) = \{T \in \rho(\mathcal{F}) \mid S \subseteq T, \nexists T' \in \rho(\mathcal{F}) : S \subseteq T' \subset T\}$ *denote the minimal completion-sets of* S *in* $\mathcal{F}$.

If $|\mathtt{C}_{\rho(\mathcal{F})}(S)| = 1$ we slightly abuse notation and write $\mathtt{C}_{\rho(\mathcal{F})}(S)$ to denote the unique minimal completion-set of S.

Principle 5.1.62 (Unique completion)**.** *A semantics* ρ *satisfies* unique completion *in* $\mathfrak{C}$ *iff for every CAF* $\mathcal{F} \in \mathfrak{C}$*, for every* $S, T \in \rho(\mathcal{F})$, $|\mathtt{C}_{\rho(\mathcal{F})}(S \cup T)| \leq 1$.

Proposition 5.1.63. *Cautious closure implies unique completion.*

Proof. The unique completion of two extensions $S, T \in \rho(\mathcal{F})$ in question is given by the union $T \cup S$. In case there are several completions of $T \cup S$, we have that $S \subseteq \mathtt{up}_{\rho(\mathcal{F})}(T)$ and thus $S \cup T \in \rho(\mathcal{F})$. ☐

We thus obtain that unique completion is satisfied by grounded, h-naive, and h-preferred semantics in the general case and additionally by admissible, stable, and both versions of semi-stable and stage semantics in $\mathfrak{C}_{wf}$.

Proposition 5.1.64. *Grounded, h-naive, and h-preferred semantics satisfy unique completion in $\mathfrak{C}_u$. Moreover, admissible, preferred, stable, h-naive, and both variants of semi-stable and stage semantics satisfy unique completion in $\mathfrak{C}_{wf}$.*

For general CAFs, the principle is not satisfied by any of the remaining semantics: a counter-example is given by Example 5.1.54, here, $\{a\}$ has two minimal completions $\{a, b\}$ and $\{a, c\}$.

Likewise, neither i-naive nor conflict-free semantics satisfy unique completion in $\mathfrak{C}_{wf}$: in Example 5.1.49, the sets $\{a, b\}, \{a, c\}, \{b, c\}$ as well as the singletons $\{a\}, \{b\}, \{c\}$ are conflict-free and i-naive claim-sets, thus each singleton has two minimal completions.

We end this section by showing that for well-formed CAFs, unique completion is satisfied by complete semantics.

Proposition 5.1.65. *Complete semantics satisfy unique completion in $\mathfrak{C}_{wf}$.*

Proof. Recall that in well-formed CAFs, each realization of a claim-set attacks the same arguments. Thus, every realization of $T \cup S$ for two extensions $S, T \in co_i(\mathcal{F})$ in a well-formed CAF $\mathcal{F}$ defends the same arguments. It follows that $S \cup T$ admits a unique completion in case $T \cup S$ is ad-realizable in $\mathcal{F}$. □

We summarize our results in Table 5.5 and Table 5.6 for general and well-formed CAFs, respectively. Apart from grounded semantics which satisfies almost all set-theoretical principles under consideration by definition, only h-naive and h-preferred semantics satisfy I-maximality, (weak) cautious closure, and unique completion in the general case. Tightness and conflict-sensitivity are also not satisfied in the well-formed case. Cautious closure, on the other hand, is satisfied by almost all admissible-based semantics except for complete semantics. Weak cautious closure and unique completion are satisfied by complete semantics as well.

5.2 Expressiveness

In this section, we investigate the expressive power of claim semantics. As already observed in the previous section, CAF semantics are in general more expressive than their AF counterparts: several semantics violate I-maximality, moreover, it is possible to construct (well-formed) CAFs that violate tightness and conflict-sensitivity which is impossible for e.g., preferred resp. admissible semantics as shown by Dunne et al. [84].

In order to study the expressive power of the considered semantics, we provide characterizations of the *signatures* of the semantics [84]. The signature captures all possible outcomes which can be obtained by argumentation frameworks when evaluated under a semantics and thus characterizes the expressiveness of a semantics.

Formally, the signature Σ_σ^{AF} of an AF-semantics σ is defined as $\Sigma_\sigma^{AF} = \{\sigma(F) \mid \mathcal{F}$ is an AF$\}$. We adapt the concept to CAFs resp. well-formed CAFs as follows.

	I-Max.	Downw. Closure	Tight	Conflict-sensitive	Cautious Closure	w-Cautious Closure	Unique Compl.
cf_i	✗	✓	✗	✗	✗	✗	✗
ad_i	✗	✗	✗	✗	✗	✗	✗
gr_i	✓	✗	✓	✓	✓	✓	✓
co_i	✗	✗	✗	✗	✗	✗	✗
pr_i	✗	✗	✗	✗	✗	✗	✗
pr_h	✓	✗	✗	✗	✓	✓	✓
stb_i	✗	✗	✗	✗	✗	✗	✗
$cf\text{-}stb_h$	✗	✗	✗	✗	✗	✗	✗
$ad\text{-}stb_h$	✗	✗	✗	✗	✗	✗	✗
ss_i	✗	✗	✗	✗	✗	✗	✗
ss_h	✗	✗	✗	✗	✗	✗	✗
na_i	✗	✗	✗	✗	✗	✗	✗
na_h	✓	✗	✗	✗	✓	✓	✓
stg_i	✗	✗	✗	✗	✗	✗	✗
stg_h	✗	✗	✗	✗	✗	✗	✗

Table 5.5: Set-theoretical principles w.r.t. general CAFs.

	I-Max.	Downw. Closure	Tight	Conflict-sensitive	Cautious Closure	w-Cautious Closure	Unique Compl.
cf_i	✗	✓	✗	✗	✗	✗	✗
ad_i	✗	✗	✗	✗	✓	✓	✓
gr_i	✓	✗	✓	✓	✓	✓	✓
co_i	✗	✗	✗	✗	✗	✓	✓
pr_i	✓	✗	✗	✗	✓	✓	✓
stb_i	✓	✗	✗	✗	✓	✓	✓
ss_i	✓	✗	✗	✗	✓	✓	✓
ss_h	✓	✗	✗	✗	✓	✓	✓
na_i	✗	✗	✗	✗	✗	✗	✗
na_h	✓	✗	✗	✗	✓	✓	✓
stg_i	✓	✗	✗	✗	✓	✓	✓
stg_h	✓	✗	✗	✗	✓	✓	✓

Table 5.6: Set-theoretical principles w.r.t. well-formed CAFs.

Definition 5.2.1. *The* signature *of a semantics ρ w.r.t. general and well-formed CAFs, respectively, is given by*

$$\begin{aligned}\Sigma_\rho^{CAF} &= \{\rho(\mathcal{F}) \mid \mathcal{F} \textit{ is a CAF}\}\\ \Sigma_\rho^{wf} &= \{\rho(\mathcal{F}) \mid \mathcal{F} \textit{ is a well-formed CAF}\}.\end{aligned}$$

Note that Σ_σ^{AF} yields a collection of sets of arguments while Σ_ρ^{CAF} and Σ_ρ^{wf} yield a collection of sets of claims. In order to compare argument-based signatures with their claim-based variants, we identify AFs with CAFs where each argument is assigned its unique argument name (i.e., $cl = id$) as done in Section 5.1. For any AF-semantics σ,

$$\Sigma_\sigma^{AF} \subseteq \Sigma_{\sigma_i}^{wf} \subseteq \Sigma_{\sigma_i}^{CAF} \quad \text{and} \quad \Sigma_\sigma^{AF} \subseteq \Sigma_{\sigma_h}^{wf} \subseteq \Sigma_{\sigma_h}^{CAF}$$

since each AF corresponds to a (well-formed) CAF with an unique claim per argument; moreover, each well-formed CAF is indeed a CAF.

5.2.1 Expressiveness of CAF Semantics

We begin our investigations with the class of general CAFs. As we will see, almost every extension-set can be expressed with only very soft restrictions, i.e., CAF semantics are in general very expressive, as the following theorem shows:

Theorem 5.2.2. *The following characterizations hold:*

$$\begin{aligned}
\Sigma_{gr_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid |\mathbb{S}| = 1\}\\
\Sigma_{cf_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \mathbb{S} \textit{ is downwards closed}\}\\
\Sigma_{ad_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \emptyset \in \mathbb{S}\}\\
\Sigma_{co_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \textstyle\bigcap_{S\in\mathbb{S}} S \in \mathbb{S}\}\\
\Sigma_{\rho}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \mathbb{S} \textit{ is I-maximal}\},\ \ \rho \in \{pr_h, na_h\}\\
\Sigma_{\rho}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} = \{\emptyset\} \textit{ or } \emptyset \notin \mathbb{S}\},\ \ \rho \in \{stb_i, cf\text{-}stb_h, ad\text{-}stb_h\}\\
\Sigma_{\rho}^{CAF} &= \Sigma_{stb_c}^{CAF} \setminus \{\emptyset\},\ \ \rho \in \{pr_i, na_i, ss_i, ss_h, stg_i, stg_h\}
\end{aligned}$$

From Section 5.1, we know that conflict-free semantics are downwards closed and that h-preferred and h-naive semantics satisfy I-maximality. This confirms that it is impossible to construct CAFs where conflict-free extension-sets are not downwards-closed or, e.g., h-naive semantics violate I-maximality, as postulated in the theorem. Moreover, the grounded extension is always unique, the empty set is always admissible, the intersection of all complete sets is complete, and stable semantics might return empty extension-sets.

In the remaining part of this section, we show that for each extension-set $\mathbb{S}$ which obeys the ρ-specific requirements, we can construct a CAF $\mathcal{F}$ that returns exactly $\mathbb{S}$ as ρ-extensions, i.e., $\rho(\mathcal{F}) = \mathbb{S}$.

First, each extension-set $\mathbb{S}$ with $|\mathbb{S}| = 1$ is expressible under grounded semantics: it suffices to consider the CAF $\mathcal{F} = (\{c \in S \mid S \in \mathbb{S}\}, \emptyset, id)$ with no attacks. Second, in order to obtain $\mathbb{S} = \{\emptyset\}$ we consider the empty framework $\mathcal{F} = (\emptyset, \emptyset, cl)$ which satisfies $\rho(\mathcal{F}) = \mathbb{S}$ for all considered semantics. Third, stable semantics can express $\mathbb{S} = \emptyset$: as for AFs, it suffices to consider a single self-attacking argument; the CAF $\mathcal{F} = (\{a\}, \{(a, a)\}, id)$ thus yields an example for $stb_i(\mathcal{F}) = ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F}) = \emptyset$.

Next, we define a method which can be used to construct CAFs that return each non-empty extension-set $\mathbb{S}$ that obeys the semantics-specific requirements for all apart from admissible and complete semantics. Note that we have used the construction already in Section 5.1 in Example 5.1.54 to show that grounded, h-naive, and h-preferred semantics do not satisfy cautious closure in general. The basic idea is to add an argument c_i for each claim c from claim-set S_i in a given extension-set $\mathbb{S}$ that attacks all arguments not associated to claims in S_i. In this way, each claim-set realizes itself in the resulting CAF.

Construction 5.2.3. *Given a non-empty extension-set* $\mathbb{S} = \{S_1, \dots, S_n\} \subseteq 2^{\mathsf{C}}$*, we define* $\mathcal{F}^u_{\mathbb{S}} = (A, R, cl)$ *with*

$$A = \{c_i \mid S_i \in \mathbb{S}, c \in S_i\},$$
$$R = \{(c_i, d_j) \mid c_i, d_j \in A, i \neq j\},$$

and $cl(c_i) = c$ *for all* $c_i \in A$.

Proposition 5.2.4. *Given a non-empty extension-set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$, $\emptyset \notin \mathbb{S}$*, let* $\mathcal{F}^u_{\mathbb{S}}$ *be defined as in Construction 5.2.3, and let* `Sem` *denote the set of all considered semantics. Then*

1. *if* $\emptyset \notin \mathbb{S}$, $\rho(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$ *for* $\rho \in \texttt{Sem} \setminus \{cf_i, ad_i, co_i, gr_i, pr_h, na_h\}$*;*
2. *if* $\mathbb{S}$ *is I-maximal,* $\rho(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$ *for* $\texttt{Sem} \setminus \{cf_i, ad_i, co_i, gr_i\}$*;*
3. *if* $\mathbb{S}$ *is downward closed,* $\rho(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$ *for* $\{cf_i, ad_i, co_i\}$*.*

Proof. We let $E_i = \{c_i \mid c \in S_i\}$ denote the corresponding realization of a set $S_i \in \mathbb{S}$.

(1) To show that $\rho(\mathcal{F}) = \mathbb{S}$ for each of the considered semantics, we first observe that each attack is symmetric. Hence $pr_i(\mathcal{F}^u_{\mathbb{S}}) = na_i(\mathcal{F}^u_{\mathbb{S}})$ and $ss_i(\mathcal{F}^u_{\mathbb{S}}) = stg_i(\mathcal{F}^u_{\mathbb{S}})$; also, $ss_h(\mathcal{F}^u_{\mathbb{S}}) = stg_h(\mathcal{F}^u_{\mathbb{S}})$ and $cf\text{-}stb_h(\mathcal{F}^u_{\mathbb{S}}) = ad\text{-}stb_h(\mathcal{F}^u_{\mathbb{S}})$ (since $cf_i(\mathcal{F}^u_{\mathbb{S}}) = ad_i(\mathcal{F}^u_{\mathbb{S}})$).

Second, we observe that for each $S_i \in \mathbb{S}$, the realization E_i is stable in the underlying AF, therefore, $stb_i(\mathcal{F}^u_{\mathbb{S}}) \neq \emptyset$ and thus $stb_i(\mathcal{F}^u_{\mathbb{S}}) = ss_i(\mathcal{F}^u_{\mathbb{S}}) = stg_i(\mathcal{F}^u_{\mathbb{S}})$. We moreover obtain $\mathbb{S} \subseteq stb_i(\mathcal{F}^u_{\mathbb{S}})$. As the CAF possesses a stable extension, we furthermore conclude that $cf\text{-}stb_h(\mathcal{F}^u_{\mathbb{S}}) = stg_h(\mathcal{F}^u_{\mathbb{S}})$ (by Lemma 4.2.9) and thus $ss_h(\mathcal{F}^u_{\mathbb{S}}) = stg_h(\mathcal{F}^u_{\mathbb{S}}) = cf\text{-}stb_h(\mathcal{F}^u_{\mathbb{S}}) = ad\text{-}stb_h(\mathcal{F}^u_{\mathbb{S}})$.

Third, we observe that all stable variants coincide. It suffices to show that $cf\text{-}stb_h(\mathcal{F}^u_{\mathbb{S}}) \subseteq stb_i(\mathcal{F}^u_{\mathbb{S}})$. Consider a h-$cf$-stable set S and its $cf\text{-}stb_h$-realization E

in $\mathcal{F}^u_{\mathbb{S}}$. We first observe that $S \subseteq S_i$ for some $S_i \in \mathbb{S}$ because all other claim-sets do not have a conflict-free realization in $\mathcal{F}^u_{\mathbb{S}}$. Moreover, $E \subseteq E_i$ because all other realizations of E are not conflict-free. E attacks all arguments with claims $c \notin S$. Now, assume there is an argument $a \in A \setminus E$ with $\mathit{cl}(a) \in S$ that is not attacked by E. This is the case only if $\mathit{cl}(a) \in S_i$. As each claim of the claim-set S_i has exactly one realization in E_i we have found a claim that is neither defeated nor contained in E, contradiction to our assumption E cf-stb_h-realizes S in $\mathcal{F}^u_{\mathbb{S}}$.

Finally, we observe that $\mathit{pr}_i(\mathcal{F}^u_{\mathbb{S}}) = \mathit{stb}_i(\mathcal{F}^u_{\mathbb{S}})$ since each $\subseteq$-maximal admissible set in F attacks all other arguments. As there are no other $\subseteq$-maximal admissible sets in the underlying AF we obtain $\mathit{pr}_i(\mathcal{F}^u_{\mathbb{S}}) \subseteq \mathbb{S}$. By $\mathbb{S} \subseteq \mathit{stb}_i(\mathcal{F}^u_{\mathbb{S}}) = \mathit{pr}_i(\mathcal{F}^u_{\mathbb{S}}) \subseteq \mathbb{S}$ we have shown that $\rho(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$ for all considered semantics as required.

(2) Now assume that $\mathbb{S}$ is I-maximal. By (1), we obtain the statement for all semantics in $\mathtt{Sem} \setminus \{\mathit{cf}_i, \mathit{ad}_i, \mathit{co}_i, \mathit{gr}_i, \mathit{pr}_h, \mathit{na}_h\}$. Since h-preferred and h-naive semantics can be equivalently defined based on preferred and naive argument-extensions, respectively (cf. Proposition 4.1.4), it holds that $\rho(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$ for $\rho \in \{\mathit{pr}_h, \mathit{na}_h\}$.

(3) Finally, let us assume that $\mathbb{S}$ is downward-closed. By (1), we obtain that $\mathbb{S} \setminus \{\emptyset\} = \rho(\mathcal{F}^u_{\mathbb{S}})$ for all semantics in $\mathtt{Sem} \setminus \{\mathit{cf}_i, \mathit{ad}_i, \mathit{co}_i, \mathit{gr}_i, \mathit{pr}_h, \mathit{na}_h\}$. As each subset of i-naive claim-sets is conflict-free, we obtain $\mathit{cf}_i(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$ as required. As observed in (1), conflict-free and admissible semantics coincide in $\mathcal{F}^u_{\mathbb{S}}$; moreover, $\emptyset = \bigcap_{S \in \mathbb{S}} S$ is contained in $\mathbb{S}$, furthermore, each realization E_i of S_i contains all arguments it defends, consequently, we furthermore obtain $\mathit{co}_i(\mathcal{F}^u_{\mathbb{S}}) = \mathbb{S}$. □

Evaluating $\mathcal{F}^u_{\mathbb{S}}$ under admissible and complete semantics might yield additional claim-sets. As observed in the proof of Proposition 5.2.4, $\mathit{ad}_i(\mathcal{F}^u_{\mathbb{S}})$ is downwards-closed for each extension-set $\mathbb{S}$. Moreover, the grounded extension is always empty in $\mathcal{F}^u_{\mathbb{S}}$ since there are no arguments that are unattacked. Consequently, $\mathbb{S} \cup \{\emptyset\} \subseteq \mathit{co}_i(\mathcal{F}^u_{\mathbb{S}})$ for each extension-set $\mathbb{S}$. We observe, however, that in both cases, the construction produces a CAF that accepts *at least* all claim-sets in $\mathbb{S}$ with respect to admissible and complete semantics.

Proposition 5.2.5. *Consider an extension-set $\mathbb{S}$ and let $\mathcal{F}^u_{\mathbb{S}}$ be defined as in Construction 5.2.3. It holds that $\mathbb{S} \subseteq \rho(\mathcal{F}^u_{\mathbb{S}})$ for $\rho \in \{\mathit{ad}_i, \mathit{co}_i\}$.*

For complete semantics, we adapt the construction appropriately. It suffices to apply Construction 5.2.3 to $\mathbb{S} \setminus \{\bigcap_{S \in \mathbb{S}} S\}$ and add isolated arguments for all claims in $\bigcap_{S \in \mathbb{S}} S$.

Proposition 5.2.6. *Given a non-empty extension-set $\mathbb{S} \subseteq 2^{\mathsf{C}}$ with $\bigcap_{S \in \mathbb{S}} S \in \mathbb{S}$. Let $\mathbb{T} = \mathbb{S} \setminus \{\bigcap_{S \in \mathbb{S}} S\}$ and let $\mathcal{F}^u_{\mathbb{T}} = (A, R, \mathit{cl})$ be defined as in Construction 5.2.3. We define $\mathcal{F} = (A \cup A', R, \mathit{cl}')$ with $A' = \{a_c \mid c \in \bigcap_{S \in \mathbb{S}} S\}$ and $\mathit{cl}'(a_c) = c$ for $a_c \in A'$ and $\mathit{cl}'(a) = \mathit{cl}(a)$ otherwise. It holds that $\mathit{co}_i(\mathcal{F}) = \mathbb{S}$.*

Proof. Consider an extension-set $\mathbb{S} = \{S_1, \dots, S_n\}$. We first observe that all arguments in A' are not attacked and thus contained in each complete set in F.

Second, we show that each claim-set $S_i \in \mathbb{S}$ is co_i-realized in $\mathcal{F}$: For $S_i = \bigcap_{S\in\mathbb{S}} S$, we observe that $\mathcal{F}$ contains precisely one argument a_c with claim c for all claims $c \in \bigcap_{S\in\mathbb{S}} S$. The set that contains all these arguments—the set A'—defends itself as it is unattacked; moreover, it does not defend any other arguments as it has no outgoing attacks. Consequently, $\bigcap_{S\in\mathbb{S}} S \in co_i(\mathcal{F})$. Note that no subset of $\bigcap_{S\in\mathbb{S}} S$ is complete.

In case $S_i \neq \bigcap_{S\in\mathbb{S}} S$, we consider the realization $E_i = \{c_i \mid c \in S_i\} \cup A'$ of S_i. Observe that E_i is conflict-free and attacks all remaining arguments by construction, thus it is stable and in particular complete in F. Moreover, no subset of S_i is complete since each argument in E_i attacks all arguments in $A \setminus E_i$ and thus defends all arguments in E_i. Finally, we note that no superset of E_i is complete in F. Consequently, $co(F) = \{E_i \mid i \leq n\}$. We thus obtain $co_i(\mathcal{F}) = \mathbb{S}$, as desired. □

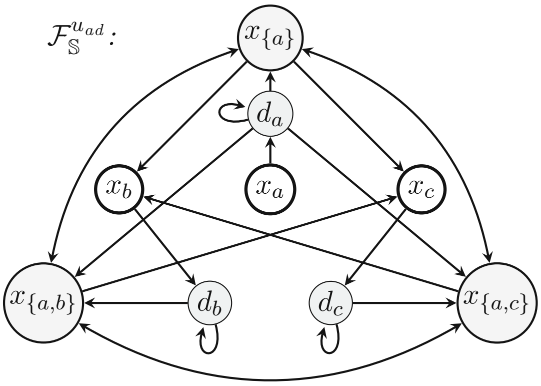
It remains to give a construction for admissible semantics. We let $[\mathbb{S}] = \bigcup_{S\in\mathbb{S}} S$ denote the set of all claims that appear in $\mathbb{S}$.

Construction 5.2.7. *Given a set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$, *we define* $\mathcal{F}^{u_{ad}}_{\mathbb{S}} = (A, R, cl)$ *with*

$$\begin{aligned} A &= \{x_S \mid S \in \mathbb{S}, S \neq \emptyset\} \cup \{x_c, d_c \mid c \in [\mathbb{S}]\}, \\ R &= \{(x_S, x_T) \mid S, T \in \mathbb{S}, S \neq T\} \ \cup \{(x_S, x_c) \mid S \in \mathbb{S}, c \in [\mathbb{S}] \setminus S\} \cup \\ &\quad \{(x_c, d_c), (d_c, d_c) \mid c \in [\mathbb{S}]\} \cup \{(d_c, x_S) \mid S \in \mathbb{S}, c \in S\}, \end{aligned}$$

$cl(x_c) = cl(d_c) = c$ *and* $cl(x_S) \in S$, *i.e., for* x_S *we pick an arbitrary claim from the set* S.

Example 5.2.8. *Let* $\mathbb{S} = \{\emptyset, \{a\}, \{a,b\}, \{a,c\}\}$. *Following Construction 5.2.7, we introduce an argument* x_S *for each* $S \in \mathbb{S}$, *moreover, we add attacks between all arguments* x_S *and* x_T, $T \neq S$. *Each such argument belongs to the admissible extension that realizes* S *in the resulting CAF. Next, we add two arguments for each of the claims* a, b, c *in* $[\mathbb{S}]$: *argument* x_c *with claim* c *and a self-attacking argument* d_c. *We obtain the following CAF* $\mathcal{F}^{u_{ad}}_{\mathbb{S}}$ *(claims are omitted, arguments that represent claims are filled white):*



The set $\{x_{\{a,b\}}, x_a, x_b\}$ *is admissible in* $\mathcal{F}^{u_{ad}}_{\mathbb{S}}$: $x_{\{a,b\}}$ *defends* x_b *against the attacks from* $x_{\{a\}}$ *and* $x_{\{a,c\}}$. *Moreover, the arguments* x_a *and* x_b *attack* d_a *and* d_b, *resp., and thus defend* $x_{\{a,b\}}$. *Hence* $\{a,b\}$ *is* ad_i*-realizable in* $\mathcal{F}^{u_{ad}}_{\mathbb{S}}$. *It can be checked that* $ad_i(\mathcal{F}^{u_{ad}}_{\mathbb{S}}) = \mathbb{S}$.

Proposition 5.2.9. *Given a set $\mathbb{S} \subseteq 2^{\mathsf{C}}$ such that $\emptyset \in \mathbb{S}$, and let $\mathcal{F}_{\mathbb{S}}^{u_{ad}}$ be defined as in Construction 5.2.7. It holds that $ad_i(\mathcal{F}_{\mathbb{S}}^{u_{ad}}) = \mathbb{S}$.*

Proof. We denote the underlying AF of $\mathcal{F}_{\mathbb{S}}^{u_{ad}}$ by F. First, let us show that each $S \in \mathbb{S}$ is admissible realizable in F. Indeed, the set $E = \{x_S\} \cup \{x_c \mid c \in S\}$ is admissible in F and satisfies $cl(E) = S$: E is conflict-free by construction, moreover, each argument x_c defends x_S against the attack from d_c. Furthermore, x_S attacks all remaining set-arguments. Thus E is admissible in F.

Next, we show that no proper superset of E is admissible in F: as each other set-argument is attacked, it holds that $E \cup \{x_T\}$ is conflicting for each x_T, $T \neq S$. Moreover, each dummy argument d_c is self-attacking, thus $E \cup \{d_c\}$ is conflicting for each $c \in [\mathbb{S}]$. Finally, since each claim-argument x_c with $c \notin S$ is attacked by $x_S \in E$, we obtain that no proper superset of E is conflict-free.

It remains to show that no proper subset of E is admissible. First, we observe that $E \setminus \{x_S\}$ is not admissible as it does not defend itself. In case we remove some argument x_c for some $c \in S$, we have that x_S is no longer defended against the attack from d_c. Consequently, we obtain $ad_i(\mathcal{F}_{\mathbb{S}}^{u_{ad}}) = \mathbb{S}$. ☐

5.2.2 Expressiveness of well-formed CAFs

Turning now to well-formed CAFs, we have already seen in Sections 4.1, 4.2, and 5.1 that the semantics under considerations admit a different behavior compared to the general case when restricted to this CAF-class. I-maximality is satisfied by preferred, h-naive, stable, and all variants of semi-stable and stage semantics; moreover, admissible and complete semantics satisfy cautious respectively weak cautious closure, indicating that not all extension-sets are expressible with respect to well-formed CAFs.

Our characterization results for well-formed CAFs can be summarized as follows:

Theorem 5.2.10. *The following characterizations hold:*

$$\begin{aligned}
\Sigma_{gr_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid |\mathbb{S}| = 1\} \\
\Sigma_{cf_i}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \mathbb{S} \textit{ is downwards-closed}\} \\
\Sigma_{ad_i}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \emptyset \in \mathbb{S}, \mathbb{S} \textit{ is cautiously closed}\} \\
\Sigma_{co_i}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \textstyle\bigcap_{S \in \mathbb{S}} S \in \mathbb{S}, \mathbb{S} \textit{ is weak-cautiously closed} \\
&\qquad \textit{and satisfies unique completion}\} \\
\Sigma_{\rho}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \textit{ is I-maximal}\},\ \rho \in \{stb_i, cf\text{-}stb_h, ad\text{-}stb_h\} \\
\Sigma_{\rho}^{wf} &= \Sigma_{stb_c}^{wf} \setminus \{\emptyset\},\ \rho \in \{pr_i, pr_h, na_h, ss_i, ss_h, stg_i, stg_h\}
\end{aligned}$$

Remark 5.2.11. *We remark that signature characterizations for well-formed CAFs for some of the semantics, i.e., for conflict-free, h-naive, grounded, admissible, complete, preferred, stable, h-semi-stable, and h-stage semantics, can also be obtained through recent* *expressiveness results for SETAFs and their relation to well-formed CAFs: SETAF signature characterizations provided in [85] translate to well-formed CAFs via the semantics-preserving transformation presented in Section 3.4 (adapted to h-semantics in Section 4.3.4). Hence the signatures for the aforementioned semantics coincide with their SETAF counter-part. However, in order to obtain a well-formed CAF having specific extensions, it is necessary to first construct a SETAF, determine its normal form, and apply Translation 3.4.10. In order to avoid this detour over SETAFs, we will present genuine signature constructions for well-formed CAFs from Theorem 5.2.10 in the subsequent part of this section. We moreover note that for admissible and complete semantics, the formulations of the signature characterizations slightly differ: in [85], the distinctive characteristics of admissible and complete semantics are* set-conflict-sensitivity *and* set-com-closure*, respectively. The definitions are equivalent to our formulations, as we demonstrate with our constructions. Hence our formulation yields an alternative view on admissible and complete semantics in SETAFs.*

As the attentive reader might have noticed, Theorem 5.2.10 does not speak about i-naive semantics. Indeed, the characterization of the signature for well-formed CAFs for i-naive semantics remains an open problem. We discuss several observations and known (im)possibility-results at the end of this section.

Signatures for grounded and conflict-free semantics coincide with those for general CAFs using $\Sigma_{\sigma}^{AF} \subseteq \Sigma_{\sigma_i}^{wf} \subseteq \Sigma_{\sigma_i}^{CAF}$ and the coincidence of $\Sigma_{\sigma}^{AF} = \Sigma_{\sigma_i}^{CAF}$ for $\sigma \in \{cf, gr\}$.

I-maximality characterizes stable, preferred, h-naive, and both variants of semi-stable and stage semantics, as we show next. To do so, we consider a construction that has been used already in Section 5.1 in Example 5.1.49 to show that tightness and conflict-sensitivity is not satisfied by any of the (non-single-status) semantics under consideration.

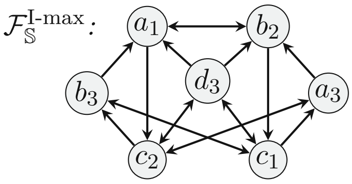
Construction 5.2.12. *For a set $\mathbb{S} = \{S_1, \dots, S_n\}$, we define $\mathcal{F}_{\mathbb{S}}^{\text{I-max}} = (A, R, cl)$ with*

$$A = \{c_i \mid c \in S_i, 1 \leq i \leq n\},$$
$$R = \{(c_i, d_j) \mid 1 \leq i, j \leq n, c \notin S_j\},$$

and $cl(c_i) = c$ for all $c_i \in A$.

We obtain well-formed CAFs as arguments with the same claim attack the same arguments.

Example 5.2.13. *Consider the extension-set $\mathbb{S} = \{\{a, c\}, \{b, c\}, \{a, b, d\}\}$. Applying Construction 5.2.12 yields the following CAF:*



Next we show that each I-maximal non-empty extension-set can be obtained under preferred, stable, h-naive, and both variants of semi-stable and stage semantics when applying Construction 5.2.12. For the case $\mathbb{S} = \emptyset$, we consider again the CAF that contains a single self-attacking argument only. The following proposition thus proves signature characterizations from Theorem 5.2.10 for all of the aforementioned semantics.

Proposition 5.2.14. *Given an I-maximal non-empty extension-set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$*, let* $\mathcal{F}_{\mathbb{S}}^{\text{I-max}}$ *be defined as in Construction 5.2.12. It holds that* $\rho(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = \mathbb{S}$ *for each semantics* $\rho \in \{stb_i, cf\text{-}stb_h, ad\text{-}stb_h, pr_i, pr_h, na_h, ss_i, ss_h, stg_i, stg_h\}$.

Proof. Let $\mathbb{S} = \{S_1, \dots, S_n\}$. For each claim-set $S_i \in \mathbb{S}$ we denote its canonical realization in $\mathcal{F}_{\mathbb{S}}^{\text{I-max}}$ by $E_i = \{c_i \mid c \in S_i\}$. We write F to denote the underlying AF of $\mathcal{F}_{\mathbb{S}}^{\text{I-max}}$.

First, we prove the statement for admissible-based semantics. Since $stb(F) \subseteq pr(F)$ holds, it suffices to show (1) $\{E_i \mid i \leq n\} \subseteq stb(F)$ and (2) $pr(F) \subseteq \{E_i \mid i \leq n\}$.

(1) By construction, E_i is conflict-free in F for each $S_i \in \mathbb{S}$. Moreover, E_i attacks all d_j with $j \neq i$ since S_i and S_j are incomparable, hence there is an $c \in S_i$ which does not occur in S_j. Thus E_i is a stable extension of F.

(2) Consider a preferred set $E \in pr(F)$. We show that E is a subset of E_i for some $i \leq n$. First, we observe that $cl(E) \subseteq S_i$ for some $S_i \in \mathbb{S}$, otherwise, E is conflicting: if E realizes a claim d that does not occur in S_i then each argument $c_i \in S_i$ is attacked by arguments with claim d. Thus $cl(E) \subseteq S_i$ for some $i \leq n$.

    Now, towards a contradiction, assume that there is an argument $c_j \in E$ with $i \neq j$. As S_i and S_j are incomparable there is a claim $d \in S_i \setminus S_j$ that attacks c_j (i.e., each argument with claim d attacks c_j), in particular, the argument d_i attacks c_j. Since $cl(E) \subseteq S_i$, there is no argument in E that attacks d_i, otherwise S_i would be conflicting. Consequently, $E \subseteq E_i$.

    From (1), we already know that $E_i \in pr(F)$ for each $S_i \in \mathbb{S}$ (since each stable extension is preferred). Hence, by the $\subseteq$-maximality of preferred extensions, it holds that $E = E_i$.

By (1) & (2) we obtain $\mathbb{S} \subseteq stb_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) \subseteq pr_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) \subseteq \mathbb{S}$, thus

$$stb_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = ss_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = ss_h(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = pr_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = \mathbb{S}.$$

Recall that in well-formed CAFs, all variants of stable semantics coincide. Likewise, all variants of preferred semantics yield the same outcome.

Next, we show that (3) $na_h(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) \subseteq \mathbb{S}$. First, we observe that each $S_i \in \mathbb{S}$ is cf_i-realizable via E_i. Second, there is no $E \subseteq A_F$ with $cl(E) \supset S_i$: as already observed in (2), there is no set of arguments $E \subseteq A_F$ with $cl(E) \supset S_i$ that is conflict-free in F.

By (1) & (3) we obtain $\mathbb{S} \subseteq stb_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) \subseteq na_h(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) \subseteq \mathbb{S}$, thus

$$stb_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = stg_i(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = stg_h(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = na_h(\mathcal{F}_{\mathbb{S}}^{\text{I-max}}) = \mathbb{S}.$$

This concludes the proof of the proposition. □

It remains to provide proofs for the signature characterizations for admissible and complete semantics for well-formed CAFs. We show that the signature for admissible semantics is characterized by cautious closure and empty-set-acceptance; moreover, we show that complete semantics can express each extension-set $\mathbb{S}$ that is weakly cautiously closed, satisfies unique completion and contains $\bigcap_{S\in\mathbb{S}} \mathbb{S}$.

We start by introducing a construction that will serve as basis to express extension-sets under admissible and complete semantics. For this, it will be convenient to introduce a function $\mathtt{min}_{\mathbb{S}}(c)$ that returns, for a given extension-set $\mathbb{S}$ and a claim $c \in [\mathbb{S}]$, the $\subseteq$-minimal sets in $\mathbb{S}$ that contain c.

Definition 5.2.15. *Given an extension-set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$ *and a claim* $c \in [\mathbb{S}]$, *we define* $\mathtt{min}_{\mathbb{S}}(c) = \{M \in \mathbb{S} \mid c \in M, \nexists S \in \mathbb{S}(S \subset M \wedge c \in S)\}$.

For I-maximal extension-sets, the function $\mathtt{min}_{\mathbb{S}}(c)$ will return all sets in extension-set $\mathbb{S}$ that contain the claim $c \in [\mathbb{S}]$. Indeed, if $\mathbb{S} \setminus \{\emptyset\}$ is incomparable, then $\mathtt{min}_{\mathbb{S}}(c) = \{M \in \mathbb{S} | c \in M\}$ for each $M \in \mathbb{S}$.

Example 5.2.16. *Consider the extension-set* $\mathbb{S} = \{\emptyset, \{a,c\}, \{b,c\}, \{c\}, \{a,b,d\}\}$. *The* $\subseteq$*-minimal sets relative to claims in* $[\mathbb{S}]$ *are given by*

$$\mathtt{min}_{\mathbb{S}}(a) = \{\{a,c\},\{a,b,d\}\} \qquad \mathtt{min}_{\mathbb{S}}(b) = \{\{b,c\},\{a,b,d\}\}$$
$$\mathtt{min}_{\mathbb{S}}(c) = \{\{c\}\} \qquad \mathtt{min}_{\mathbb{S}}(d) = \{\{a,b,d\}\}$$

Now, consider the I-maximal extension-set $\mathbb{S}' = \mathbb{S} \setminus \{\emptyset, \{c\}\}$. *We obtain*

$$\mathtt{min}_{\mathbb{S}'}(a) = \{\{a,c\},\{a,b,d\}\} \qquad \mathtt{min}_{\mathbb{S}'}(b) = \{\{b,c\},\{a,b,d\}\}$$
$$\mathtt{min}_{\mathbb{S}'}(c) = \{\{a,c\},\{b,c\}\} \qquad \mathtt{min}_{\mathbb{S}'}(d) = \{\{a,b,d\}\}$$

We are ready to present our construction that will serve as basis to characterize admissible and complete semantics.

Construction 5.2.17. *Given an extension-set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$, *we define* $\mathcal{F}_{\mathbb{S}} = (A, R, cl)$ *with*

$$A = \{c_M \mid c \in [\mathbb{S}], M \in \mathtt{min}_{\mathbb{S}}(c)\},$$
$$R = \{(c_M, c'_{M'}) \mid c_M, c'_{M'} \in A, c \notin \mathtt{up}_{\mathbb{S}}(M')\},$$

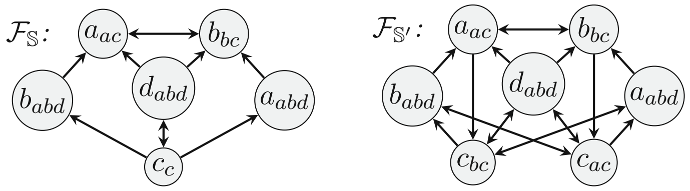
and $cl(c_M) = c$ *for all* $c_M \in A$.

$\mathcal{F}_{\mathbb{S}}$ is well-formed since each attack depends on the claim of the attacking argument. Moreover, in case $\mathbb{S} \setminus \{\emptyset\}$ is incomparable, we have $\mathtt{min}_{\mathbb{S}}(c) = \{M \in \mathbb{S} | c \in M\}$ and $\mathtt{up}_{\mathbb{S}}(M) = M$ for each $M \in \mathbb{S}$, thus $\mathcal{F}_{\mathbb{S}}$ can be written as

$$A = \{c_S \mid S \in \mathbb{S}, c \in S\},$$
$$R = \{(c_S, c'_{S'}) \mid c_S, c'_{S'} \in A, c \notin S'\},$$

with cl as defined above. Here, we obtain the CAF $\mathcal{F}_{\mathbb{S}}^{\text{I-max}}$ from Construction 5.2.3. Hence $\mathcal{F}_{\mathbb{S}}$ generalizes $\mathcal{F}_{\mathbb{S}}^{\text{I-max}}$ which extends to extension-sets that are not I-maximal.

Example 5.2.18. *Consider the extension-sets* $\mathbb{S} = \{\emptyset, \{a, c\}, \{b, c\}, \{c\}, \{a, b, d\}\}$ *and* $\mathbb{S}' = \mathbb{S} \setminus \{\emptyset, \{c\}\}$ *from Example 5.2.16. We note that both* $\mathbb{S}$ *and* $\mathbb{S}'$ *are cautiously closed. Construction 5.2.17 yields the following CAFs:*



Note that $\mathcal{F}_{\mathbb{S}'}$ *corresponds to the CAF from Example 5.2.13. We observe that there is only one single argument* c_c *in* $\mathcal{F}_{\mathbb{S}}$ *with claim* c *while* $\mathcal{F}_{\mathbb{S}'}$ *yields two arguments* c_{bc} *and* c_{ac} *with claim* c.

Attacks of $\mathcal{F}_{\mathbb{S}}$ *and* $\mathcal{F}_{\mathbb{S}'}$ *are constructed as follows: For each minimal set* M *that induces an argument* c_M, c_M *is attacked by all claims that are* not *contained in* $\mathtt{up}_{\mathbb{S}}(M)$. *For* $M = \{a, c\}$, *we have* $\mathtt{up}_{\mathbb{S}}(\{a, c\}) = \{a, c\}$ *as there are no proper supersets of* $\{a, c\}$, *thus the argument* a_{ac} *is attacked by all arguments having claim* b *or* d. *The set* $\{c\}$ *on the other hand, is contained in all non-empty sets of* $\mathbb{S}$ *except* $\{a, b, d\}$, *yielding* $\mathtt{up}_{\mathbb{S}}(\{c\}) = \{a, b, c\}$; *consequently,* c_c *is attacked only by the unique argument* d_{abd} *having claim* d.

We show that each set $S \in \mathbb{S}$ is admissible in $\mathcal{F}_{\mathbb{S}}$ in case $\mathbb{S}$ is weakly cautiously closed and contains $\emptyset$.

Proposition 5.2.19. *For a weakly cautiously closed set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$ *with* $\emptyset \in \mathbb{S}$, *let* $\mathcal{F}_{\mathbb{S}}$ *be as in Construction 5.2.17. It holds that* $\mathbb{S} \subseteq ad_i(\mathcal{F}_{\mathbb{S}})$.

Proof. Let $S \in \mathbb{S}$, and let $E = \{c_M \in A | M \subseteq S\}$. Clearly, $cl(E) = S$; moreover, E is conflict-free since $c \in \mathtt{up}_{\mathbb{S}}(M')$ for each $c_M, c'_{M'} \in E$ using $M' \subseteq S \subseteq \mathtt{up}_{\mathbb{S}}(M')$. It remains to show that S defends itself. Let c_N denote an argument with claim c that attacks E. We proceed by case distinction: (i) $S \subseteq \mathtt{up}_{\mathbb{S}}(N)$ and (ii) $S \not\subseteq \mathtt{up}_{\mathbb{S}}(N)$.

(i) In case $S \subseteq \mathtt{up}_{\mathbb{S}}(N)$, there is $T \in \mathbb{S}$ such that $N \cup S \subseteq T$ since $\mathbb{S}$ is weakly cautiously closed. Thus we obtain a contradiction to c_N attacks E by construction of $\mathcal{F}_{\mathbb{S}}$.

(ii) In case $S \not\subseteq \mathtt{up}_{\mathbb{S}}(N)$, there is some $d \in S$ such that $d \notin T$ for all upper sets $T \supseteq N$ of N in $\mathbb{S}$, i.e., $d \notin \mathtt{up}_{\mathbb{S}}(N)$. Thus, by construction of $\mathcal{F}_{\mathbb{S}}$, all arguments with claim d attack c_N. It remains to show that E contains an argument with claim d. Again, by construction of $\mathcal{F}_{\mathbb{S}}$, each claim in S appears as claim of some subset S' of S, thus there is an argument $d_{S'}$, $d \in S'$ for some $S' \subseteq S$, with claim d that attacks c_N. □

As cautious closure is a special case of weak cautious closure, the statement also holds true if $\mathbb{S}$ is cautiously closed. The other direction does not hold as the CAF $\mathcal{F}_{\mathbb{S}}$ in Example 5.2.18 demonstrates: Here, the argument d_{abd} defends itself, thus $\{d\}$ is admissible in $\mathcal{F}_{\mathbb{S}}$ although $\{d\} \notin \mathbb{S}$.

Next we show a property of $\mathcal{F}_{\mathbb{S}}$ that is crucial towards expressing suitable extension-sets under complete semantics: If $\mathbb{S}$ is weakly cautiously closed, then each admissible set E in $F_{\mathbb{S}}$ satisfies $\bigcup_{c_M \in E} M \subseteq S$ for some $S \in \mathbb{S}$.

Proposition 5.2.20. *For a weakly cautiously closed extension-set $\mathbb{S} \subseteq 2^{\mathsf{C}}$, it holds that for all $E \in ad(F_{\mathbb{S}})$, there is $S \in \mathbb{S}$ such that $\bigcup_{c_M \in E} M \subseteq S$.*

Proof. Consider some $E \in ad(\mathcal{F}_{\mathbb{S}})$. Then $cl(E) \subseteq \mathtt{up}_{\mathbb{S}}(M)$ for each $M \in \mathbb{S}$ with $c_M \in E$, otherwise there is $d \in cl(E)$ that attacks c_M, contradiction to conflict-freeness of E.

We show that for all arguments $c_M \in E$, for each claim $d \in M$, it holds that d does not attack E. Consider an argument $c_M \in E$. We proceed by case distinction: (i) $M \subseteq cl(E)$ and (ii) $M \not\subseteq cl(E)$.

(i) First assume $M \subseteq cl(E)$. As observed above, $cl(E) \subseteq \mathtt{up}_{\mathbb{S}}(M')$ for each argument $c'_{M'} \in E$, thus $d \in \mathtt{up}_{\mathbb{S}}(M')$ for each $d \in M$ and each argument $c'_{M'} \in E$. By construction of $\mathcal{F}_{\mathbb{S}}$, no $d \in M$ attacks E.

(ii) Now assume $M \not\subseteq cl(E)$. Towards a contradiction, let us assume that there is a claim $d \in M \setminus cl(E)$ that attacks E. That is, there is some argument d_N with claim d that attacks E and $N \subseteq M$ (since $d \in M$, there is $N \subseteq M$ such that N is a $\subseteq$-minimal set containing d in $\mathbb{S}$). Since E defends itself, there is some argument having claim $e \in cl(E)$ satisfying $e \notin \mathtt{up}_{\mathbb{S}}(N)$ (then e attacks d_N by construction of $\mathcal{F}_{\mathbb{S}}$). But then we obtain $e \notin \mathtt{up}_{\mathbb{S}}(N) \subseteq \mathtt{up}_{\mathbb{S}}(M)$, contradiction to $cl(E) \subseteq \mathtt{up}_{\mathbb{S}}(M)$.

We have shown that for all arguments $c_M \in E$, for each claim $d \in M$, it holds that d does not attack E. This means that for every two arguments $c_M, c'_{M'} \in E$, it holds that $M \subseteq \mathtt{up}_{\mathbb{S}}(M')$. By successive application of the weak cautious closure criteria, we obtain that there is $S \in \mathbb{S}$ with $\bigcup_{c_M \in E} M \subseteq S$. □

Moreover, in case $\mathbb{S}$ furthermore satisfies unique completion, then each union of two sets in $\mathbb{S}$ defends all 'missing elements' of its completion-set in $\mathcal{F}_{\mathbb{S}}$.

Proposition 5.2.21. *For a weakly cautiously closed set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$ *that satisfies unique completion, let* $S, T \in \mathbb{S}$ *and* $\mathcal{F}_{\mathbb{S}}$ *be as in Construction 5.2.17. It holds* $S \cup T$ *defends all arguments* c_M *that satisfy (1)* $c \in \mathtt{C}_{\mathbb{S}}(S \cup T) \setminus (S \cup T)$ *and (2)* $M \subseteq \mathtt{C}_{\mathbb{S}}(S \cup T)$.

Proof. Given $S, T \in \mathbb{S}$ and consider an argument c_M with $c \in \mathtt{C}_{\mathbb{S}}(S \cup T) \setminus (S \cup T)$ and $M \subseteq \mathtt{C}_{\mathbb{S}}(S \cup T)$, and let $c'_{M'}$ be and attacker of c_M in $\mathcal{F}_{\mathbb{S}}$. Consequently, $c' \notin \mathtt{up}_{\mathbb{S}}(M)$. Now assume c_M is not defended against the attack from $c'_{M'}$ by $S \cup T$. This is the case only if $S \cup T$ is contained in the union of all upper sets of M', i.e., $S \cup T \subseteq \mathtt{up}_{\mathbb{S}}(M')$. Since $\mathbb{S}$ is weakly closed, there is some set $U \in \mathbb{S}$ that contains $S \cup T \cup M'$; by unique completion we may furthermore assume that $\mathtt{C}_{\mathbb{S}}(S \cup T) \subseteq U$. But then we have $c' \in U \subseteq \mathtt{up}_{\mathbb{S}}(M)$, contradiction to our initial assumption $c'_{M'}$ attacks c_M. □

Next we show that each weakly closed extension-set $\mathbb{S}$ that satisfies unique completion and contains $\bigcap \mathbb{S}$ is a superset of $co_i(\mathcal{F})$. A crucial property is that arguments that correspond to the same minimal set (i.e., they possess the same subscript) are attacked by the same arguments.

Proposition 5.2.22. *For a weakly cautiously closed set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$ *which satisfies unique completion and contains* $\bigcap \mathbb{S}$*, let* $\mathcal{F}_{\mathbb{S}}$ *be as in Construction 5.2.17, we have* $\mathbb{S} \supseteq co_i(\mathcal{F}_{\mathbb{S}})$.

Proof. Assume there is $S \in co_i(\mathcal{F}_{\mathbb{S}})$ such that $S \notin \mathbb{S}$. Let E be a *co*-realization of S in $\mathcal{F}_{\mathbb{S}}$, then by Proposition 5.2.20, there is $T \in \mathbb{S}$ such that $\bigcup_{c_M \in E} M \subseteq T$.

Since E is complete, we have $S = \bigcup_{c_M \in E} M$: Consider some argument $c_M \in E$. By design of $\mathcal{F}_{\mathbb{S}}$, each argument d_M, $d \in M$, possesses the same attacker as c_M thus d_M is defended by E because c_M is defended by E. It is evident that d_M is not attacked by any argument $a \in E$ (otherwise, a attacks c_M); moreover, d_M does not attack any argument $c'_{M'} \in E$ since in this case, E attacks d_M and thus also c_M, contradiction to conflict-freeness of E. By Proposition 5.2.21, it holds that $S = \bigcup_{c_M \in E} M$ contains all arguments $c'_{M'}$ with $c' \in \mathtt{C}_{\mathbb{S}}(S) \setminus S$ and $M' \subseteq \mathtt{C}_{\mathbb{S}}(S)$, hence $S = \mathtt{C}_{\mathbb{S}}(S)$ and thus $S \in \mathbb{S}$. □

Although $\mathcal{F}_{\mathbb{S}}$ possesses characteristics that are necessary for realizing admissible and complete extension-sets, we observe that the construction is not sufficient to express all suitable extension-sets under admissible or complete semantics, respectively:

- $\mathcal{F}_{\mathbb{S}}$ does not realize admissible extension-sets (assuming $\mathbb{S}$ is cautiously closed and contains $\emptyset$): As already mentioned, constructing $\mathcal{F}_{\mathbb{S}}$ might yield additional admissible claim-sets that are not contained in $\mathbb{S}$ (cf. Example 5.2.18, here, $S = \{d_{a,b,d}\}$ is admissible in $\mathcal{F}_{\mathbb{S}}$ but $S \notin \mathbb{S}$).
- $\mathcal{F}_{\mathbb{S}}$ does not realize complete extension-sets (assuming $\mathbb{S}$ is weakly cautiously closed, satisfies unique completion, and contains $\bigcap \mathbb{S}$): While $\mathcal{F}_{\mathbb{S}}$ might produce too many extensions for admissible semantics, the opposite is the case for complete semantics: Let $\mathbb{S} = \{\emptyset, \{a\}, \{b\}, \{a, b, c\}\}$, then $\mathcal{F}_{\mathbb{S}} = (\{a_a, b_b, c_{abc}\}, \emptyset, cl)$ which yields $co_i(\mathcal{F}_{\mathbb{S}}) = \{\{a, b, c\}\}$. Hence the challenge is to separate all complete subsets.

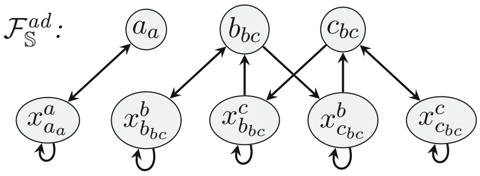
First, we extend $\mathcal{F}_{\mathbb{S}}$ to capture admissible claim-sets.

Construction 5.2.23. *Given a set* $\mathbb{S} \subseteq 2^{\mathsf{C}}$ *and let* $\mathcal{F}_{\mathbb{S}} = (A, R, cl)$ *be defined as in Construction 5.2.17. We define* $\mathcal{F}_{\mathbb{S}}^{ad} = (A^{ad}, R^{ad}, cl^{ad})$ *with*

$$A^{ad} = A \cup \{x^d_{c_M} | c_M \in A, d \in M\},$$
$$R^{ad} = R \cup \{(d_{M'}, x^d_{c_M}), (x^d_{c_M}, x^d_{c_M}), (x^d_{c_M}, c_M) | c_M \in A, d \in M\},$$

and $cl^{ad}(c_M) = cl(c_M) = c$ *for all* $c \in [\mathbb{S}]$ *and* $cl^{ad}(x^d_{c_M}) = x^d_{c_M}$ *otherwise.*

Example 5.2.24. *Let* $\mathbb{S} = \{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\}$*. First, we construct the corresponding CAF* $\mathcal{F}_{\mathbb{S}}$ *that contains no attacks; additionally, we get* $|M|$ *new (self-attacking) arguments for each* $c_M \in A$ *that attack* c_M *and are attacked by each argument having claim* $d \in M$*. The resulting framework is thus given as follows:*



Lemma 5.2.25. *For an extension-set* $\mathbb{S}$*, let* $\mathcal{F}_{\mathbb{S}}^{ad} = (A, R, cl)$ *be defined as in Construction 5.2.23, and let* $E \subseteq A$*. Then*

1. *if an argument* $c_M \in A$ *is defended by* E *then it holds that* $M \subseteq cl(E)$*;*
2. $ad_i(\mathcal{F}_{\mathbb{S}}^{ad}) \subseteq ad_i(\mathcal{F}_{\mathbb{S}})$*;*
3. $E \in ad(F_{\mathbb{S}}^{ad})$ *implies* $cl(E) = \bigcup_{c_M \in E} M$*.*

Proof. (1) follows since only arguments with claim d defend c_M against the attack from $x^d_{c_M}$ for all $d \in M$. To show (2), consider a set $S \in ad_i(\mathcal{F}_{\mathbb{S}}^{ad})$ and an ad-realization E of S in F. Then E defends itself against all attackers in A, thus $S \in ad_i(\mathcal{F}_{\mathbb{S}})$.

For (3), let us first observe that each admissible set $E \in ad(\mathcal{F}_{\mathbb{S}})$ is contained in the union of all minimal sets M that are associated to arguments in E, i.e., $cl(E) \subseteq \bigcup_{c_M \in E} M$. This follows from the fact that $c \in M$ for every argument $c_M \in A$. Moreover, $E \in ad(F_{\mathbb{S}}^{ad})$ implies $E \in ad(F_{\mathbb{S}})$ implies that $cl(E) \subseteq \bigcup_{c_M \in E} M$. By (1) we obtain equality since each argument c_M requires $d \in cl(E)$ for all $d \in M$. □

Proposition 5.2.26. *Let* $\mathbb{S}$ *be cautiously closed and contain* $\emptyset$*. Then* $\mathbb{S} = ad_i(\mathcal{F}_{\mathbb{S}}^{ad})$*.*

Proof. Let $\mathcal{F}_{\mathbb{S}}^{ad} = (A, R, cl)$. We first prove that each set $S \in \mathbb{S}$ is indeed admissible: First, in case $S = \emptyset$ we are done since the empty set is always admissible. Now, let $S \in \mathbb{S}$ be non-empty. We show that $E = \{c_M \in A | M \subseteq S, c \in S\}$ is an admissible realization of

$\mathbb{S}$ in $\mathcal{F}_{\mathbb{S}}^{ad}$. It is easy to see that $cl(E) = S$. Moreover, E is conflict-free since for every two arguments $c_M, c'_{M'} \in E$, it holds that $c \in \mathtt{up}_{\mathbb{S}}(M')$ since $M' \subseteq S \subseteq \mathtt{up}_{\mathbb{S}}(M')$. Moreover, E defends itself: Consider some argument $x \in A$ that attacks an argument $c_M \in E$. In case x is of the form $x^d_{c_M}$, it holds that E defends itself since $M \subseteq S$. In case x is of the form $c'_{M'}$ for some claim c', we proceed analogous as in the proof of Proposition 5.2.19 and obtain that E defends itself against each attack.

For the other direction, consider an admissible set $E \in ad(F_{\mathbb{S}}^{ad})$. We have $cl(E) = \bigcup_{c_M \in E} M$. By Proposition 5.2.20, there is some $S \in \mathbb{S}$ that contains $cl(E)$; since $\mathbb{S}$ is cautiously closed, we obtain that $cl(E) \in \mathbb{S}$ since S serves as witness for $M \in \mathtt{up}_{\mathbb{S}}(M')$ for every sets $M, M' \in \mathbb{S}$ that are associated to arguments in $c_M, c'_{M'} \in E$. ☐

Proposition 5.2.27. *Let $\mathbb{S}$ be cautiously closed and contain $\emptyset$. Then $\mathbb{S} = co_i(\mathcal{F}_{\mathbb{S}}^{ad})$.*

Proof. Let $\mathcal{F}_{\mathbb{S}}^{ad} = (A, R, cl)$. We have shown in Lemma 5.2.25 that each admissible set $S \in \mathcal{F}_{\mathbb{S}}^{ad}$ is realized by $E = \{c_M \in A \mid M \subseteq S\}$. In case E defends some argument $c_M \notin E$, we have $M \not\subseteq S$, that is, there is some argument $x^d_{c_M}$ that attacks c_M and is defended by $d \in M \setminus S$ but not by S. Thus the statement follows. ☐

In case $\mathbb{S}$ is *weakly* cautiously closed we observe that $\mathbb{S} \neq co_i(\mathcal{F}_{\mathbb{S}}^{ad})$: the empty set is complete in $\mathcal{F}_{\mathbb{S}}^{ad}$ since each argument has an attacker; moreover, in case the minimal completion set of $S \cup T$ contains additional arguments for two sets $S, T \in \mathbb{S}$, i.e., in case $\mathsf{C}_{\mathbb{S}}(S \cup T) \notin \{\emptyset, S \cup T\}$, we have that $S \cup T$ is also complete in $\mathcal{F}_{\mathbb{S}}^{ad}$.

In order to deal with this issue, we adapt a concept from [84]. We use defense formulae to determine which arguments are needed to defend a given claim c.

Definition 5.2.28. *For $\mathbb{S} \subseteq 2^{\mathsf{C}}$ and $c \in [\mathbb{S}]$, we let* $\mathrm{def}_{\mathbb{S}}(c) = \{S \cup T \mid S, T \in \mathbb{S}, c \in \mathsf{C}_{\mathbb{S}}(S \cup T) \setminus (S \cup T)\}$. *The* DNF defense formula *of c is defined as $\mathcal{D}_{\mathbb{S}}^{c} = \bigvee_{S \in \mathrm{def}_{\mathbb{S}}(c)} \bigwedge_{d \in S} d$.*

Example 5.2.29. *We consider a set $\mathbb{S} = \{\{a\}, \{a, c\}, \{a, b\}, \{a, b, c, d\}\}$. $\mathbb{S}$ is weakly cautiously closed, moreover, $\bigcap_{S \in \mathbb{S}} S = \{a\}$ is contained in $\mathbb{S}$. We obtain* $\mathrm{def}_{\mathbb{S}}(a) = \mathrm{def}_{\mathbb{S}}(b) = \mathrm{def}_{\mathbb{S}}(c) = \emptyset$ *and* $\mathrm{def}_{\mathbb{S}}(d) = \{\{a, b, c\}\}$. *For a, b, and c, the corresponding DNF formula corresponds to $\bot$; for d, we have $\mathcal{D}_{\mathbb{S}}^{d} = (a \wedge b \wedge c)$.*

We are ready to present the construction for complete semantics.

Construction 5.2.30. *Given a set $\mathbb{S} \subseteq 2^{\mathsf{C}}$ and let $\mathcal{F}_{\mathbb{S}} = (A, R, cl)$ be defined as in Construction 5.2.17. For every argument $c_M \in A$, we consider the extended DNF defense formula $\mathcal{D}_{\mathbb{S}}^{c} \vee \bigwedge_{d \in M} d$ and denote by $\mathcal{CD}_{\mathbb{S}}^{c_M}$ the corresponding CNF formula. We define $\mathcal{F}_{\mathbb{S}}^{co} = (A^{co}, R^{co}, cl^{co})$ as follows*

$$A^{co} = A \cup \{x^{\gamma}_{c_M} | c_M \in A, M \neq \bigcap_{S \in \mathbb{S}} \mathbb{S}, \gamma \in \mathcal{CD}_{\mathbb{S}}^{c_M}\},$$

$$R^{co} = R \cup \{(d_{M'}, x^{\gamma}_{c_M}), (x^{\gamma}_{c_M}, x^{\gamma}_{c_M}), (x^{\gamma}_{c_M}, c_M) | c_M \in A, d \in \gamma\},$$

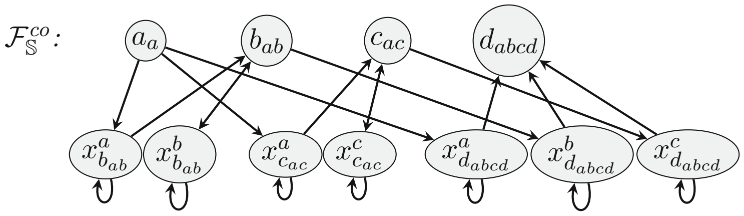
and $cl^{co}(c_M) = cl(c_M) = c$ for all $c \in [\mathbb{S}]$ and $cl^{co}(x^{\gamma}_{c_M}) = x^{\gamma}_{c_M}$ otherwise.

Observe that the grounded extension is realized by arguments that are unattacked in case it is non-empty: auxiliary arguments for an argument C_M are only constructed in case $M \neq \bigcap_{S \in \mathbb{S}} \mathbb{S}$. In every other case, c_M is attacked by argument(s) $x^{\gamma}_{c_M}$ determined by the extended attack formula. Let us consider an example.

Example 5.2.31. *Let us consider the set* $\mathbb{S} = \{\{a\}, \{a, c\}, \{a, b\}, \{a, b, c, d\}\}$ *from Example 5.2.29. First, when constructing* $\mathcal{F}_{\mathbb{S}}$*, we generate four arguments, one for each claim:* a_a*,* b_{ab}*,* c_{ac}*, and* d_{abcd}*. Note that none of these arguments attack each other.*

We proceed by generating the auxiliary arguments: For the claims a*,* b*, and* c*, the DNF defense formula is empty. The extended DNF defense formula for the arguments* a_a*,* b_{ab}*, and* c_{ac} *thus corresponds to the conjunction of the respective sets in the subscript:* $\mathcal{D}^a_{\mathbb{S}} = (a)$*,* $\mathcal{D}^b_{\mathbb{S}} = (a \wedge b)$ *, and* $\mathcal{D}^c_{\mathbb{S}} = (a \wedge c)$*. The corresponding CNF formulae are thus* $\{\{a\}\}$*,* $\{\{a\}, \{b\}\}$*, and* $\{\{a\}, \{c\}\}$*, respectively. For claim* d*, the DNF defense formula is given by* $\mathrm{def}_{\mathbb{S}}(d) = \{\{a, b, c\}\}$*, thus the extended DNF defense formula corresponding to the argument* d_{abcd} *is given by* $\mathcal{D}^d_{\mathbb{S}} \vee \bigwedge_{x \in M} x = (a \wedge b \wedge c) \vee (a \wedge b \wedge c \wedge d)$*. Clearly, this formula can be simplified to the single clause* $(a \wedge b \wedge c)$*. The corresponding CNF is* $\mathcal{CD}^{d_{abcd}}_{\mathbb{S}} = \{\{a\}, \{b\}, \{c\}\}$*.*

We are ready to give the construction. Note that no auxiliary arguments are generated for a_a *since* $\{a\} = \bigcap_{S \in \mathbb{S}} S$*. The resulting CAF is depicted below:*



The argument a_a *is unattacked and does not defend any other argument, thus* $\{a\}$ *is the grounded extension as desired. It can be checked that the complete claim-sets coincide with* $\mathbb{S}$ *(e.g.,* a_a *and* b_{ab} *jointly defend the argument* b_{ab}*).*

Observe that the only difference between $\mathcal{F}^{ad}_{\mathbb{S}}$ *and* $\mathcal{F}^{co}_{\mathbb{S}}$ *for the extension-set* $\mathbb{S}$ *is that* $\mathcal{F}^{ad}_{\mathbb{S}}$ *would contain an additional self-attacking node* $x^d_{d_{abcd}}$ *that attacks and is counter-attacked* d_{abcd}*. In* $\mathcal{F}^{ad}_{\mathbb{S}}$*, the set* $\{a_a, b_{ab}, c_{ac}\}$ *does therefore not defend* d_{abcd}*, consequently,* $\{a, b, c\}$ *is complete in* $\mathcal{F}^{ad}_{\mathbb{S}}$*. In* $\mathcal{F}^{co}_{\mathbb{S}}$*, on the other hand,* d_{abcd} *is defended by* $\{a_a, b_{ab}, c_{ac}\}$ *in* $\mathcal{F}^{co}_{\mathbb{S}}$ *and we obtain* $co_i(\mathcal{F}^{co}_{\mathbb{S}}) = \mathbb{S}$*.*

In case $\mathbb{S}$ is cautiously closed and $\bigcap \mathbb{S} = \emptyset$, we obtain a CAF identical to $\mathcal{F}^{ad}_{\mathbb{S}}$. In this sense the construction refines Construction 5.2.23. We lose a useful property of $\mathcal{F}^{ad}_{\mathbb{S}}$: in $\mathcal{F}^{ad}_{\mathbb{S}}$, each complete set S is realized by $\{c_M \mid M \subseteq S\}$, the extended construction might cause the defense of additional arguments c_M for $M \not\subseteq S$. By Lemma 5.2.20, this affects only arguments c_M such that $c \in S$ and $M \cup S$ possesses a completion-set in $\mathbb{S}$ (all other arguments c_M with claim $c \in S$ are attacked by some arguments in $\{c_M \mid M \subseteq S\}$).

We are ready to prove our last characterization result.

Proposition 5.2.32. *Let $\mathbb{S}$ be weakly cautiously closed, satisfy unique completion and contain $\bigcap \mathbb{S}$. Then $\mathbb{S} = co_i(\mathcal{F}^{co}_{\mathbb{S}})$.*

Proof. We let $\mathcal{F}^{co}_{\mathbb{S}} = (A, R, cl)$; $S \subseteq \mathbb{S}$ and $E' = \{c_M \in A \mid M \subseteq S\}$; moreover, let $E = E' \cup \{c_M \in A \mid c \in S, \mathtt{C}_{\mathbb{S}}(S \cup M) = 1, \exists T, U \subseteq S : c \in \mathtt{C}_{\mathbb{S}}(T \cup U) \setminus (T \cup U)\}$. Observe that E is conflict-free since for every two arguments $c_M, c'_{M'} \in E$ we have $c \in S \subseteq \mathtt{up}_{\mathbb{S}}(M')$ (in case $M' \notin S$ we have $\mathtt{C}_{\mathbb{S}}(S \cup M') = 1$).

Next we show that E defends itself: Consider some argument $x \in A$ that attacks an argument $c_M \in E$. The case x is of the form $c'_{M'}$ for some $c' \in [\mathbb{S}]$ is analogous to the case distinction in the proof of Proposition 5.2.19. In case x is of the form $x^{\gamma}_{c_M}$ and $M \subseteq S$, E defends itself since $\gamma \cap M \neq \emptyset$. In case $M \not\subseteq S$, there are $T, U \subseteq S$ with $c \in \mathtt{C}_{\mathbb{S}}(T \cup U) \setminus (T \cup U)$; by construction of $\mathcal{F}^{co}_{\mathbb{S}}$, $T \cup U \in \mathrm{def}^{c}_{\mathbb{S}}$, we thus obtain $\gamma \cap (T \cup U) \neq \emptyset$. We obtain that E defends itself against all attacker.

Moreover, E contains all arguments it defends: Assume there is an argument $c_M \in A$ that is not contained in E but defended by E. We show that there is $\gamma \in \mathcal{CD}^{c_M}_{\mathbb{S}}$ such that $\gamma \cap S = \emptyset$. It suffices to show that for all $T \in \mathrm{def}_{\mathbb{S}}(c)$, there is $d \in T$ such that $d \notin S$ (we note that by definition of E, we have $M \not\subseteq S$, thus there is a claim $d \in M \setminus S$).

First note that in case $c \in S$ and there is $T \in \mathrm{def}_{\mathbb{S}}(c)$ with $T \subseteq S$ we have $c_M \in E$: By assumption c_M is defended by E we have (1) E does not attack c_M thus $S \subseteq \mathtt{up}_{\mathbb{S}}(M)$ and therefore $\mathtt{C}_{\mathbb{S}}(S \cup M) = 1$ is satisfied; and (2) there are sets $A, B \subseteq S$ with $T = A \cup B$ that defend c.

In case $c \in S$ and there is no $T \in \mathrm{def}_{\mathbb{S}}(c)$ with $T \subseteq S$ we are done: In this case, there is $\gamma \in \mathcal{CD}^{c_M}_{\mathbb{S}}$ such that $\gamma \cap S = \emptyset$ and thus c_M is not defended against the attack $x^{\gamma}_{c_M}$.

Let us now consider the case $c \notin S$. In case there is no $T \in \mathrm{def}_{\mathbb{S}}(c)$ with $T \subseteq S$ we are done: In this case, there is $\gamma \in \mathcal{CD}^{c_M}_{\mathbb{S}}$ such that $\gamma \cap S = \emptyset$ and thus c_M is not defended against the attack $x^{\gamma}_{c_M}$.

In case $c \notin S$ and there is $T \in \mathrm{def}_{\mathbb{S}}(c)$ with $T \subseteq S$. Thus there are sets $A, B \subseteq S$ with $T = A \cup B$ that defend c. Hence $\mathtt{C}_{\mathbb{S}}(A \cup B) \not\subseteq S$ contradiction to unique completion.

For the other direction, consider a set $E \in co(F^{co}_{\mathbb{S}})$. We show that $cl(E) \in \mathbb{S}$. In case $E = \emptyset$, there is no argument in E that is unattacked. By construction of $\mathcal{F}^{co}_{\mathbb{S}}$, this is the case only if $\bigcap_{S \in \mathbb{S}} S = \emptyset$, i.e., if $\emptyset \in \mathbb{S}$.

Now assume $E \neq \emptyset$. It holds that E contains all arguments c_M with $M \subseteq cl(E)$ since each such argument is defended by M. Thus there is some $S \in \mathbb{S}$ such that $cl(E) \subseteq S$ by Lemma 5.2.20. Now assume $cl(E) \notin \mathbb{S}$. In this case, $T = \mathtt{C}_{\mathbb{S}}(\bigcup_{c_M \in E} M)$ is a proper superset of $cl(E)$. Observe that E is not constructed from a single $\subseteq$-minimal set M, i.e., E contains arguments c_M, $c'_{M'}$ with $M \neq M'$ (since no proper subset of such a set M is complete). Now, by design of $\mathcal{F}^{co}_{\mathbb{S}}$, there are sets $U, V \in \mathbb{S}$ with $U, V \subseteq cl(E)$ and there is $c \in T \setminus cl(E)$ such that $U \cup V$ defend all arguments with claim c against the attacks of arguments of the form $x^{\gamma}_{c_M}$ for an arbitrary $\subseteq$-minimal set $M \subseteq T$ containing c. Now, let $M \subseteq \mathtt{C}_{\mathbb{S}}(U \cup V)$ be a $\subseteq$-minimal set in $\mathbb{S}$ that contains c. Then c_M is defended by $U \cup V$ against attacks from arguments in A by Proposition 5.2.21 (since $c \in \mathtt{C}_{\mathbb{S}}(U \cup V) \setminus (U \cup V)$ and $M \subseteq \mathtt{C}_{\mathbb{S}}(U \cup V)$ is satisfied).

Consequently, E defends c_M against all attacks, moreover, $E \cup \{c_M\}$ is conflict-free since $M \subseteq \mathtt{C}_{\mathbb{S}}(U \cup V)$, thus E is not complete in $F^{co}_{\mathbb{S}}$, contradiction to our assumption. ☐

Inherited Naive Semantics

Naive semantics are often perceived as the conflict-free counter-part of preferred semantics as they have many common characteristics. It is thus surprising that the semantics admit several differences when considered with respect to the claims of the arguments. The variants of naive semantics differ even on well-formed CAFs while preferred semantics suggest that maximization on argument-level and maximization on claim-level coincide in this case (recall that both variants of preferred semantics coincide on well-formed CAFs).

We recall that i-naive semantics does not satisfy I-maximality, not even on well-formed CAFs (cf. Example 4.1.2). On the other hand, it is not possible to express all i-maximal extension-sets, as we show next. We first note that, for each well-formed CAF, the set of all (non-self-attacking) occurrences of a claim is contained in some naive argument-extension.

Proposition 5.2.33. *Let $\mathcal{F}$ be a well-formed CAF. Then, for each $c \in \bigcup_{S \in na_i(\mathcal{F})} S$ there is an extension $E \in na(F)$ such that all (non-self-attacking) $a \in A_{\mathcal{F}}$ with $cl(a) = c$ are contained in E.*

Proof. As $c \in \bigcup_{S \in na_i(\mathcal{F})} S$, there is an argument with claim c that is not self-attacking in F. As $\mathcal{F}$ is well-formed, the set $\{a \in A \mid cl(a) = c, (a,a) \notin R\}$ is conflict-free in F and thus contained in some $E \in na(F)$. ☐

Lemma 5.2.34. *For well-formed CAFs, the set $\mathbb{S} = \{\{a,b\},\{a,c\},\{b,c\}\}$ cannot be realized with i-naive semantics, i.e. $\mathbb{S} \notin \Sigma^{wf}_{na_i}$.*

Proof. Towards a contradiction assume there is a CAF $\mathcal{F}$ with $na_i(\mathcal{F}) = \mathbb{S}$. By Proposition 5.2.33 there are sets $E_a, E_b, E_c \in na(\mathcal{F})$ containing all arguments with claim a, b, and c respectively. Let us first assume that all three sets E_a, E_b, E_c are different and have different claim sets, i.e. $cl(E_a)$, $cl(E_b)$, $cl(E_c)$ are mutually distinct. W.l.o.g. we can assume that $cl(E_a) = \{a,b\}$, $cl(E_b) = \{b,c\}$ and $cl(E_c) = \{a,c\}$. That is, (a) there is an argument $b_i \in E_a$ that is not in conflict with any argument with claim a; (b) there is $c_j \in E_b$ that is not in conflict with any argument with claim b; and (c) there is $a_k \in E_c$ that is not in conflict with any argument with claim c. Now consider the set $\{a_k, b_i\}$ which is conflict-free by (a). As $\{a,b,c\} \notin \mathbb{S}$ the set $\{a_k, b_i\}$ has a conflict with c_j. By (c) the conflict has to be between b_i and c_j. However, from (b) we have that c_j is not in conflict with b_i. That is, $\{a_k, b_i, c_j\} \in cf(\mathcal{F})$ and thus $\{a,b,c\} \in na_i(\mathcal{F})$, a contradiction to $na_i(\mathcal{F}) = \mathbb{S}$.

The remaining cases, i.e. (i) E_a, E_b, E_c are different but two of the sets have the same claim-set, and (ii) at least two of the sets E_a, E_b, E_c coincide, can be shown to lead to a contradiction by similar arguments. □

Although i-naive semantics are not I-maximal, it is not possible to express all extension-sets under naive semantics, in particular, it is not possible to express each I-maximal extension-set. This shows that the signatures of i-naive and h-naive semantics are incomparable. As summarized in Table 5.6, i-naive semantics satisfy none of the known principles for AF or CAF semantics. The precise characterization of naive semantics remains an open problem.

5.3 Summary

In this chapter, we studied various principles for claim-focused reasoning and characterized the expressiveness of (almost all) CAF semantics considered in this work.

Our principle-based analysis includes a wide range of genuine principles for claim-focused reasoning. Moreover, we study many well-investigated principles in the context of CAFs. Our results show that well-formed CAFs retain many desired properties like (CF-)reinstatement and I-maximality. Set-theoretical principles like conflict-sensitivity and tightness are, however, violated, which already indicates the higher expressiveness of (well-formed) CAFs when compared to AFs. Our findings moreover reveal that the behavior of CAF semantics with respect to general CAFs is more difficult to capture by means of existing principles; in particular inherited semantics successfully withstand traditional analysis methods. Exceptions are those principles that require the existence of a set of arguments with specific properties (e.g., the defense principle which requires that a set of claims has a realization that defends itself); notable is also the justified rejection principle which is satisfied by stable and conflict-free-based semantics also in the general case. The difficulty indicates that the 'right' principles that characterize the behavior of some of the inherited semantics when considered with respect to general CAFs have yet to be found; we consider this as an important point on our future agenda.

Our signature results confirm that CAF semantics are more expressive than their AF counterpart. In general CAFs, the restrictions are minimal. Indeed, almost each extension-set can be expressed by most of the semantics apart from h-preferred and h-naive which are constrained by I-maximality. This property also characterizes many semantics in well-formed CAFs. We have furthermore identified generalizing properties (i.e., (weak) cautious closure and unique completion) that are characteristic for admissible and complete semantics, respectively, and presented constructions to realize extension-sets confirming to these properties. By doing so, we provide explicit algorithms to construct a (well-formed) CAF that models a desired situation. Moreover, our signature results can prove useful when considering changes in argumentation frameworks or their underlying knowledge bases following certain constraints since expressiveness characterizations are the basis for certain (im-)possibility results regarding changes of the extensions (cf. [27]).

CHAPTER 6

Computational Complexity of Conclusion-focused Reasoning

In this chapter, we analyze the computational complexity of claim-focused reasoning and analyze the different variants of CAF semantics from a complexity-theoretic perspective. We study decision problems which are well-established in the field of computational argumentation with respect to hybrid semantics and consider computational aspects of comparing different variants of claim semantics. Our results complement and extend the complexity analysis for inherited CAF semantics presented in [92].

While a semantics returns a set of a set of extensions, one is often interested in the acceptability of a particular argument or claim. Two commonly studied reasoning modes which specify the acceptance status are *credulous* and *skeptical acceptance* [89]. Following [92], we adapt them to the realm of claim-focused reasoning as follows: a claim is credulously accepted with respect to a particular semantics if it is contained in some extension; likewise, it is skeptically accepted if it is contained in each extension. We also study the computational complexity of *verifying* that a given claim-set is realizable with respect to a given semantics. Moreover, we consider the *non-emptiness problem*: although most of the semantics considered in this work are guaranteed to return *some* extension it is not clear if the extension is 'reasonable' in the sense that no argument might be accepted. For instance, the empty set can be preferred in case no other admissible extension exists. We study the computational complexity of deciding whether a given CAF has at least one extension which is non-empty. We furthermore investigate the well-known *coherence problem* for hybrid semantics, i.e., we investigate the complexity of deciding whether h-stable and h-preferred semantics coincide. Finally, we investigate the differences of the inherited and hybrid semantics from a computational point of view. To do so, we consider the problem of *concurrence* of inherited and hybrid semantics, i.e., we study the computational complexity of deciding whether the hybrid and the inherited variant of a semantics yield the same extensions in a given CAF.

Among our findings is that the complexity of the verification problem with respect to all considered cl-semantics as well as the skeptical acceptance problem for h-naive semantics admit a higher complexity than for the AF semantics counter-part. A similar behavior has been observed in [92] for inherited semantics. We identify notable differences but also similarities between hybrid and inherited semantics. On the one hand, we show that the skeptical acceptance problem with respect to h-naive semantics admit a higher complexity than for i-naive semantics; moreover, the complexity of verifying a claim-set with respect to preferred semantics drops while the complexity rises for naive semantics when compared to their inherited counter-parts; on the other hand, both variants admit the same complexity for well-formed CAFs. Furthermore, it can be surprisingly hard to decide concurrence, ranging up to the third level of the polynomial hierarchy.

Background & Notation. *We make use of concepts and results from Chapter 4.*

6.1 Complexity Theory: A Brief Introduction

We give a brief summary of the basic terms used in this work; for a comprehensive overview we refer to, e.g., [9, 145].

The analysis of the computational complexity of a problem is concerned with, roughly speaking, grouping problems into different categories to make them comparable in terms of computational difficulty. Complexity classes impose bounds on certain resources, e.g., time, usually measured in the number of steps required by a machine, or space. We consider a formal model of computation, e.g., Turing machines[1], a suitable representation of objects (encoding), and a unified formulation in terms of *decision problems*.

Formally, objects (problem instances) are encoded as strings $x \in A^*$ over a finite alphabet A. For instance, a CAF $\mathcal{F}$ can be represented as the concatenation of two binary strings, one encodes the adjacency matrix of (A, R) and the other encodes the function mapping arguments to claims. We usually abstract from the particular representation and measure the input size in terms of the factor determining the length of the string instead. In the context of the decision problems we study, the input is often a CAF, and the size of the CAF depends polynomially on the number of arguments (i.e., a CAF (A, R, cl) has at most $|A|^2$ attacks).

A *decision problem* Q separates instances of the set of all possible input objects D_Q into two disjoint classes: *positive* instances, i.e., those instances returning ‘yes’ (True, or 1) and *negative* instances which return the answer ‘no’ (False, or 0). We identify a decision problem Q with the set of all positive instances, i.e., $x \in Q$ iff x is a yes-instance of Q. That is, a decision problem is a subset $Q \subseteq D_Q$ of a set of input objects D_Q. By $|x|$ we denote the length of the instance $x \in Q$.

For instance, to investigate credulous reasoning with respect to a particular semantics ρ we ask, for a given CAF $\mathcal{F}$ and a claim c: *Is there some set $S \in \rho(\mathcal{F})$ such that $c \in S$?*

---

[1] There are many different formulations of computational models, most of them equivalent to each other with respect to the problems we consider in this work.

Here, the tuple $(\mathcal{F}, c)$ belongs to the set of input objects; it is a positive instance if the answer to this question is 'yes', i.e., if there is some $S \in \rho(\mathcal{F})$ such that $c \in S$, and a negative instance otherwise. We formulate the problem as follows:

$Cred_{\rho}^{CAF}$

Input: A CAF $\mathcal{F} = (A, R, cl)$ and a claim $c \in cl(A)$.

Output: TRUE iff there is a claim-set $S \in \rho(\mathcal{F})$ with $c \in S$.

We measure the running time of a decision problem in terms of counting the number of steps which are required to solve the problem in dependency of the input.

Definition 6.1.1. *A decision problem Q is solvable in* polynomial time *iff there is a Turing machine T and a polynomial p such that T stops after at most $p(|x|)$ steps and returns 'yes' if $x \in Q$, or 'no' if $x \notin Q$.*

By P we denote the class of all decision problems solvable in polynomial time. We also say that Q is tractable if $Q \in \mathsf{P}$.

Let us introduce the class of problems solvable in non-deterministic polynomial time.

Definition 6.1.2. *A decision problem Q is solvable in* non-deterministic polynomial time *iff there is a Turing machine T and a polynomial p such that, for each $x \in Q$, there exists a* certificate *$y \in C_Q$, for a set C_Q of finite objects, with $|y| < p(|x|)$ s.t. T stops after at most $p(|(x, y)|)$ steps on input (x, y) and returns 'yes' if $x \in Q$, or 'no' if $x \notin Q$.*

By NP we denote the class of all decision problems solvable in non-deterministic polynomial time. The canonical decision problem for the class NP is the *boolean satisfiability problem*.

SAT

Input: A boolean formula φ.

Output: TRUE iff φ is satisfiable.

NP-membership is often verified via so-called *guess-and-check procedures*: given an instance $x \in D_Q$, we *guess* a potential certificate $y \in C_Q$; and provide an algorithm to *check* in polynomial time whether y is a certificate for x. For instance, a guess-and-check procedure for SAT is as follows: given a boolean formula φ over atoms in X, we guess an interpretation $\mathcal{I}$ of φ and check (in P) whether φ is true under $\mathcal{I}$.

We furthermore consider the complement of a complexity class.

Definition 6.1.3. *For a complexity class C over problems in A^*, we denote by $\mathsf{coC} = \{A^* \setminus Q \mid Q \in \mathsf{C}\}$ the set of complementary problems of C.*

This gives rise to the class coNP which contains all problems complementary to NP. A classical decision problem for this class is the *unsatisfiability problem*.

$$\begin{tikzcd}
 & \mathsf{NP} = \Sigma_1^{\mathsf{P}} \arrow[dr] & & \Sigma_2^{\mathsf{P}} \arrow[dr] & & \Sigma_3^{\mathsf{P}} \arrow[dr] & & \Sigma_4^{\mathsf{P}} & \\
\mathsf{P} = \Delta_1^{\mathsf{P}} \arrow[ur] \arrow[dr] & & \Delta_2^{\mathsf{P}} \arrow[ur] \arrow[dr] & & \Delta_3^{\mathsf{P}} \arrow[ur] \arrow[dr] & & \Delta_4^{\mathsf{P}} \arrow[ur] \arrow[dr] & & \cdots \\
 & \mathsf{coNP} = \Pi_1^{\mathsf{P}} \arrow[ur] & & \Pi_2^{\mathsf{P}} \arrow[ur] & & \Pi_3^{\mathsf{P}} \arrow[ur] & & \Pi_4^{\mathsf{P}} &
\end{tikzcd}$$

Figure 6.1: Relations between complexity classes. Arrows denote set inclusion.

UNSAT

Input: A boolean formula φ.
Output: TRUE iff φ is unsatisfiable.

The problem is equivalent to the tautology problem TAUT which takes as input a boolean formula and returns TRUE iff φ is a tautology, i.e., satisfied under each assignment.

We furthermore recall the complexity class DP which contains all problems which are the intersection of a problem in NP and coNP.

Definition 6.1.4. $\mathsf{DP} = \mathsf{NP} \wedge \mathsf{coNP} = \{A \cap B \mid A \in \mathsf{NP}, B \in \mathsf{coNP}\}$.

A canonical problem for the class DP is SAT-UNSAT:

SAT-UNSAT

Input: A tuple (φ, ψ) of two boolean formulae.
Output: TRUE iff φ is valid and ψ is not valid.

The polynomial hierarchy. Let us recall the notion of an *oracle*. Consider a complexity class C. A C-*oracle* decides a problem Q from C in one computation step. It can be seen as black box which returns the correct answer when asked whether $x \in Q$.

Given a complexity class B, we let B^{C} denote the class of problems that can be decided in B with access to a C-oracle. For instance, the class P^{NP} is the class of all problems which can be decided in polynomial time with polynomially many calls to an NP-oracle. This gives rise to the *polynomial hierarchy* by setting $\Delta_0^{\mathsf{P}} = \Sigma_0^{\mathsf{P}} = \Pi_0^{\mathsf{P}} = \mathsf{P}$; $\Delta_{k+1}^{\mathsf{P}} = \mathsf{P}^{\Sigma_k^{\mathsf{P}}}$; $\Sigma_{k+1}^{\mathsf{P}} = \mathsf{NP}^{\Sigma_k^{\mathsf{P}}}$; and $\Pi_{k+1}^{\mathsf{P}} = \mathsf{NP}^{\Pi_k^{\mathsf{P}}}$.

Note that $\Sigma_1^{\mathsf{P}} = \mathsf{NP}$, $\Pi_1^{\mathsf{P}} = \mathsf{coNP}$, $\Delta_1^{\mathsf{P}} = \mathsf{P}$, and $\Delta_2^{\mathsf{P}} = \mathsf{P}^{\mathsf{NP}}$. Moreover, $\Sigma_k^{\mathsf{P}} = \mathsf{co}\Pi_k^{\mathsf{P}}$. Figure 6.1 depicts the known relations between the classes.

We recall the canonical problems for the classes. For this, we consider *quantified boolean formulae (QBF)* which are of the form

$$Q_1 X_1 \ldots Q_k X_k \varphi(X_1, \ldots, X_k)$$

where $X_1, \ldots, X_k$ are propositional atoms and $Q_i \in \{\forall, \exists\}$, moreover, the quantifiers are alternating, i.e., $Q_i \neq Q_{i+1}$ for all $i \leq k$.

$\mathsf{QSAT}_k^{\forall}$

Input: A QBF $\Psi = Q_1X_1 \ldots Q_kX_k\varphi(X_1, \ldots, X_k)$ with $Q_i \neq Q_{i+1}$ for all $i \leq k$ and $Q_1 = \forall$.

Output: TRUE iff Ψ is valid.

$\mathsf{QSAT}_k^{\exists}$

Input: A QBF $\Psi = Q_1X_1 \ldots Q_kX_k\varphi(X_1, \ldots, X_k)$ with $Q_i \neq Q_{i+1}$ for all $i \leq k$ and $Q_1 = \exists$.

Output: TRUE iff Ψ is valid.

The problem $\mathsf{QSAT}_k^{\forall}$ is canonical for $\Pi_\mathsf{k}^\mathsf{P}$, and $\mathsf{QSAT}_k^{\exists}$ is canonical for $\Sigma_\mathsf{k}^\mathsf{P}$.

Completeness and polynomial reductions. We have defined several complexity classes with (presumably) rising complexity. We end this brief overview with the notions of *completeness* and *polynomial-time reductions* of problems.

Definition 6.1.5. *A decision problem Q' is polynomial time reducible to problem Q ($Q' \leq_P Q$), iff there is a polynomial time computable function R s.t. $x \in Q'$ iff $R(x) \in Q$.*

Definition 6.1.6. *A decision problem Q is C-hard iff $Q' \leq_P Q$ for each $Q' \in \mathsf{C}$. Q is C-complete iff it is C-hard and in C.*

C-completeness indicates that the problem is among the hardest problems for the class C.

The problem SAT is NP-complete, that is, each problem in NP can be reduced to SAT. Likewise, UNSAT is coNP-complete; moreover, the problem $\mathsf{QSAT}_k^{\forall}$ is $\Pi_\mathsf{k}^\mathsf{P}$-complete if k is even, i.e., for each $k = 2n$ where $n \geq 1$, $n \in \mathbb{N}$; and $\mathsf{QSAT}_k^{\exists}$ is $\Sigma_\mathsf{k}^\mathsf{P}$-complete if k is odd, i.e., for each $k = 2n + 1$ where $n \geq 0$, $n \in \mathbb{N}$.

The strategy to prove C-completeness of a decision problem Q consists of two parts: verifying membership, that is, presenting a procedure to verify that the problem lies in C; and proving C-hardness of the problem by reducing a C-complete problem to Q.

Remark 6.1.7. *We use the following conventions. First, we assume that each formula φ appearing as input of a decision problem is in conjunctive normal form (CNF) (unless stated otherwise). We remark that the restricted problem has the same complexity as* SAT. *Second, we identify the clauses with sets, e.g., we identify clause $(x_1 \wedge x_2 \wedge \neg y)$ with the set of literals $\{x_1, x_2, \neg y\}$. Third, we identify models of a formula φ with the set of atoms assigned to true in the model. That is, for a satisfying assignment $\mathcal{I}$ of φ over atoms in X, we call the set $X' = \{x \in X \mid \mathcal{I}(x) = true\}$ a model of φ. Moreover, we sometimes use $\bar{x}$ to denote $\neg x$ for an atom x.*

6.2 Complexity of Reasoning Problems

In this section, we study the computational complexity of claim-focused reasoning with respect to our novel hybrid semantics. We consider *credulous* and *skeptical acceptance*, i.e., deciding whether a given claim is contained in at least one or all extensions under a given semantics, respectively; *verifying* that a given set of claims is a ρ-claim-set for a given semantics ρ; and deciding whether a *non-empty extension* exists (with respect to a given semantics). We define them as follows (with respect to a CAF semantics ρ):

- *Credulous Acceptance* ($Cred_\rho^{CAF}$): Given a CAF $\mathcal{F}$ and a claim $c \in cl(A_\mathcal{F})$, is c contained in some $S \in \rho(\mathcal{F})$?
- *Skeptical Acceptance* ($Skept_\rho^{CAF}$): Given a CAF $\mathcal{F}$ and claim $c \in cl(A_\mathcal{F})$, is c contained in each $S \in \rho(\mathcal{F})$?
- *Verification* (Ver_ρ^{CAF}): Given a CAF $\mathcal{F}$ and a set $S \subseteq cl(A_\mathcal{F})$, is $S \in \rho(\mathcal{F})$?
- *Non-emptiness* (NE_ρ^{CAF}): Given a CAF $\mathcal{F}$, is there a non-empty set $S \subseteq cl(A_\mathcal{F})$ such that $S \in \rho(\mathcal{F})$?

We furthermore consider these reasoning problems restricted to well-formed CAFs and denote them by $Cred_\rho^{wf}$, $Skept_\rho^{wf}$, Ver_ρ^{wf}, and NE_ρ^{wf}. Moreover, we denote the corresponding decision problems for AFs (which can be obtained by defining cl as the identity function) by $Cred_\sigma^{AF}$, $Skept_\sigma^{AF}$, Ver_σ^{AF}, and NE_σ^{AF}.

For AF and inherited CAF semantics, the computational complexity of these problems has been already established. Tables 6.1 and 6.2 depict known complexity results for Dung AF semantics [74, 81, 90, 89]; and for inherited CAF semantics [92, 117], respectively.

The forthcoming analysis yields the following high level picture: Credulous and skeptical reasoning as well as deciding existence of a non-empty extension under hybrid semantics is

σ	$Cred_\sigma^{AF}$	$Skept_\sigma^{AF}$	Ver_σ^{AF}	NE_σ^{AF}
cf	in P	trivial	in P	in P
ad	NP-c	trivial	in P	NP-c
co	NP-c	in P	in P	NP-c
gr	in P	in P	in P	in P
stb	NP-c	coNP-c	in P	NP-c
na	in P	in P	in P	in P
pr	NP-c	Π_2^P-c	coNP-c	NP-c
ss	Σ_2^P-c	Π_2^P-c	coNP-c	NP-c
stg	Σ_2^P-c	Π_2^P-c	coNP-c	in P

Table 6.1: Computational complexity of reasoning with respect to AF semantics.

ρ	$Cred_\rho^\Delta$	$Skept_\rho^\Delta$	Ver_ρ^{CAF} / Ver_ρ^{wf}	NE_ρ^Δ
cf_i	in P	trivial	**NP**-c / in P	in P
ad_i	NP-c	trivial	**NP**-c / in P	NP-c
co_i	NP-c	in P	**NP**-c / in P	NP-c
gr_i	in P	in P	in P	in P
stb_c	NP-c	coNP-c	**NP**-c / in P	NP-c
na_i	in P	**coNP**-c	**NP**-c / in P	in P
pr_i	NP-c	Π_2^P-c	$\boldsymbol{\Sigma}_2^{\mathbf{P}}$-c / coNP-c	NP-c
ss_i	Σ_2^P-c	Π_2^P-c	$\boldsymbol{\Sigma}_2^{\mathbf{P}}$-c / coNP-c	NP-c
stg_i	Σ_2^P-c	Π_2^P-c	$\boldsymbol{\Sigma}_2^{\mathbf{P}}$-c / coNP-c	in P

Table 6.2: Computational complexity results for inherited semantics, with $\Delta \in \{CAF, wf\}$. Results that deviate from the corresponding results for AFs are bold-face.

of the same complexity as in AFs except for the notable difference that skeptical reasoning with respect to h-naive semantics goes up two levels in the polynomial hierarchy and is thus also more expensive than deciding skeptical acceptance for i-naive semantics which has been shown to be coNP-complete. For well-formed CAFs, skeptical reasoning admits the same complexity for both hybrid and inherited naive semantics but remains more expensive than in AFs.

For general CAFs, the verification problem is more expensive than for AFs for all of the considered semantics. Comparing hybrid and inherited semantics we observe that the complexity of the verification problem for h-preferred semantics drops while the complexity for h-naive semantics admits a higher complexity than their inherited counterparts; the hybrid and inherited variants of stable, semi-stable and stage semantics admit the same complexity. For well-formed CAFs, the complexity of the verification problem coincides with the known results for AFs.

6.2.1 Complexity Results for General CAFs

We start our analysis with general CAFs. First, we discuss upper bounds before we present hardness results yielding the corresponding lower bounds for the decision problems. An overview of our results is given in Table 6.3.

Remark 6.2.1. *The complexity results for h-semi-stable and h-stage semantics have been settled in the scope of Alexander Gressler's Masters Thesis [117]. We include them in order to provide a complete picture of the complexity of claim-focused reasoning.*

Membership Results. We will first discuss the membership proofs. To begin with, we will give poly-time respectively coNP procedures for deciding whether a given set of arguments E is a ρ-realization for $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h, ss_h, stg_h\}$. This lemma yields upper bounds for the respective reasoning problems; notice that the complexity goes up one level in the polynomial hierarchy since one requires an additional guess for E.

Lemma 6.2.2. *Given a CAF $\mathcal{F}$ and some $E \subseteq A_{\mathcal{F}}$. Deciding whether E realizes (1) a τ-h-stable claim-set in $\mathcal{F}$ for $\tau \in \{ad, cf\}$ is in* P*; (2) a h-semi-stable (h-stage) claim set in $\mathcal{F}$ is in* coNP*.*

Proof. Checking admissibility (conflict-freeness) of E is in P (cf. Table 6.1); moreover, $E^*_{\mathcal{F}}$ can be computed in polynomial time by looping over all claims $c \in cl(A_{\mathcal{F}})$ and adding each c to $E^*_{\mathcal{F}}$ if E attacks each occurrence of c in $\mathcal{F}$. For τ-h-stable semantics, it remains to check whether $cl(E) \cup E^*_{\mathcal{F}} = cl(A_{\mathcal{F}})$. For h-semi-stable (h-stage) semantics, we have to check that each $E' \subseteq A_{\mathcal{F}}$ with $cl(E') \cup {E'}^*_{\mathcal{F}} \supset cl(E) \cup E^*_{\mathcal{F}}$ is not admissible (conflict-free). This can be solved in coNP by a standard guess & check algorithm, i.e. guess a set and verify that it is admissible (conflict-free), compute the claims and verify that they are a proper superset of the claims of the original set, yielding a coNP algorithm to verify that E realizes a h-semi-stable (h-stage) claim-set in $\mathcal{F}$. ☐

Proposition 6.2.3. *The following membership results hold for the verification problem:*

1. Ver^{CAF}_{ρ} *is in* NP *for* $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h\}$*,*
2. Ver^{CAF}_{ρ} *is in* Σ^{P}_2 *for* $\rho \in \{ss_h, stg_h\}$*,*
3. Ver^{CAF}_{ρ} *is in* DP *for* $\rho \in \{pr_h, na_h\}$*.*

Proof. Consider a CAF $\mathcal{F} = (A, R, cl)$ and a set $S \subseteq cl(A)$ that has to be verified against a semantics ρ. 1 & 2) Here we can apply a guess and check algorithm. That is, one can verify $S \in \rho(\mathcal{F})$ by guessing a set of arguments $E \subseteq A$ with $cl(E) = S$ and checking whether E is a ρ-realization of S. The latter is in P, respectively coNP by Lemma 6.2.2, yielding NP- and Σ^{P}_2-procedures for the respective semantics.

3) DP-membership of Ver^{CAF}_{ρ} for $\rho \in \{pr_h, na_h\}$ is by (a) checking that a given claim-set S is admissible (conflict-free) and (b) verifying subset-maximality of S. The former has been shown to be NP-complete (cf. Table 6.2); the latter is in coNP: Guess a set of arguments E such that $S \subset cl(E)$ and check admissibility (conflict-freeness) of E. Thus Ver^{CAF}_{ρ} can be represented as the intersection of a NP-complete problem and a problem in coNP and lies therefore in DP. ☐

We next turn the reasoning problems, starting with the skeptical acceptance problem.

Proposition 6.2.4. *The following membership results hold for skeptical acceptance:*

1. $Skept^{CAF}_{\rho}$ *is in* coNP *for* $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h\}$*,*
2. $Skept^{CAF}_{\rho}$ *is in* Π^{P}_2 *for* $\rho \in \{pr_h, na_h, ss_h, stg_h\}$*.*

Proof. Membership proofs for $Skept_\rho^{CAF}$ are by standard guess-and-check-algorithms for the the complementary problem: For a CAF $\mathcal{F} = (A, R, cl)$ and claim $c \in cl(A)$, guess a set $E \subseteq A$ such that $c \notin cl(E)$ and check $cl(E) \in \rho(\mathcal{F})$. 1) For $\rho \in \{\tau\text{-}stb_h\}$ the latter can be verified in P by Lemma 6.2.2, which yields coNP-membership; 2) By the same lemma, testing for $\sigma \in \{ss_h, stg_h\}$, is coNP, thus showing Π_2^P-membership; for $\rho \in \{pr_h, na_h\}$, we use the result for Ver_ρ^{CAF}, i.e., $cl(E) \in \rho(\mathcal{F})$ can be verified via two NP-oracle calls, which shows that $Skept_\rho^{CAF}$ is in Π_2^P. ☐

Proposition 6.2.5. *The following membership results hold for credulous acceptance:*

1. $Cred_\rho^{CAF}$ *is in* P *for* $\rho \in \{na_h\}$,
2. $Cred_\rho^{CAF}$ *is in* NP *for* $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h, pr_h\}$,
3. $Cred_\rho^{CAF}$ *is in* Σ_2^P *for* $\rho \in \{ss_h, stg_h\}$.

Proof. Membership for $Cred_\rho^{CAF}$ and $\rho \in \{\tau\text{-}stb_h, ss_h, stg_h\}$ are by standard guess-and-check-algorithms: For a CAF $\mathcal{F} = (A, R, cl)$ and claim $c \in cl(A)$, guess a set $E \subseteq A$ such that $c \in cl(E)$ and check $cl(E) \in \rho(\mathcal{F})$. For h-preferred and h-naive semantics, we exploit the fact a claim $c \in cl(A)$ is credulously accepted with respect to h-preferred (h-naive) semantics iff it is contained in some i-admissible (i-conflict-free) claim-set and thus the complexity of $Cred_\theta^{CAF}$ for $\theta \in \{cf_i, ad_i\}$ (cf. Table 6.2) applies. ☐

Proposition 6.2.6. *The following membership results hold for the non-empty problem:*

1. NE_ρ^{CAF} *is in* P *for* $\rho \in \{na_h, stg_h\}$;
2. NE_ρ^{CAF} *is in* NP *for* $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h, pr_h, ss_h\}$.

Proof. NE_ρ^{CAF} for $\rho \in \{ss_i, stg_i, pr_h, na_h,\ ss_h, stg_h\}$ can be reduced to the respective problem for AFs: for h-preferred (h-naive) semantics and both variants of semi-stable (stage) semantics, we have that a CAF has a non-empty claim-set iff a non-empty admissible (conflict-free) set of argument exists, i.e., NE_ρ^{CAF} $\rho \in \{pr_h, ss_h, na_h, stg_h\}$, coincides with either NE_{ad}^{AF} or NE_{cf}^{AF} and we get the complexity directly from Table 6.1. For $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h\}$, NE_ρ^{CAF} can be verified by guessing a non-empty set $E \subseteq A$ and utilizing Lemma 6.2.2 (1) for checking that $cl(E)$ is a τ-h-stable claim-set of $\mathcal{F}$. ☐

Hardness Results. We now turn to the hardness results for the considered decision problems. First observe that one can reduce AF decision problems to the corresponding problems for CAFs by assigning each argument a unique claim. Thus CAF decision problems generalize the corresponding problems for AFs and are therefore at least as hard. It remains to provide hardness proofs for the decision problems with higher complexity. Hence it remains to show hardness for $Skept_{na_h}^{CAF}$ and the verification problems Ver_ρ^{CAF} for all semantics ρ under consideration.

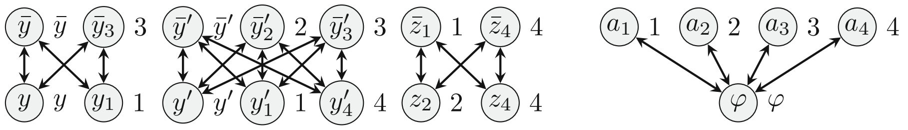


Figure 6.2: Example CAF from the proof of Proposition 6.2.10 (Reduction 6.2.7) for the formula $\forall yy'\exists z\varphi$, where φ is given by clauses $\{\{y, y', \neg z\}, \{\neg y', z\}, \{\neg y, \neg y'\}, \{y', z, \neg z\}\}$.

We will first present a reduction from $\mathsf{QSAT}_2^\forall$ to show Π_2^P-hardness of $Skept_{na_h}^{CAF}$ before we address the verification problems. In this reduction, starting from a QBF $\Psi = \forall Y \exists Z \varphi(Y, Z)$ where φ is a 3-CNF given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $X = Y \cup Z$, we construct a CAF as follows (cf. Figure 6.2):

- For each clause cl_i, we introduce three arguments representing the literals contained in cl_i and assign them claim i;
- we add arguments representing literals over Y and assign them unique claims;
- furthermore, we add arguments $a_1, \dots, a_n$ with claims $1, \dots, n$ and an argument φ with unique claim φ;
- we introduce conflicts between each argument representing a variable $x \in X$ and arguments representing its negation; moreover, we add symmetric attacks between φ and each argument a_i.

Reduction 6.2.7. *Let $\Psi = \forall Y \exists Z \varphi(Y, Z)$ be an instance of $\mathsf{QSAT}_2^\forall$, where φ is a 3-CNF given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $X = Y \cup Z$. We let $\bar{Y} = \{\bar{y} \mid y \in Y\}$ and construct a CAF $\mathcal{F} = (A, R, cl)$ as follows (cf. Figure 6.2):*

$$\begin{aligned} A &= \{x_i \mid x \in cl_i, i \le n\} \cup \{\bar{x}_i \mid \neg x \in cl_i, i \le n\} \ \cup Y \cup \bar{Y} \cup \{a_1, \dots, a_n, \varphi\} \\ R &= \{(a_i, \varphi), (\varphi, a_i) \mid i \le n\} \cup \{(x_i, \bar{x}_j)(\bar{x}_j, x_i), \mid i, j \le n\} \cup \\ &\quad \{(y, \bar{y}_i), (\bar{y}_i, y), (y_i, \bar{y}), (\bar{y}, y_i), (y, \bar{y}), (\bar{y}, y) \mid y \in Y\} \end{aligned}$$

where $cl(x_i) = cl(\bar{x}_i) = cl(a_i) = i$, $cl(y) = \ y$, $cl(\bar{y}) = \bar{y}$, and $cl(\varphi) = \varphi$.

We will show that Ψ is valid iff the claim φ is skeptically accepted with respect to h-naive semantics in $\mathcal{F}$. The main observation is that for every $Y' \subseteq Y$, the set of arguments $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{a_1, \dots, a_n\}$ is conflict-free, thus $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n\} \in cf_i(\mathcal{F})$. Consequently, φ is skeptically accepted with respect to h-naive semantics iff for every $Y' \subseteq Y$, the set $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n, \varphi\}$ is h-naive. It suffices to check that for every $Y' \subseteq Y$, the set $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n, \varphi\}$ is h-naive iff there is $Z' \subseteq Z$ such that $Y' \cup Z'$ is a model of φ. This is addressed in the following lemma.

Lemma 6.2.8. *For every $Y' \subseteq Y$, $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n, \varphi\} \in na_h(\mathcal{F})$ iff there is $Z' \subseteq Z$ such that $M = Y' \cup Z'$ is a model of φ.*

Proof. Let $S = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n, \varphi\}$.

First assume $S \in na_h(\mathcal{F})$. Consider a cf_i-realization E of S. We have $\varphi \in E$ because φ is the unique argument having claim φ. Consequently, $a_i \notin E$ and thus each claim i is represented by x_i for some $x \in X \cup \bar{X}$. Let $Z' = \{z \in Z \mid z_i \in E\}$. Then $M = Y' \cup Z'$ is a model of φ: Consider an arbitrary clause cl_i. Since $\{1, \dots, n\} \subseteq S$, there is some argument with claim i in E, that is, either $a_i \in E$ or $x_i \in E$ or $\bar{x}_i \in E$ for some $x \in X$ (observe that $y_i \in E$ iff $y \in E$ and $\bar{y}_i \in E$ iff $\bar{y} \in E$, thus a further case distinction for $y \in Y$, $\bar{y} \in \bar{Y}$ is not required). We have that $a_i \notin E$ since $n \in S$ and for each argument b with $cl(b) = n$ we have $(a_i, b) \in R$. Thus there is $x \in X$ such that either $x_i \in E$ or $\bar{x}_i \in E$. In the former case, we have $x \in M$ and thus M satisfies cl_i, in the latter case $x \notin M$ and thus cl_i is satisfied. We obtain that M is a model of φ.

Now assume there is $Z' \subseteq Z$ such that $M = Y' \cup Z'$ is a model of φ. Let $E = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{x_i \mid x \in M\} \cup \{\bar{x}_i \mid x \notin M\} \cup \{\varphi\}$. E is conflict-free since $a_i \notin E$ for all $i < n$; other conflicts appear only between arguments x_i, $\bar{x}_j$ referring to the same atom x. Moreover, as M is a model of φ, we have that for each clause cl_i, there is either a positive literal $x \in cl_i$ with $x \in M$ or a negative literal $\bar{x} \in cl_i$ with $x \notin M$. Thus $\{1, \dots, n\} \subseteq cl(E)$; moreover, $Y' \cup \{\bar{y} \mid y \notin Y'\} \subseteq cl(E)$ and therefore $cl(E) = S$. S is a maximal h-conflict-free claim-set since $S \cup \{c\} \notin cf_i(\mathcal{F})$ for any $c \in (Y \cup \bar{Y}) \setminus S$ as each realization of $S \cup \{c\}$ contains y, $\bar{y}$ for some $y \in Y$. Thus $S \in na_h(\mathcal{F})$. □

We are now ready to prove the correctness of the reduction.

Lemma 6.2.9. *The formula Ψ is valid iff the claim n is skeptically accepted with respect to h-naive semantics in $\mathcal{F}$.*

Proof. Assume Ψ is not valid. Then there is $Y' \subseteq Y$ such that for all $Z' \subseteq Z$, $M = Y' \cup Z'$ does not satisfy φ. Let $S = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n\}$. Observe that S is i-conflict-free, witnessed by the cf_i-realization $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{a_1, \dots, a_n\}$. S is h-naive since $S \cup \{\varphi\} \notin cf_i(\mathcal{F})$ by (1) and $S \cup \{c\} \notin cf_i(\mathcal{F})$ for any $c \in (Y \cup \bar{Y}) \setminus S$ as each realization of $S \cup \{c\}$ contains y, $\bar{y}$ for some $y \in Y$. Thus φ is not skeptically accepted with respect to h-naive semantics in $\mathcal{F}$.

Assume φ is not skeptically accepted with respect to h-naive semantics in $\mathcal{F}$. Then there is a set $S \in na_h(\mathcal{F})$ such that $\varphi \notin S$. Observe that S contains $Y' \cup \{\bar{y} \mid y \notin Y\}$ for some $Y' \subseteq Y$ by construction. Let $Y' = S \cup Y$. We show that for all $Z' \subseteq Z$, $Y' \cup Z'$ is not a model of φ: Towards a contradiction assume there is $Z' \subseteq Z$ such that $M = Y' \cup Z'$ is a model of φ. By (1), $T = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n, \varphi\} \in na_h(\mathcal{F})$. Thus $T \supset S$ since $\varphi \notin S$, contradiction to S being h-naive in $\mathcal{F}$. It follows that Ψ is not valid. □

By the above lemma and the fact that the reduction can be performed in polynomial time we obtain Π_2^P-hardness.

Proposition 6.2.10. *$Skept_{na_h}^{CAF}$ is Π_2^P-hard.*

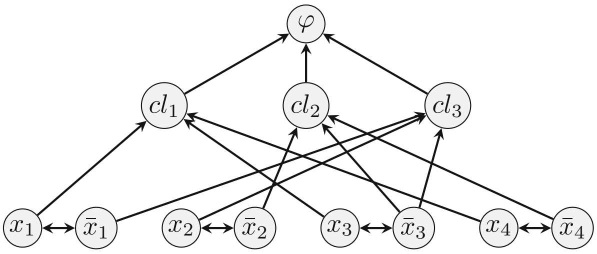


Figure 6.3: An AF constructed from Reduction 6.2.11 for a formula φ with clauses $\{\{x_1, x_3, x_4\}, \{\neg x_3, \neg x_4, \neg x_2)\}, \{\neg x_1, \neg x_3, x_2\}\}$.

Hardness results for verification problems admit a higher complexity compared to AFs for all of the considered semantics. DP-hardness with respect to h-preferred and h-naive semantics is by reductions from SAT-UNSAT; Σ_2^{P}- hardness with respect to i-semi-stable and i-stage semantics are by reductions from credulous reasoning for AFs with the respective semantics; the remaining hardness results are shown via reductions from appropriate decision problems for inherited semantics.

We recall the standard reduction [89, Reduction 3.6] that provides the basis for DP-hardness of verification for h-preferred semantics and reappears in Section 6.3.

Reduction 6.2.11. *Let φ be given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in X and let $\bar{X} = \{\bar{x} \mid x \in X\}$. We construct AF $F = (A, R)$ with*

$$\begin{aligned} A &= X \cup \bar{X} \cup C \cup \{\varphi\} \\ R &= \{(x, cl) \mid cl \in C, x \in cl\} \cup \{(\bar{x}, cl) \mid cl \in C, \neg x \in cl\} \cup \\ &\quad \{(x, \bar{x}), (\bar{x}, x) \mid x \in X\} \cup \{(cl_i, \varphi) \mid i \leq n\} \end{aligned}$$

Intuitively, each conflict-free set of literal-arguments that defend the argument φ corresponds to a satisfying assignment of φ. An example is given in Figure 6.3.

We next present a reduction from SAT-UNSAT to $Ver_{pr_h}^{CAF}$ which shows DP-hardness. For a SAT-UNSAT instance (φ_1, φ_2) we apply Reduction 6.2.11 to both formulae; moreover, we let the clause-arguments be self-attacking. The resulting verification instance consists of the (disjoint) union of the two CAFs.

Reduction 6.2.12. *Let (φ_1, φ_2) be an instance of* SAT-UNSAT*, where each of the propositional formulae φ_i (for $i = 1, 2$) is given over a set of clauses $C_i = \{cl_1^i, \dots, cl_n^i\}$ over atoms in X_i. Moreover, we assume $X_1 \cap X_2 = \emptyset$. Let (A_i, R_i) be the AFs that we obtain when applying Reduction 6.2.11 to the formulae φ_i and adding attacks $\{(cl, cl) \mid cl \in C_i\}$. We construct the CAF $\mathcal{F}_{(\varphi_1, \varphi_2)} = (A_1 \cup A_2, R_1 \cup R_2, cl)$ with $cl(x) = cl(\bar{x}) = x$ for all $x \in X_i$, $cl(cl) = d$ for all $cl \in C_i$ and $cl(\varphi_i) = \varphi_i$.*

We refer to Figure 6.4 for an illustrative example. Next we show that a formula φ_i is satisfiable iff $X_i \cup \{\varphi_i\}$ is a h-preferred claim-set of (A_i, R_i, cl) which yields the correctness of the reduction.

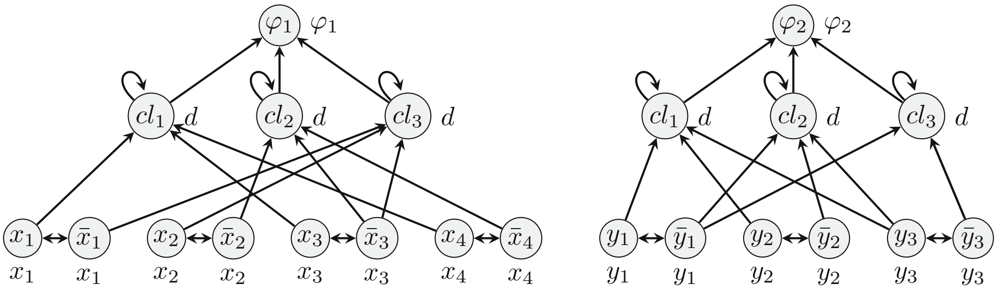


Figure 6.4: Reduction 6.2.12 for formulae (φ_1, φ_2) given by clauses $\{\{x_1, x_3, x_4\}, \{\neg x_3, \neg x_4, \neg x_2)\}, \{\neg x_1, \neg x_3, x_2\}\}$ and $\{\{y_1, y_2, y_3\}, \{\neg y_1, \neg y_2, y_3)\}, \{\neg y_1, y_3, \neg y_3\}\}$, resp.

Lemma 6.2.13. *(φ_1, φ_2) is a valid* SAT-UNSAT *instance iff $X_1 \cup X_2 \cup \{\varphi_1\}$ is a h-preferred claim-set of $\mathcal{F}_{(\varphi_1,\varphi_2)}$.*

Proof. We show that $X_1 \cup X_2 \cup \{\varphi_1\}$ is h-preferred in $\mathcal{F}_{(\varphi_1,\varphi_2)}$ iff φ_1 is satisfiable and φ_2 is unsatisfiable. We construct the CAF $\mathcal{F}_{(\varphi_1,\varphi_2)}$ as the disjoint union of the CAFs $\mathcal{F}_1 = (A_1, R_1, cl)$ and $\mathcal{F}_2 = (A_2, R_2, cl)$. Since $\mathcal{F}_1$ and $\mathcal{F}_2$ have no common arguments (and thus $pr_h(\mathcal{F}) = \{S \cup T \mid S \in pr_h(\mathcal{F}_1), T \in pr_h(\mathcal{F}_2)\}$), it suffices to show that

(a) φ_i is satisfiable iff $X_i \cup \{\varphi_i\}$ is a h-preferred claim-set of $\mathcal{F}_i$, and

(b) φ_i is unsatisfiable iff X_i is a h-preferred claim-set of $\mathcal{F}_i$.

We have that (b) follows from (a) since X_i is i-admissible in $\mathcal{F}_i$ independently of the satisfiability of φ_i (for an ad_i-realization, consider $X' \cup \{\bar{x} \mid x \notin X'\}$ for any $X' \subseteq X_i$) and no argument $cl \in C_i$ can appear in an admissible set. We show φ_i is satisfiable iff $X_i \cup \{\varphi_i\}$ is a h-preferred claim-set of $\mathcal{F}_i$:

Assume φ_i is satisfiable and consider a model M of φ_i. Let $E = M \cup \{\bar{x} \mid x \notin M\}$. We show that $E' = E \cup \{\varphi_i\}$ is admissible in (A_i, R_i'): First observe that E is admissible since each $a \in X_i \cup \bar{X}_i$ defends itself. Since M satisfies φ_i, we have that for any clause $cl \in C_i$, there is either $x \in cl$ with $x \in M$ or $\bar{x} \in cl$ with $x \notin M$, thus E attacks each $cl \in C$. Consequently, E defends φ_i; we conclude that E' is admissible in (A_i, R_i'). Moreover, $cl(E')$ is a subset-maximal i-admissible claim-set since $cl(E') = A_i \setminus \{d\}$, that is, $cl(E')$ contains every claim $c \in cl(A_i)$ which is assigned to non-self-attacking arguments. Thus $cl(E') = X_i \cup \{\varphi_i\}$ is h-preferred in $\mathcal{F}_i$.

Now assume $X_i \cup \{\varphi_i\}$ is h-preferred in $\mathcal{F}_i$. Then there is an ad_i-realization E of $X_i \cup \{\varphi_i\}$ which attacks each clause-argument $cl \in C_i$. Hence $M = E \cap X_i$ is a model of φ_i. $\square$

By the above lemma and since the reduction can be performed in polynomial time we obtain DP-hardness of the verification problem with respect to h-preferred semantics.

Proposition 6.2.14. *$Ver_{pr_h}^{CAF}$ is* DP*-hard.*

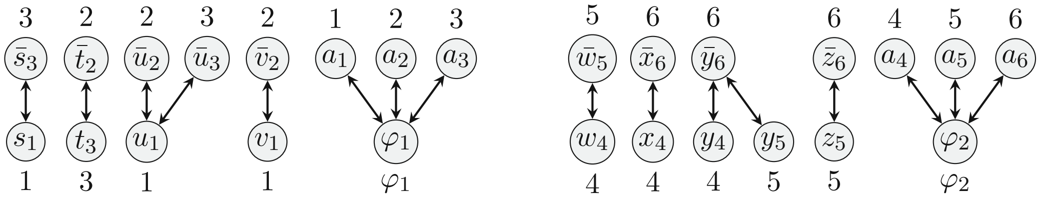


Figure 6.5: Reduction 6.2.15 for formulae (φ_1, φ_2) given by the sets of clauses $\{\{s, u, v\}, \{\neg u, \neg v, \neg t)\}, \{\neg s, \neg u, t\}\}$ and $\{\{w, x, y\}, \{\neg w, y, z\}, \{\neg x, \neg y, \neg z\}\}$, resp.

DP-hardness of verification for h-naive semantics can be shown via a reduction from SAT-UNSAT by combining ideas from the previous propositions. As in Proposition 6.2.14, we construct two independent frameworks $\mathcal{F}_1$, $\mathcal{F}_2$ representing the formulae (3-CNFs) φ_1, φ_2 given by clauses $C_1 = \{cl_1, \dots, cl_m\}$ resp. $C_2 = \{cl_{m+1}, \dots, cl_n\}$. The construction is similar to the one in Proposition 6.2.10: for each literal in a clause $cl_i \in C_j$ we introduce an argument with claim i; moreover, we add arguments φ_1, φ_2 to represent both formulae; finally, we add $|C_j|$ arguments with claims $1, \dots, m$, and $m+1, \dots, n$. We will show that $\{1, \dots, n, \varphi_1\}$ is h-naive in $\mathcal{F}_1 \cup \mathcal{F}_2$ iff φ_1 is satisfiable and φ_2 is unsatisfiable.

Reduction 6.2.15. *Let* (φ_1, φ_2) *be an instance of* SAT-UNSAT*, where each of the formulae* φ_j *(for* $j = 1, 2$*) is given over a set of clauses* C_j *over atoms in* X_j*. Moreover, we assume* $X_1 \cap X_2 = \emptyset$*,* $C_1 = \{cl_1, \dots, cl_m\}$*,* $C_2 = \{cl_{m+1}, \dots, cl_n\}$*, and define* $A'_1 = \{a_1, \dots, a_m\}$ *and* $A'_2 = \{a_{m+1}, \dots, a_n\}$*. We construct CAF* $\mathcal{F}_{(\varphi_1,\varphi_2)} = (A, R, cl)$ *with*

$$\begin{aligned}
A &= \{x_i \mid x \in cl_i, 1 \leq i \leq n\} \cup \{\bar{x}_i \mid \bar{x} \in cl_i, 1 \leq i \leq n\} \cup A'_1 \cup A'_2 \cup \{\varphi_1, \varphi_2\}\\
R &= \{(x_i, \bar{x}_j)(\bar{x}_j, x_i), \mid i, j \leq n\} \cup \{(a_i, \varphi_1), (\varphi_1, a_i) \mid i \leq m\} \cup\\
&\quad \{(a_i, \varphi_2), (\varphi_2, a_i) \mid m < i \leq n\}
\end{aligned}$$

with $cl(x_i) = cl(\bar{x}_i) = cl(a_i) = i$ *and* $cl(\varphi_i) = \varphi_i$*.*

See Figure 6.5 for an example illustrating the reduction.

Lemma 6.2.16. (φ_1, φ_2) *is a valid* SAT-UNSAT *instance iff* $\{1, \dots, n, \varphi_1\} \in na_h(\mathcal{F})$*.*

Proof. For the purpose of this proof we consider the CAF $\mathcal{F}_{(\varphi_1,\varphi_2)}$ as disjoint union of two CAFs $\mathcal{F}_1$ and $\mathcal{F}_2$. To this end let $\mathcal{F}_1$ be the projection of $\mathcal{F}_{(\varphi_1,\varphi_2)}$ on the arguments $\{x_i \mid x \in cl_i, 1 \leq i \leq m\} \cup \{\bar{x}_i \mid \bar{x} \in cl_i, 1 \leq i \leq m\} \cup A'_1 \cup \{\varphi_1\}$ and $\mathcal{F}_2$ be the projection of $\mathcal{F}_{(\varphi_1,\varphi_2)}$ on the arguments $\{x_i \mid x \in cl_i, m+1 \leq i \leq n\} \cup \{\bar{x}_i \mid \bar{x} \in cl_i, m+1 \leq i \leq n\} \cup A'_2 \cup \{\varphi_2\}$. Notice that $\mathcal{F}_{(\varphi_1,\varphi_2)} = \mathcal{F}_1 \cup \mathcal{F}_2$ and that $\mathcal{F}_1$ and $\mathcal{F}_2$ are isomorphic.

We show φ_1 is satisfiable and φ_2 is unsatisfiable iff $\{1, \dots, n, \varphi_1\} \in na_h(\mathcal{F})$ by proving

(a) φ_1 is satisfiable iff $\{1, \dots, m, \varphi_1\} \in na_h(\mathcal{F}_1)$.

(b) φ_2 is unsatisfiable iff $\{m+1, \dots, n\} \in na_h(\mathcal{F}_2)$.

We have $na_i(\mathcal{F}) = \{S \cup T \mid S \in na_i(\mathcal{F}_1), T \in na_i(\mathcal{F}_2)\}$ since $\mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset$ and $cl(A_1) \cap cl(A_2) = \emptyset$. Thus φ_1 is satisfiable and φ_2 is unsatisfiable iff $\{1, \dots, n, \varphi_1\} \in na_h(\mathcal{F})$.

Proof of (a): First assume φ_1 is satisfiable and consider a model M of φ_1. Let $E = \{x_i \mid x \in M, i \leq m\} \cup \{\bar{x}_i \mid x \notin M, i \leq m\} \cup \{\varphi_1\}$. E is conflict-free by construction; moreover, $\varphi_1 \in cl(E)$ and $i \in cl(E)$ for each $i \leq m$: For each clause $cl_i \in C_1$, there is either $x \in M \cap cl_i$ or $\bar{x} \in cl_i$ such that $x \notin M$, consequently there is either $x_i \in E$ with $cl(x_i) = i$ or $\bar{x}_i \in E$ with $cl(\bar{x}_i) = i$. Hence $\{1, \dots, m, \varphi_1\}$ has a cf_i-realization in $\mathcal{F}_1$.

Now assume $\{1, \dots, m, \varphi_1\} \in na_h(\mathcal{F})$. Let E be a cf_i-realization of $\{1, \dots, m, \varphi_1\}$ and let $M = \{x \mid \exists i \leq m : x_i \in E\}$. Now, consider an arbitrary clause $cl_i \in C_1$. Then E contains an argument with claim i, that is, either $x_i \in E$ or $\bar{x}_i \in E$. In the former case, $x \in M$ and thus cl_i is satisfied. In the latter case, $x \notin M$ as $\bar{x}_i$ is in conflict with all arguments x_j and thus cl_i is satisfied. Hence M is a model of φ_1, i.e., φ_1 is satisfiable.

Proof of (b): First notice that $cl(A_2') = \{m+1, \dots, n\}$ is i-conflict-free by construction. By (a), φ_2 is unsatisfiable iff $\{m+1, \dots, n, \varphi_2\} \notin na_h(\mathcal{F}_2')$. We thus obtain φ_2 is unsatisfiable iff $\{m+1, \dots, n, \varphi_2\} \notin na_h(\mathcal{F}_2)$ iff $\{m+1, \dots, n\} \in na_h(\mathcal{F}_2)$. ☐

We obtain DP-hardness of the verification problem with respect to h-naive semantics.

Proposition 6.2.17. *$Ver^{CAF}_{na_h}$ is DP-hard.*

Finally, we provide hardness results for h-semi-stable, τ-h-stable and h-stage semantics. We will present reductions from the verification problem of suitable inherited semantics.

Reduction 6.2.18. *For a CAF $\mathcal{F}$, let $A' = A \cup \{a' \mid a \in A\}$ and $cl' : A' \to cl(A) \cup \{c_a \mid a \in A\}$ with $cl'(a) = cl(a)$ and $cl'(a') = cl(c_a)$ for fresh claims $c_a \notin cl(A)$ for all $a \in A$. We let $Tr_1(\mathcal{F})$, $Tr_2(\mathcal{F})$, and $Tr_3(\mathcal{F})$ be defined as follows:*

- *$Tr_1(\mathcal{F}) = (A', R', cl')$ with $R' = R \cup \{(a, a'), (a', a') \mid a \in A\}$;*
- *$Tr_2(\mathcal{F}) = (A', R_2', cl')$ with $R_2' = R' \cup \{(a, b') \mid (a, b) \in R\}$;*
- *$Tr_3(\mathcal{F}) = (A', R_3', cl')$ with $R_3' = R_2' \cup \{(b, a) \mid (a, b) \in R\} \cup \{(a, b) \mid a \in A, (b, b) \in R\}$.*

Translations Tr_1 and Tr_3 already appear in [117]. It has been shown that Tr_1 maps i-preferred to h-semi-stable semantics and Tr_3 maps i-stage to h-stage semantics. We show that Tr_2 maps i-stable to h-stable semantics. We summarize the results below.

Lemma 6.2.19. *For a CAF $\mathcal{F}$, it holds that*

(1) $pr_i(\mathcal{F}) = pr_i(Tr_1(\mathcal{F})) = ss_h(Tr_1(\mathcal{F}))$;

(2) $stb_i(\mathcal{F}) = stb_i(Tr_2(\mathcal{F})) = \tau\text{-}stb_h(Tr_2(\mathcal{F}))$ for $\tau \in \{ad, cf\}$; moreover,

(3) $stg_i(\mathcal{F}) = stg_i(Tr_3(\mathcal{F})) = stg_h(Tr_3(\mathcal{F}))$.

ρ	$Cred_{\rho}^{CAF}$	$Skept_{\rho}^{CAF}$	Ver_{ρ}^{CAF}	NE_{ρ}^{CAF}
$ad\text{-}stb_h$	NP-c	coNP-c	**NP**-c	NP-c
$cf\text{-}stb_h$	NP-c	coNP-c	**NP**-c	NP-c
na_h	in P	$\underline{\boldsymbol{\Pi}_{\mathbf{2}}^{\mathbf{P}}\text{-c}}$	$\underline{\mathbf{DP}\text{-c}}$	in P
pr_h	NP-c	Π_2^{P}-c	$\underline{\mathbf{DP}\text{-c}}$	NP-c
ss_h	Σ_2^{P}-c	Π_2^{P}-c	$\boldsymbol{\Sigma}_{\mathbf{2}}^{\mathbf{P}}$-c	NP-c
stg_h	Σ_2^{P}-c	Π_2^{P}-c	$\boldsymbol{\Sigma}_{\mathbf{2}}^{\mathbf{P}}$-c	in P

Table 6.3: Complexity of CAF semantics. Results that deviate from AF semantics are bold-face; results that deviate from those w.r.t. inherited semantics are underlined.

Proof. Proofs for (1) and (3) can be found in [117]. To verify (2), let $Tr_2(\mathcal{F}) = \mathcal{F}' = (A', R', cl')$. Since $stb_i(\mathcal{F}) \subseteq ad\text{-}stb_h(\mathcal{F}) \subseteq cf\text{-}stb_h(\mathcal{F})$ holds for any CAF $\mathcal{F}$, it suffices to show that (i) $stb_i(\mathcal{F}) \subseteq stb_i(\mathcal{F}')$ and (ii) $cf\text{-}stb_h(\mathcal{F}') \subseteq stb_i(\mathcal{F})$.

First observe that (a) for every set $E \subseteq A$, E attacks argument a' in $\mathcal{F}'$ iff $a \in E \cup E_F^+$. Indeed, E attacks an argument a' iff either $a \in E$ or if there is $b \in E$ such that $(b, a) \in R$.

(i) Let $S \in stb_i(\mathcal{F})$ and consider a stb_i-realization $E \subseteq A$. We show that E is stable in $\mathcal{F}'$: First notice that E is conflict-free since we introduced no attacks between existing arguments in $\mathcal{F}'$. Moreover, E attacks every argument $a \in A' \setminus E$: E attacks every argument $a \in A \setminus E$; moreover, E attacks every $a' \in \{a' \mid a \in A\}$ by (a) since $E \cup E_F^+ = A$.

(ii) Let $S \in cf\text{-}stb_h(\mathcal{F}')$, then there is a set $E \in A'$ such that $E \in cf(F)$ and $cl(E) \cup E_{\mathcal{F}'}^{\circledast} = cl(A')$. We show that $E \in stb(F)$. First observe that $E \subseteq A$ since each argument $a' \in \{a' \mid a \in A\}$ is self-attacking; moreover, E is conflict-free in F. We show that E attacks every argument $a \in A \setminus E$: We have $\{c_a \mid a \in A\} \subseteq E_{\mathcal{F}'}^{\circledast}$ since $cl(E) \cup E_{\mathcal{F}'}^{\circledast} = cl(A')$. Thus E attacks each argument a' in $\mathcal{F}'$. We conclude by (a) that $a \in E \cup E_F^+$ for every argument $a \in A$. We have shown that $E \in stb(F)$ and, consequently, $S \in stb_i(\mathcal{F})$. ☐

Lower bounds for Ver_{ρ}^{CAF}, $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h, ss_h, stg_h\}$, thus follow from the results of the respective inherited semantics: For a given CAF $\mathcal{F} = (A, R, cl)$ and a set of claims $S \subseteq cl(A)$, one can check $S \in \rho'(\mathcal{F})$, $\rho' \in \{stb, pr, stg\}$, by applying the respective translation and checking whether S is a ρ-realization in the resulting CAF.

Proposition 6.2.20. *Ver_{ρ}^{CAF} is* NP*-hard for $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h\}$ and Σ_2^{P}-hard for $\rho \in \{ss_h, stg_h\}$.*

Proof. The NP-hardness of Ver_{ρ}^{CAF} for $\rho \in \{ad\text{-}stb_h, cf\text{-}stb_h\}$ is by the fact that Ver_{stb_i} is NP-hard and translation Tr_2. The Σ_2^{P}-hardness of $Ver_{ss_h}^{CAF}$ is by the fact that Ver_{pr_i} is Σ_2^{P}-hard and translation Tr_1. Finally, the Σ_2^{P}-hardness of $Ver_{stg_h}^{CAF}$ is by the fact that Ver_{stg_i} is Σ_2^{P}-hard and translation Tr_3. ☐

This concludes our complexity analysis of general CAFs. The full complexity landscape for hybrid semantics is summarized in Table 6.3. While both stable variants as well as semi-stable and stage semantics admit the same complexity as their inherited counter-part, we observe different behavior of h-naive and h-preferred semantics: verifying h-preferred and h-naive claim-extensions is DP-hard which is of lower complexity than i-preferred semantics but of higher complexity than i-naive semantics. Moreover, the skeptical acceptance problem for h-naive semantics is surprisingly hard, one level higher than its inherited counter-part and even two level higher than naive AF semantics.

6.2.2 Complexity Results for well-formed CAFs

We now turn to the complexity of well-formed CAFs. First observe that all upper bounds from the previous section carry over since well-formed CAFs are a special case of CAFs. It remains to give improved upper bounds for verification with respect to all of the considered semantics as well as for $Skept^{wf}_{na_h}$. The latter also requires a genuine hardness proof as it remains harder than the corresponding problem for AFs even in the well-formed case. For the remaining semantics, we obtain hardness results from the corresponding problems for AFs since they constitute a special case of the respective problems for CAFs.

We first discuss improved upper bounds for verification. For preferred as well as for both variants of h-stable semantics, membership is immediate by the corresponding results for inherited semantics as the respective semantics collapse in the well-formed case.

Proposition 6.2.21. *Ver^{wf}_{ρ} is in* P *for $\rho \in \{cf\text{-}stb_h, ad\text{-}stb_h\}$ and in* coNP *for $\rho = pr_h$.*

For the remaining semantics, we exploit the following observation [92].

Lemma 6.2.22. *Let $\mathcal{F}$ be well-formed. For $S \subseteq cl(A)$, let*

$$\begin{aligned} E_0(S) =& \{a \in A \mid cl(a) \in S\} \\ E_1(S) =& E_0(S) \setminus E_0(S)^+_F \\ E_2(S) =& \{a \in E_1(S) \mid b \in E_1(S)^+_F \text{ for all } (b,a) \in R\}. \end{aligned}$$

Then $S \in cf_i(\mathcal{F})$ iff $S = cl(E_1(S))$ and $S \in ad_i(\mathcal{F})$ iff $S = cl(E_2(S))$.

To check whether a set $S \subseteq cl(A)$ is h-naive in a given well-formed CAF $\mathcal{F}$, we utilize Lemma 6.2.22 to test (i) $S \in cf_i(\mathcal{F})$ and (ii) $S \cup \{c\} \notin cf_i(\mathcal{F})$ for all $c \in cl(A) \setminus S$, which yields a poly-time procedure for Ver^{wf}_{na}. For h-semi-stable and h-stage semantics, we first compute $E_1(S)$, resp. $E_2(S)$ in P (cf. Lemma 6.2.22) and utilize Lemma 6.2.2 to check in coNP whether $E_2(S)$ ($E_1(S)$) realizes a h-semi-stable (h-stage) claim set.

Proposition 6.2.23. *Ver^{wf}_{ρ} is in* coNP *for $\rho \in \{na_h, ss_h, stg_h\}$.*

It remains to discuss coNP-completeness of skeptical reasoning in well-formed CAFs w.r.t. h-naive semantics. To show hardness, we make use of a small adaption of the standard reduction (cf. Reduction 6.2.11) by removing the argument φ and all associated attacks.

ρ	$Cred_{\rho}^{wf}$	$Skept_{\rho}^{wf}$	Ver_{ρ}^{wf}	NE_{ρ}^{wf}
$cf\text{-}stb_h$	NP-c	coNP-c	in P	NP-c
$ad\text{-}stb_h$	NP-c	coNP-c	in P	NP-c
na_h	in P	**coNP**-c	in P	in P
pr_h	NP-c	Π_2^{P}-c	coNP-c	NP-c
ss_h	Σ_2^{P}-c	Π_2^{P}-c	coNP-c	NP-c
stg_h	Σ_2^{P}-c	Π_2^{P}-c	coNP-c	in P

Table 6.4: Complexity of semantics in well-formed CAFs. Results that deviate from AFs (cf. Table 6.1) are highlighted in bold-face.

Proposition 6.2.24. *$Skept_{na_h}^{wf}$ is* coNP*-complete.*

Proof. For a well-formed CAF $\mathcal{F} = (A, R, cl)$, one can verify skeptical acceptance of a claim $c \in cl(A)$ by (1) guessing a set $E \subseteq A$ such that $c \notin cl(E)$; (2) checking if $cl(E)$ is a h-naive claim-set of $\mathcal{F}$. The latter can be verified in polynomial time, yielding a NP-procedure for the complementary problem.

We reduce from UNSAT: For a formula φ with clauses $C = \{cl_1, \dots, cl_n\}$ over the atoms X, let (A', R') be as in Reduction 6.2.11. We define $\mathcal{F} = (A, R, cl)$ with $A = A' \setminus \{\varphi\}$ and $R = R' \setminus \{(cl_i, \varphi) \mid i \leq n\}$, and set $cl(x) = x$, $cl(\bar{x}) = \bar{x}$, and $cl(cl_i) = \bar{\varphi}$. $\mathcal{F}$ is well-formed. We show φ is satisfiable iff $\bar{\varphi}$ is not skeptically accepted in $\mathcal{F}$.

In case φ is satisfiable, then there is a model $M \subseteq X$ of φ. Consider $E = M \cup \{\bar{x} \mid x \notin M\}$, which is conflict-free and cannot be extended by any argument cl_i assigned with claim $\bar{\varphi}$: Indeed, since each clause cl_i is satisfied by M, there is either a positive literal $x \in cl_i$ with $x \in M$ or a negative literal $\bar{x} \in cl_i$ with $x \notin M$, thus cl_i is attacked by E in F. Moreover, we have that for each $x \in X$, either $x \in E$ (and thus $x \in cl(E)$) or $\bar{x} \in E$ (and thus $\bar{x} \in cl(E)$) and $(x, \bar{x}) \in R$. Consequently, $cl(E)$ is maximal among i-conflict-free claim-sets and thus $cl(E) \in na_h(\mathcal{F})$. It follows that $\bar{\varphi}$ is not skeptically accepted in $\mathcal{F}$.

If $\bar{\varphi}$ is not skeptically accepted in $\mathcal{F}$, there is a set $S \in na_h(\mathcal{F})$ with $\bar{\varphi} \notin S$. Then the set $E \cap X$ for a na_h-realization E of S is a model of φ since it attacks each $cl \in C$. ☐

This concludes our complexity analysis of well-formed CAFs. We summarize our results in Table 6.4. While verification admits the same complexity as the corresponding AF semantics, we observe a rise in complexity for credulous acceptance for h-naive semantics. When comparing our findings with the corresponding inherited CAF semantics we see that, in contrast to general CAFs, the complexity of the variants coincide.

6.2.3 Coherence of Hybrid Semantics

The *coherence problem* asks whether preferred and stable extensions coincide. The problem is Π_2^{P}-complete for AFs [81]. It was studied for inherited semantics [92] showing that complexity remains on the second level. Formally, we study the following problem:

- *Coherence* (Coh_τ^{CAF}): Given a CAF $\mathcal{F}$, does it hold that $pr_h(\mathcal{F}) = \tau\text{-}stb_h(\mathcal{F})$?

The coherence problem restricted to well-formed CAFs is denoted Coh_τ^{wf}. The forthcoming result shows that, although the complexity of the verification task increases for h-preferred semantics, testing coherence for CAFs in terms of h-semantics is of the same complexity as in the AF setting, as well.

Proposition 6.2.25. *Coh_τ^Δ is Π_2^{P}-complete for $\tau \in \{ad, cf\}$, $\Delta \in \{CAF, wf\}$.*

Proof. We present a Σ_2^{P}-procedure for the complementary problem: (1) Guess a set $S \subseteq cl(A)$; and (2) check $S \in (\tau\text{-}stb_h(\mathcal{F}) \setminus pr_h(\mathcal{F})) \cup (pr_h(\mathcal{F}) \setminus \tau\text{-}stb_h(\mathcal{F}))$. Verifying that S is h-preferred is DP-complete, verifying that S is h-stable is NP-complete. Hardness follows from the corresponding result for AFs, hence we obtain Π_2^{P}-completeness. ☐

6.3 Complexity of Concurrence

We study the collapse of inherited and hybrid semantics. As observed in Section 4.1, preferred and stable variants coincide on well-formed CAFs; which is not the case for the remaining semantics. The goal of this section is to study the complexity of deciding whether the semantics collapse. Formally, we are interested in the following problem:

- *Concurrence* (Con_σ^{CAF}): Given a CAF $\mathcal{F}$, does it hold that $\sigma_i(\mathcal{F}) = \sigma_h(\mathcal{F})$?

For stable semantics, we write $Con_{\tau\text{-}stb}^{CAF}$ to specify the considered h-stable variant ($\tau \in \{ad, cf\}$). The concurrence problem restricted to well-formed CAFs is denoted Con_σ^{wf}.

Our results are summarized in Table 6.5 and show that deciding concurrence is in general computationally hard; observe that for semi-stable and stage semantics, the problem is complete for the third level of the polynomial hierarchy. For preferred and stable semantics on the other hand, the question becomes trivial for well-formed CAFs as the claim-based versions of this semantics coincide with their inherited counter-parts. We furthermore show that deciding whether $cf\text{-}stb_h(\mathcal{F}) = ad\text{-}stb_h(\mathcal{F})$ for a given CAF $\mathcal{F}$ is Π_2^{P}-complete and conclude the section with a brief discussion of the well known coherence problem when applied to claim-based semantics. Let us start with the collection of results concerning concurrence which will be proven in the forthcoming two subsections.

Theorem 6.3.1. *The complexity results depicted in Table 6.5 hold.*

	pr	na	$\tau\text{-}stb$	ss	stg
Con_σ^{CAF}	Π_2^{P}-c	coNP-c	Π_2^{P}-c	Π_3^{P}-c	Π_3^{P}-c
Con_σ^{wf}	trivial	coNP-c	trivial	Π_2^{P}-c	Π_2^{P}-c

Table 6.5: Complexity of deciding Con_σ^{CAF} and Con_σ^{wf}.

6.3.1 Concurrence of General CAFs

We start with a rather straight-forward observation for preferred and naive semantics which will be useful for both membership and hardness arguments.

Proposition 6.3.2. *For a CAF $\mathcal{F}$, for $\sigma \in \{pr, na\}$, $\sigma_i(\mathcal{F}) = \sigma_h(\mathcal{F})$ if and only if $\sigma_i(\mathcal{F})$ is incomparable.*

Proof. Let $\sigma = pr$ (the proof for $\sigma = na$ is analogous). Assume $pr_i(\mathcal{F})$ is incomparable and let $S \in pr_i(\mathcal{F})$. Then $S \in ad_i(\mathcal{F})$. Now assume there is $T \in ad_i(\mathcal{F})$ with $T \supset S$. Consider a ad_i-realization E of T in $\mathcal{F}$ and let $E' \in pr(F)$ with $E \subseteq E'$. But then $cl(E') \in pr_i(\mathcal{F})$ and $cl(E') \supseteq T \supset S$, contradiction to $pr_i(\mathcal{F})$ being incomparable. □

To get upper bounds for preferred and naive semantics, it thus suffices to verify incomparability of $\sigma_i(\mathcal{F})$. We give a Σ_2^{P} (NP resp.) procedure for the complementary problem: Guess $E, G \subseteq A$ and check (i) $E, G \in \sigma(F)$ and (ii) $cl(E) \subset cl(G)$. The former is in coNP for pr (in P for na) by Table 6.1.

Membership for the remaining semantics is by the following generic guess and check procedure for the complementary problem: To verify $\sigma_i(\mathcal{F}) = \sigma_h(\mathcal{F})$ for a given CAF $\mathcal{F}$ one guesses a set of claims $S \subseteq cl(A)$ and checks whether $S \in \sigma_i(\mathcal{F})$ and $S \notin \sigma_h(\mathcal{F})$ or vice versa. The complexity of the procedure thus follows from the corresponding results for verification with respect to the considered semantics, i.e. NP-membership for stable semantics; Σ_2^{P}-membership for semi-stable and stage semantics, cf. Tables 6.2 and 6.3.

Before turning to the results for the matching lower bounds in general CAFs, let us point out that for all except naive semantics, deciding concurrence admits a lower complexity for well-formed CAFs than for general CAFs. When presenting our work on the concurrence problem at the AAAI 2021 conference, the problem of deciding concurrence for naive semantics for well-formed CAFs has been left open. This gap has been recently closed in [126]. In the present work, we omit the original hardness proof for general CAFs presented in [99] and recall the construction from [126] instead as it covers both cases.

Reduction 6.3.3. *Let φ be given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in X. Let (A, R) be defined as in Reduction 6.2.11. We construct a CAF $\mathcal{F}$ with $A_{\mathcal{F}} = A \cup \{a_1, a_2\}$, $R_{\mathcal{F}} = R \cup \{(\varphi, a_2)\}$, and $cl(x) = x$, $cl(\bar{x}) = \bar{x}$, $cl(cl_i) = cl_i$, $cl(\varphi) = \varphi$ and $cl(a_i) = a$.*

Proposition 6.3.4 ([126])**.** *Con_{na}^{CAF} is* coNP*-hard.*

Proposition 6.3.5. *Con_{pr}^{CAF} is Π_2^{P}-hard.*

Proof. We present a reduction from $Skept_{pr}^{AF}$: Given an instance $F = (A, R)$, $a \in A$ from $Skept_{pr}^{AF}$. W.l.o.g. we can assume that the preferred extensions of F are non-empty (otherwise add an isolated argument). We construct $\mathcal{F} = (A', R', cl)$ with $A' = A \cup \{i, j\}$, $R' = R \cup \{(j, b), (b, j) \mid b \in A\}$, and $cl(a) = cl(j) = c_1$, $cl(b) = c_2$ for $b \in (A \setminus \{a\}) \cup \{i\}$.

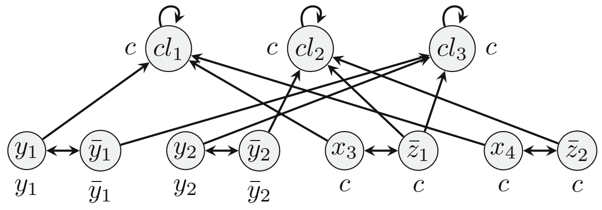


Figure 6.6: A CAF illustrating Reduction 6.3.6 for the formula $\Psi = \forall Y \exists Z \varphi(Y,Z)$ where $\varphi(Y,Z)$ is given by the clauses $\{\{y_1, z_1, z_2\}, \{\bar{z}_1, \bar{z}_2, \bar{y}_2)\}, \{\bar{y}_1, \bar{z}_1, y_2\}\}$.

Then $pr((A',R')) = \{E \cup \{i\} \mid E \in pr((A,R))\} \cup \{\{i,j\}\}$ since the argument i is isolated and thus appears in each extension; moreover, j mutually attacks each argument $b \in A$. For all extensions $D \in pr((A',R'))$ with $a \in D$ we have $cl(D) = \{c_1, c_2\}$; for all extensions $D \in pr(F')$, $D \neq \{i,j\}$, with $a \notin D$, we have $cl(D) = \{c_2\}$; moreover, $cl(\{i,j\}) = \{c_1, c_2\}$ and thus we have $\{c_1, c_2\} \in pr_i(\mathcal{F})$ independently of the considered instance. Thus a is not skeptically accepted in F with respect to preferred semantics iff $\{c_2\} \in pr_i(\mathcal{F})$ iff $pr_i(\mathcal{F})$ is not incomparable. Applying Proposition 6.3.2 concludes the proof. □

Next we present our Π_2^{P}-hardness proof for hybrid stable semantics. We will make use of the following reduction which modifies the standard reduction by (a) removing the argument φ and (b) adding self-attacks to all clause-arguments.

Reduction 6.3.6. *Let $\Psi = \forall Y \exists Z \varphi(Y,Z)$ be an instance of $\mathsf{QSAT}_2^\forall$, where φ is given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $X = Y \cup Z$ and let (A,R) be as in Reduction 6.2.11. We define a CAF (A', R', cl) with*

$$A' = A \setminus \{\varphi\}$$
$$R' = (R \cup \{(cl_i, cl_i) \mid i \leq n\}) \setminus \{(cl_i, \varphi) \mid i \leq n\}$$

and $cl(y) = y$, $cl(\bar{y}) = \bar{y}$, $cl(v) = cl(cl_i) = c$ for $i \leq n$ and $v \in Z \cup \bar{Z}$.

See Figure 6.6 for an illustrative example of the reduction.

Proposition 6.3.7. *$Con_{\tau\text{-}stb}^{CAF}$, $\tau \in \{cf, ad\}$ is Π_2^{P}-hard.*

Proof. We present a reduction from $\mathsf{QSAT}_2^\forall$. Let $\Psi = \forall Y \exists Z \varphi(Y,Z)$ be an instance of $\mathsf{QSAT}_2^\forall$, where φ is given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $X = Y \cup Z$. Let (A,R) be as in Reduction 6.3.6.

We will first show that (a) $\tau\text{-}stb_h(\mathcal{F}) = \{Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \mid Y' \subseteq Y\}$: Each τ-h-stable claim-set S contains either y or $\bar{y}$ by construction; moreover, $c \in S$ since c is not defeated by any conflict-free set of arguments $E \subseteq A$, thus each τ-h-stable claim-set is of the form $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\}$ for some $Y' \subseteq Y$. Moreover, each such set is stb_h-realizable, since for any $Y' \subseteq Y$, $z \in Z$, the set $E = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{z\}$ is admissible (conflict-free) in (A, R') and attacks every $a \in A$ such that $cl(a) \notin cl(E)$.

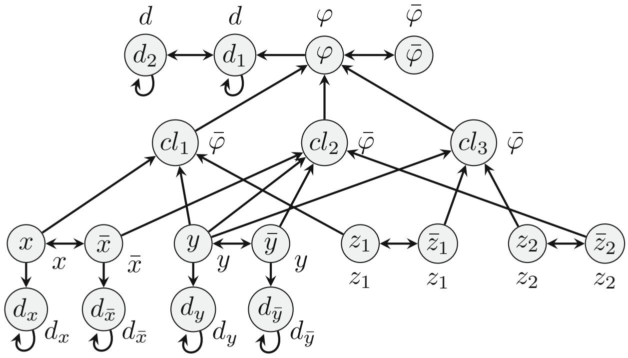


Figure 6.7: Reduction 6.3.8 for the formula $\exists X \forall Y \exists Z \varphi(X, Y, Z)$ with clauses $\{\{z_1, x, y\}, \{\neg x, \neg y, \neg z_2, y\}, \{\neg z_1, z_2, y\}\}$.

We show Ψ is valid iff $stb_i(\mathcal{F}) = \tau\text{-}stb_h(\mathcal{F})$.

Assume Ψ is valid and let $Y' \subseteq Y$. Then there is $Z' \subseteq Z$ such that φ is satisfied by $M = Y' \cup Z'$. We show that $E = M \cup \{\bar{x} \mid x \notin M\}$ is a stable extension of F: indeed, since M is a model of φ, each clause-argument is attacked; moreover, E is conflict-free. Therefore $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in stb_i(\mathcal{F})$. As Y' was arbitrary, we have that $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in stb_i(\mathcal{F})$ for all $Y' \subseteq Y$. Hence $stb_i(\mathcal{F}) = \tau\text{-}stb_h(\mathcal{F})$ by (a).

Now, assume $stb_i(\mathcal{F}) = \tau\text{-}stb_h(\mathcal{F})$. Let $Y' \subseteq Y$ and let $S = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\}$. By (a) it holds that $S \in \tau\text{-}stb_h(\mathcal{F})$ $(= stb_i(\mathcal{F}))$. Consider a stb_i-realization E of S and let $Z' = E \cap Z$. Then $M = Y' \cup Z'$ satisfies φ since each clause cl_i is attacked which encodes membership of the respective literal. Thus for every $Y' \subseteq Y$, there is $Z' \subseteq Z$ such that $Y' \cup Z'$ satisfies φ. It follows that Ψ is valid. □

We finally arrive at the Π_3^{P}-hardness proofs for concurrence in the case of semi-stable and stage semantics. We begin with defining the reduction (which will be used for both problems) and a technical lemma stating some first observations.

Reduction 6.3.8. *Let $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$ be an instance of $\mathsf{QSAT}_3^{\exists}$, where φ is given by a set of clauses $\mathcal{C} = \{cl_1, \dots, cl_n\}$ over atoms in $V = X \cup Y \cup Z$. We can assume that there is a variable $y_0 \in Y$ with $y_0 \in cl_i$ for all $i \leq n$ (otherwise we can add such a y_0 without changing the validity of Ψ). Let F be the AF constructed from φ as in Reduction 6.2.11. We define $\mathcal{F} = (A', R', cl)$ with*

$$\begin{aligned} A' &= A \cup \{d_1, d_2, \bar{\varphi}\} \cup \{d_v, d_{\bar{v}} \mid v \in X \cup Y\} \\ R' &= R \cup \{(a, d_a), (d_a, d_a), \mid a \in X \cup \bar{X} \cup Y \cup \bar{Y}\} \cup \\ &\quad \{(\varphi, \bar{\varphi}), (\bar{\varphi}, \varphi), (\varphi, d_1)\} \cup \{(d_i, d_j) \mid i, j \leq 2\} \end{aligned}$$

and $cl(v) = cl(\bar{v}) = v$ for $v \in Y \cup Z$; $cl(cl_i) = \bar{\varphi}$; $cl(d_i) = d$; $cl(a) = a$ otherwise.

An illustrative example of the reduction is given in Figure 6.7.

Lemma 6.3.9. *Let* $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$ *be an instance of* $\mathsf{QSAT}_3^{\exists}$ *and let* $\mathcal{F}$ *be defined as in Reduction 6.3.8. Then for all* $E \in ss(F)$*,*

1. $\varphi \in E \Leftrightarrow \bar{\varphi} \notin E$;
2. $\varphi \in E \Leftrightarrow E^{\oplus}_{F} = A \setminus (\{d_a \mid a \in (X \cup \bar{X} \cup Y \cup \bar{Y}) \setminus E\} \cup \{d_2\})$;
3. $\bar{\varphi} \in E \Leftrightarrow \mathcal{C} \cap E \neq \emptyset$;
4. $\bar{\varphi} \in E \Leftrightarrow E^{\oplus}_{F} = A \setminus (\{d_a \mid a \in (X \cup \bar{X} \cup Y \cup \bar{Y}) \setminus E\} \cup \{d_1, d_2\})$.

Proof. Let $F = (A, R)$ and first observe that (1) is immediate by construction.

For (2), first assume $\varphi \in E$. Then $\bar{\varphi}, d_1 \in E^{\oplus}_{F}$ since $\varphi \in E$; also, $\varphi \in E$ only if E defends φ against each cl_i, $i \leq n$, thus each cl_i is attacked by E; moreover, each $a \in V \cup \bar{V}$ is either contained or attacked by E, otherwise, $D = E \cup \{a\}$ is admissible in F with $D^{\oplus}_{F} \supset E^{\oplus}_{F}$, contradiction to $E \in ss(F)$. Thus $V \cup \bar{V} \in E^{\oplus}_{F}$ and $d_a \in E^{\oplus}_{F}$ for $a \in E \cap (X \cup \bar{X} \cup Y \cup \bar{Y})$. In case $E^{\oplus}_{F} = A \setminus (\{d_a \mid a \in (X \cup \bar{X} \cup Y \cup \bar{Y}) \setminus E\} \cup \{d_2\})$, we have $\varphi \in E$ since φ is the only argument attacking d_1.

To show (3), first assume $\bar{\varphi} \in E$. Towards a contradiction assume $\mathcal{C} \cap E = \emptyset$. Then $D = (E \cup \{\varphi\}) \setminus \{\bar{\varphi}\}$ is admissible in F and $D^{\oplus}_{F}$ is a proper subset of $E^{\oplus}_{F}$, contradiction to E being semi-stable in F. It follows that $\mathcal{C} \cap E \neq \emptyset$. The other direction is immediate since $\mathcal{C} \cap E \neq \emptyset$ implies $\varphi \notin E$. By (1) we obtain $\bar{\varphi} \in E$.

To show (4) let us again assume $\bar{\varphi} \in E$. Then $\varphi \in E^{+}_{F}$; moreover, each $a \in V \cup \bar{V}$ is either contained in E or attacked by E, otherwise, $D = (E \cup \{a\}) \setminus \{cl_i \mid i \leq n, (a, cl_i) \in R\}$ is admissible in F and satisfies $D^{\oplus}_{F} \supset E^{\oplus}_{F}$, contradiction to $E \in ss(F)$. We thus have $V \cup \bar{V} \in E^{\oplus}_{F}$ and $d_a \in E^{\oplus}_{F}$ for $a \in E \cap (X \cup \bar{X} \cup Y \cup \bar{Y})$. Also, each cl_i is either attacked by E or defended by E (by (3), there is at least one $i \leq n$ such that $cl_i \in E$). The other direction follows since $d_1 \notin E^{\oplus}_{F}$ and thus $\varphi \notin E$. □

Next we provide some properties for the reduction making use of the observation that for any instance of $\mathsf{QSAT}_3^{\exists}$, each i-semi-stable and each h-semi-stable as well as each i-stage and h-stage claim-set in the resulting CAF is of the form $S_{X'} \cup \{e\}$ where

$$S_{X'} = X' \cup \{\bar{x} \mid x \notin X'\} \cup Y \cup Z$$

for some $X' \subseteq X$ and for $e \in \{\varphi, \bar{\varphi}\}$; in fact, it can be shown that each such set is ss_h- and stg_h-realizable. Note that this is not the case for i-semi-stable and i-stage semantics (as a counter-example, consider $e = \bar{\varphi}$ and $X = \{x\}$ in Figure 6.7).

Lemma 6.3.10. *For an instance* $\exists X \forall Y \exists Z \varphi(X, Y, Z)$ *of* $\mathsf{QSAT}_3^{\exists}$ *we let* $\mathcal{F}$ *be defined as in Reduction 6.3.8; moreover, let* $\sigma \in \{ss, stg\}$*. Then,*

$$\{S_{X'} \cup \{\varphi\} \mid X' \subseteq X\} \subseteq \sigma_i(\mathcal{F}) \subseteq \sigma_h(\mathcal{F}) = \{S_{X'} \cup \{e\} \mid X' \subseteq X, e \in \{\varphi, \bar{\varphi}\}\}.$$

Proof. We present the proof for semi-stable semantics (the proof for stage semantics is analogous). To prove the statement, let us first show that

(i) each $S \in \rho(\mathcal{F})$, $\rho \in \{ss_h, ss_i\}$ is of the form $S_{X'} \cup \{e\}$ for some $X' \subseteq X$, $e \in \{\varphi, \bar{\varphi}\}$.

Proof of (i). First notice that each $S \in \rho(\mathcal{F})$ is contained in some $S_{X'} \cup \{e\}$ for some $X' \subseteq X$, for $e \in \{\varphi, \bar{\varphi}\}$: S cannot contain both $a, \bar{a}$ for $a \in X \cup \{\varphi\}$ since there is no cf_i-realization E containing both $b, \bar{b}$, for $b \in X$, nor φ, b for $b \in \{\bar{\varphi}\} \cup \mathcal{C}$. It remains to show that $S_{X'} \cup \{e\} \subseteq S$ for some $X' \subseteq X$, for $e \in \{\varphi, \bar{\varphi}\}$.

We will first prove the statement for inherited semantics. Let $V = X \cup Y \cup Z$. Let $S \in ss_i(\mathcal{F})$ and consider a ss_i-realization E of S. First notice that E contains either φ or $\bar{\varphi}$ (otherwise, we can consider $D = E \cup \{\bar{\varphi}\}$ as a proper admissible extension of E with $D^{\oplus}_{F} \supset E^{\oplus}_{F}$). Next, we show that E contains $V' \cup \{\bar{v} \mid v \notin V'\}$ for some $V' \subseteq V$: assume there is $v \in V$ such that $v, \bar{v} \notin E$. Let $D = (E \setminus \{cl_i \mid (v, cl_i) \in R\}) \cup \{v\}$. D is conflict-free since $\bar{v}, d_v \notin E$ and since $cl_i \notin E$ for each clause cl_i with $(v, cl_i) \in R$. Moreover, it holds that D defends itself since v defends itself against the attack from $\bar{v}$, also, removing the clause-arguments does not change admissibility of the remaining arguments since the only argument they defend (namely $\bar{\varphi}$) also defends itself. Furthermore, each clause-argument cl_i is attacked by D and thus $D^{\oplus}_{F} \supset E^{\oplus}_{F}$, contradiction to E being semi-stable in F.

To show the statement for hybrid semantics, let $S \in ss_h(\mathcal{F})$ and consider a ss_h-realization E of S in $\mathcal{F}$. First observe that S contains φ or $\bar{\varphi}$ (otherwise, consider $D = E \cup \{\bar{\varphi}\}$ satisfying admissibility and $D^{\circledast}_{\mathcal{F}} = E^{\circledast}_{\mathcal{F}} \cup \{\varphi, \bar{\varphi}\} \supset E^{\circledast}_{\mathcal{F}}$, contradiction to S being h-semi-stable). Next we show that S contains $X' \cup \{\bar{x} \mid x \notin X'\}$: towards a contradiction, assume there is $x \in X$ such that $x, \bar{x} \notin S$. In case $\varphi \in S$, it holds that $\varphi \in E$ and $\bar{\varphi} \notin E$, $cl_i \notin E$, $i \leq n$, since they are in conflict with φ. Then $D = E \cup \{x\}$ is admissible and properly extends E, thus $D^{\circledast}_{\mathcal{F}} \supset E^{\circledast}_{\mathcal{F}}$, contradiction to S being h-semi-stable. In case $\bar{\varphi} \in E$, let $D = (E \setminus \{cl_i \mid (x, cl_i) \in R\}) \cup \{x, \bar{\varphi}\}$, i.e., we remove all clause-arguments attacked by x and add the argument $\bar{\varphi}$ instead. Similar as above we obtain a contradiction to $S \in \sigma_h(\mathcal{F})$ since $\{x, \bar{x}\} \in D^{\circledast}_{\mathcal{F}}$ and $cl(D) = cl(E) \cup \{x\}$, moreover, D defeats all claims defeated by E in $\mathcal{F}$, implying that $D^{\circledast}_{\mathcal{F}} \supset E^{\circledast}_{\mathcal{F}}$. Finally, we show $Y \cup Z \subseteq S$: towards a contradiction, assume that there is some $v \in Y \cup Z$ such that $v \notin S$. Hence $v, \bar{v} \notin E$. Now, we can proceed analogous to above to derive a contradiction to S being h-semi-stable. $\Diamond$

Next we show that, for all $X' \subseteq X$,

(ii) each set of the form $S_{X'} \cup \{\varphi\}$ is i-semi-stable in $\mathcal{F}$; and

(iii) each set of the form $S_{X'} \cup \{e\}$, $e \in \{\varphi, \bar{\varphi}\}$ is h-semi-stable in $\mathcal{F}$.

Proof of (ii). For this, fix some set $X' \subseteq X$ and let $E = X' \cup Y' \cup Z' \cup \{\bar{v} \mid v \notin X' \cup Y' \cup Z'\} \cup \{\varphi\}$ for some $Z' \subseteq Z$ and $Y' \subseteq Y$ with $y_0 \in Y'$. E is conflict-free, moreover, E defends φ as $y_0 \in cl_i$ for all $i \leq n$, thus E is admissible. Moreover, E has $\subseteq$-maximal range since $E^{\oplus}_{F} = V \cup \bar{V} \cup \{d_a \mid a \in E \cap (X \cup \bar{X} \cup Y \cup \bar{Y})\} \cup \mathcal{C} \cup \{\varphi, \bar{\varphi}, d_1\}$:

towards a contradiction, assume there is $D \in ad(F)$ $(D \in cf(F))$ with $D_F^{\oplus} \supset E_F^{\oplus}$, that is, there is $e \in \{d_2\} \cup \{d_a \mid a \in (X \cup \bar{X} \cup Y \cup \bar{Y}) \setminus E\}$ such that $e \in D_F^{\oplus}$; in particular, $e \in D_F^{+}$ because all considered arguments are self-attacking. Observe that $d_2 \notin D_F^{+}$ since its only attacker is self-attacking. In case $e = d_a$ for some $a \in (X \cup \bar{X} \cup Y \cup \bar{Y}) \setminus E$ we have $a \in D$ and $\bar{a} \in D$ and thus D is conflicting, contradiction to D being conflict-free. Thus $cl(E) = S_{X'} \cup \{\varphi\}$ is i-semi-stable in F. $\Diamond$

Proof of (iii). Let $X' \subseteq X$. We first show that $S_{X'} \cup \{\bar{\varphi}\}$ is h-semi-stable in $\mathcal{F}$. Consider some $Y' \subseteq Y$, $Z' \subseteq Z$ and let $\mathcal{C}' \subseteq \mathcal{C}$ denote the set of clauses cl_i which are not attacked by $X' \cup Y' \cup Z' \cup \{\bar{v} \mid v \notin X' \cup Y' \cup Z'\}$. Let $E = X' \cup Y' \cup Z' \cup \{\bar{v} \mid v \notin X' \cup Y' \cup Z'\} \cup \mathcal{C}' \cup \{\bar{\varphi}\}$. Then E is admissible, $cl(E) = S_{X'} \cup \{\bar{\varphi}\}$, and $E_{\mathcal{F}}^{*} = \{d_a \mid a \in X' \cup Y' \cup Z' \cup \{\bar{v} \mid v \notin X' \cup Y' \cup Z'\}\} \cup \{\varphi\}$. Thus $cl(E) \cup E_{\mathcal{F}}^{*}$ is subset-maximal among admissible sets since it contains every claim $c \in cl(A)$ which is assigned to non-self-attacking arguments; moreover, it contains a maximal set of claims among $\{d_v \mid v \in V \cup \bar{V}\}$ since it contains precisely one of $d_v, d_{\bar{v}}$ for each $v \in V$; furthermore observe that $d \notin E_{\mathcal{F}}^{*}$ for all conflict-free sets $E \subseteq A$ since $d_2 \notin E_F^{+}$ for every $E \in cf(F)$. It follows that $S_{X'} \cup \{\bar{\varphi}\}$ is h-semi-stable. In a similar way we show that $S_{X'} \cup \{\varphi\}$ is h-semi-stable in $\mathcal{F}$. Let $E = X' \cup Y' \cup Z' \cup \{\bar{v} \mid v \notin X' \cup Y' \cup Z'\} \cup \{\varphi\}$ for some $Z' \subseteq Z$ and $Y' \subseteq Y$ with $y_0 \in Y'$. Then E defends φ as $y_0 \in cl_i$ for all $i \leq n$, thus E is admissible. Moreover, $cl(E) = S_{X'} \cup \{\varphi\}$ and $E_{\mathcal{F}}^{*} = \{d_a \mid a \in X' \cup Y' \cup Z' \cup \{\bar{v} \mid v \notin X' \cup Y' \cup Z'\}\} \cup \{\bar{\varphi}\}$. Similar as before we conclude that $E_{\mathcal{F}}^{\circledast}$ is $\subseteq$-maximal among $ad_i(\mathcal{F})$. $\Diamond$

Thus we have shown that $\{S_{X'} \cup \{\varphi\} \mid X' \subseteq X\} \subseteq \sigma_i(\mathcal{F})$ and $\sigma_h(\mathcal{F}) = \{S_{X'} \cup \{e\} \mid X' \subseteq X, e \in \{\varphi, \bar{\varphi}\}\}$, and moreover, $\sigma_i(\mathcal{F}) \subseteq \sigma_h(\mathcal{F})$ for $\sigma = ss$. $\square$

As a corollary, we obtain that both variants of semi-stable and stage semantics coincide on $\mathcal{F}$. Indeed, regarding hybrid semantics, each set $S_{X'} \cup \{e\}$ is h-stage and h-semi-stable realizable; moreover, apart from clause-arguments cl_i (which can be substituted with argument $\bar{\varphi}$ to realize the claim $\bar{\varphi}$), each other non-self-attacking argument defends itself.

Corollary 6.3.11. *Let* $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$ *be an instance of* $\mathsf{QSAT}_3^{\exists}$ *and let* $\mathcal{F}$ *be defined as in Reduction 6.3.8. Then (1)* $ss_h(\mathcal{F}) = stg_h(\mathcal{F})$*; and (2)* $ss_i(\mathcal{F}) = stg_i(\mathcal{F})$.

Proposition 6.3.12. Con_{σ}^{CAF}, $\sigma \in \{ss, stg\}$, *is* Π_3^{P}*-hard.*

Proof. Let $\mathcal{F}$ be the CAF generated by Reduction 6.3.8 from the given $\mathsf{QSAT}_3^{\exists}$ instance $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$. By Corollary 6.3.11, it suffices to prove hardness for either one of the semantics. We provide the proof for semi-stable semantics and show that Ψ is valid iff $ss_i(\mathcal{F}) \neq ss_h(\mathcal{F})$. Since $ss_i(\mathcal{F}) \subseteq ss_h(\mathcal{F})$ by Lemma 6.3.10, the latter reduces to showing that $ss_i(\mathcal{F})$ is a proper subset of $ss_h(\mathcal{F})$: we show that Ψ is valid iff there is some $X' \subseteq X$ such that $S_{X'} \cup \{\bar{\varphi}\}$ is not ss_i-realizable in $\mathcal{F}$.

Let us first assume that Ψ is valid, that is, there is $X' \subseteq X$ such that for all $Y' \subseteq Y$, there is $Z' \subseteq Z$ such that $X' \cup Y' \cup Z'$ is a model of φ. We show $S_{X'} \cup \{\bar{\varphi}\} \notin ss_i(\mathcal{F})$. Towards a contradiction, assume there is $E \in ss(F)$ with $cl(E) = S_{X'} \cup \{\bar{\varphi}\}$. Then $\bar{\varphi} \in E$. By

Lemma 6.3.9, we have $E^{\oplus}_{F} = A\setminus(\{d_a \mid a \in (X\cup\bar{X}\cup Y\cup\bar{Y})\setminus E\}\cup\{d_1, d_2\})$. Let $Y' = E\cap Y$. By assumption Ψ is valid, there is $Z' \subseteq Z$ such that $M = X'\cup Y'\cup Z'$ is a model of φ. Let $D = M\cup\{\bar{v} \mid v \notin M\}\cup\{\varphi\}$. D is conflict-free; moreover, D attacks every cl_i, $i \leq n$ since M is a model of φ, hence D is admissible in $\mathcal{F}$. Next we show that $D^{\oplus}_{F}$ is a proper superset of $E^{\oplus}_{F}$: it holds that $V\cup\bar{V} \subseteq D^{\oplus}_{F}$; also, $\mathcal{C} \subseteq D^{\oplus}_{F}$ as shown above; moreover, $\bar{\varphi}, d_1 \in D^{\oplus}_{F}$ since $\varphi \in D$. As D and E contain the same arguments $a \in X\cup\bar{X}\cup Y\cup\bar{Y}$ by construction, we furthermore have $\{d_a \mid a \in (X\cup\bar{X}\cup Y\cup\bar{Y})\setminus E\} = \{d_a \mid a \in (X\cup\bar{X}\cup Y\cup\bar{Y})\setminus D\}$. It follows that $D^{\oplus}_{F} = A\setminus(\{d_a \mid a \in (X\cup\bar{X}\cup Y\cup\bar{Y})\setminus E\}\cup\{d_2\})$. Thus D is admissible and $D^{\oplus}_{F} \supset E^{\oplus}_{F}$, contradiction to our assumption E is semi-stable in F.

Now assume Ψ is not valid. We show that for all $X' \subseteq X$, $S_{X'}\cup\{\bar{\varphi}\} \in ss_i(\mathcal{F})$. Fix $X' \subseteq X$. Since Ψ is not valid, there is $Y' \subseteq Y$ such that for all $Z' \subseteq Z$, $X'\cup Y'\cup Z'$ is not a model of φ. Fix $Z' \subseteq Z$ and let $E = X'\cup Y'\cup Z'\cup\{\bar{v} \mid v \notin X'\cup Y'\cup Z'\}\cup\mathcal{C}'\cup\{\bar{\varphi}\}$, where $\mathcal{C}' \subseteq \mathcal{C}$ contains all clauses cl_i which are not attacked by $X'\cup Y'\cup Z'\cup\{\bar{a} \mid a \notin X'\cup Y'\cup Z'\}$. Then E is admissible and $E^{\oplus}_{F} = A\setminus(\{d_a \mid a \in (X\cup\bar{X}\cup Y\cup\bar{Y})\setminus E\}\cup\{d_1, d_2\})$. We show that E is semi-stable in F: Assume there is $D \subseteq A$ with $D^{\oplus}_{F} \supset E^{\oplus}_{F}$. First observe that D attacks the same arguments d_a, $a \in X\cup\bar{X}\cup Y\cup\bar{Y}$, as E and thus $X'\cup Y' \subseteq D$. By Lemma 6.3.9 and since $D^{\oplus}_{F}$ is strictly bigger than $E^{\oplus}_{F}$, we have that $D^{\oplus}_{F} = A\setminus(\{d_a \mid a \in (X\cup\bar{X}\cup Y\cup\bar{Y})\setminus D\}\cup\{d_2\})$. It follows that $\varphi \in D$. Let $Z'' = D\cap Z$. Then $M = X'\cup Y'\cup Z''$ is a model of φ since each clause-argument is attacked. Thus φ is satisfied by M, contradiction to our initial assumption Ψ is not valid. It follows that $S_{X'}\cup\{\bar{\varphi}\} \in ss_i(\mathcal{F})$ for all $X' \subseteq X$. Thus $ss_i(\mathcal{F}) = ss_h(\mathcal{F})$ by Lemma 6.3.10. □

6.3.2 Concurrence of Well-formed CAFs

For well-formed CAFs, h-preferred and i-preferred as well as all considered variants of stable semantics coincide as shown in Section 4.1 thus the respective problems become trivial. Since for semi-stable and stage semantics, the complexity for verification drops for both variants, we obtain Π_2^{P}-membership results by using the same generic membership argument as for general CAFs. As coNP-hardness of deciding concurrence for naive semantics has been proven in [126] it remains to show matching hardness results for semi-stable and stage concurrence. This is by a reduction from $\mathsf{QSAT}_2^{\forall}$ with some appropriate adaptions of Reduction 6.2.11.

Reduction 6.3.13. *Let $\Psi = \forall Y \exists Z \varphi(Y, Z)$ be an instance of $\mathsf{QSAT}_2^{\forall}$, where φ is given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $X = Y\cup Z$. Let (A, R) be the AF constructed from φ as in Reduction 6.2.11. We define $\mathcal{F} = (A', R', cl)$ with*

$$\begin{aligned}
A' &= A\cup\{e, d_1, d_2, \bar{\varphi}_1, \bar{\varphi}_2\}\\
R' &= R\cup\{(a, d_a)(d_a, d_a) \mid a \in Y\cup\bar{Y}\}\cup\{(d_i, d_j) \mid i, j = 1, 2\}\cup\\
&\quad \{(a, b) \mid a, b \in \{\varphi, \bar{\varphi}_1, \bar{\varphi}_2\}, a \neq b\}\cup\{(\varphi, e), (e, e), (\varphi, d_1), (\bar{\varphi}_1, d_1)\}
\end{aligned}$$

and $cl(d_1) = cl(d_2) = d$ and $cl(v) = v$ otherwise.

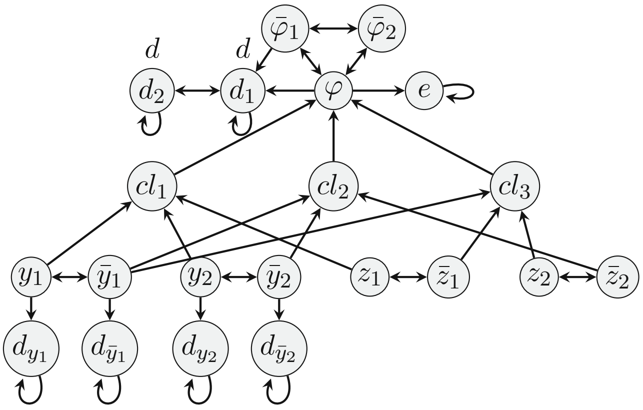


Figure 6.8: Reduction 6.3.13 for the formula $\forall Y \exists Z \varphi(Y, Z)$ where $\varphi(Y, Z)$ is given by the clauses $\{\{z_1, y_1, y_2\}, \{\bar y_1, \bar y_2, \bar z_2)\}, \{\bar z_1, \bar y_1, z_2\}\}$. Since $cl(a) = a$ for all arguments $a \in A \setminus \{d_1, d_2\}$, we omit all claims that coincide with the arguments name.

An example is given in Figure 6.8. Conflict-free claim-sets in $\mathcal{F}$ admit a close correspondence to their realizations in F since all arguments except the self-attackers d_1 and d_2 have been assigned unique claims. The following observations are easy to verify.

Lemma 6.3.14. *Let* $\Psi = \forall Y \exists Z \varphi(Y, Z)$ *be an instance of* $\mathsf{QSAT}_2^\forall$, *let* $\sigma \in \{ss, stg\}$ *and let* $\mathcal{F} = (A, R, cl)$ *be as in Reduction 6.3.13. Then*

1. *for all* $E \in cf(F)$, $(cl(E))^+_{\mathcal{F}} = E^+_F \setminus \{d_1\}$;
2. *every* $S \in cf_i(\mathcal{F})$ *admits a unique realization in* F;
3. *for all* $S \in \sigma_i(\mathcal{F}) \cup \sigma_h(\mathcal{F})$, *either* $\varphi \in S$ *or* $\varphi_1 \in S$ *or* $\varphi_2 \in S$.

The following two lemmata will be useful to prove Π_2^{P}-hardness of the concurrence problem for semi-stable and stage semantics. First, we show that each i-semi-stable (i-stage) claim-set is h-semi-stable (h-stage).

Lemma 6.3.15. *Let* $\Psi = \forall Y \exists Z \varphi(Y, Z)$ *be an instance of* $\mathsf{QSAT}_2^\forall$, *let* $\sigma \in \{ss, stg\}$ *and let* $\mathcal{F} = (A, R, cl)$ *be as in Reduction 6.3.13. Then* $\sigma_i(\mathcal{F}) \subseteq \sigma_h(\mathcal{F})$.

Proof. Let $F = (A, R)$, consider $S \in \sigma_i(\mathcal{F})$ and let E denote the unique σ_i-realization of S in F. We provide the proof for semi-stable semantics; the proof for stage semantics is analogous. As $E \in ss(F)$, we have that $E^\oplus_F$ is subset-maximal among admissible extensions. We will show that $S^\circledast_{\mathcal{F}}$ is $\subseteq$-maximal among i-admissible claim-sets. Towards a contradiction, assume there is $T \in ad_i(\mathcal{F})$ with $T^\circledast_{\mathcal{F}} \supset S^\circledast_{\mathcal{F}}$. Consider the unique ad_i-realization D of T in F, then $D^\oplus_F \setminus \{d_1\} = T^\circledast_{\mathcal{F}} \supset S^\circledast_{\mathcal{F}} = E^\oplus_F \setminus \{d_1\}$. If either $d_1 \in D^+_F$ or $d_1 \notin E^+_F$ we have $D^\oplus_F \supset E^\oplus_F$, contradiction to E being semi-stable in F. Let us assume $d_1 \in E^+_F$ but $d_1 \notin D^+_F$. By Lemma 6.3.14, we have $\varphi_2 \in D$ since φ_2 does not attack

d_1; also, $\varphi_1 \in E$ or $\varphi \in E$. In case $\varphi \in E$, we have $e \in E_F^+$, $e \notin D_F^+$ thus $e \in S_{\mathcal{F}}^{\oplus}$ but $e \notin T_{\mathcal{F}}^{\oplus}$, contradiction to the assumption $T_{\mathcal{F}}^{\circledast} \supset S_{\mathcal{F}}^{\circledast}$. In case $\varphi_2 \in D$ and $\varphi_1 \in E$, consider $D' = (D \cup \{\varphi_1\}) \setminus \{\varphi_2\}$. D' is admissible as D is admissible and exchanging φ_2 with φ_1 does neither add conflicts nor undefended arguments. Moreover, $d_1 \in (D')_F^+$ and $D_F^{\oplus} = (D')_F^{\oplus} \setminus \{d_1\}$. Therefore $(D')_F^{\oplus} \supset E_F^{\oplus}$, contradiction to $E \in ss(F)$. ☐

Lemma 6.3.16. *Let $\Psi = \forall Y \exists Z \varphi(Y, Z)$ be an instance of $\mathsf{QSAT}_2^{\forall}$, let $\sigma \in \{ss, stg\}$ and let $\mathcal{F} = (A, R, cl)$ be as in Reduction 6.3.13. Then for all $S \in \sigma_i(\mathcal{F}) \cup \sigma_h(\mathcal{F})$, $\varphi \in S$ implies $S \in \sigma_i(\mathcal{F}) \cap \sigma_h(\mathcal{F})$.*

Proof. Let $F = (A, R)$. By Lemma 6.3.15, $\sigma_i(\mathcal{F}) \subseteq \sigma_h(\mathcal{F})$ thus it suffices to prove the statement for $S \in \sigma_h(\mathcal{F})$. Let E denote the unique cf_i-realization of S in F. We will show $E \in \sigma(F)$. Towards a contradiction, assume there is $D \in ad(F)$ ($D \in cf(F)$) with $D_F^{\oplus} \supset E_F^{\oplus}$. As $\varphi \in E$ we have $d_1 \in E_F^+$ and thus $D_F^{\oplus} \setminus \{d_1\} \supset E_F^{\oplus} \setminus \{d_1\}$. By Lemma 6.3.14, $D_{\mathcal{F}}^{\circledast} = D_F^{\oplus} \setminus \{d_1\} \supset E_F^{\oplus} \setminus \{d_1\} = S_{\mathcal{F}}^{\circledast}$, contradiction to $S \in \sigma_h(\mathcal{F})$. ☐

Proposition 6.3.17. *Con_{σ}^{wf}, $\sigma \in \{ss, stg\}$, is Π_2^{P}-hard.*

Proof. Let $\Psi = \forall Y \exists Z \varphi(Y, Z)$ be an instance of $\mathsf{QSAT}_2^{\forall}$ and let $\mathcal{F} = (A, R, cl)$ be as in Reduction 6.3.13. We will show Ψ is valid iff $\sigma_i(\mathcal{F}) = \sigma_h(\mathcal{F})$.

First assume Ψ is valid. We show that in this case, $\varphi \in S$ for all $S \in \sigma_i(\mathcal{F}) \cup \sigma_h(\mathcal{F})$. By Lemma 6.3.16, this implies $S \in \sigma_i(\mathcal{F}) \cap \sigma_h(\mathcal{F})$ and thus $\sigma_i(\mathcal{F}) = \sigma_h(\mathcal{F})$. By Lemma 6.3.15, it suffices to prove the statement for every $S \in \sigma_h(\mathcal{F})$. Towards a contradiction, assume there is $S \in \sigma_h(\mathcal{F})$ such that $\varphi \notin S$. Then $e \notin S_{\mathcal{F}}^{\circledast}$. Let $Y' = S \cap Y$. Since Ψ is valid, there is $Z' \subseteq Z$ such that $Y' \cup Z'$ is a model of φ. Let $E = Y' \cup Z' \cup \{\bar{x} \mid x \notin Y' \cup Z'\} \cup \{\varphi\}$. Then $S' = cl(E)$ is i-admissible (i-conflict-free) and $S'^{\circledast}_{\mathcal{F}} = cl(A) \setminus (\{d\} \cup \{d_y \mid y \notin E\} \cup \{d_{\bar{y}} \mid \bar{y} \notin E\})$. We conclude that $S'^{\circledast}_{\mathcal{F}} \supset S_{\mathcal{F}}^{\circledast}$ since $e \notin S_{\mathcal{F}}^{\circledast}$ and $\{d\} \cup \{d_y \mid y \notin E\} \cup \{d_{\bar{y}} \mid \bar{y} \notin E\} \nsubseteq S_{\mathcal{F}}^{\circledast}$, contradiction to S is h-semi-stable (h-stage). It follows that $\varphi \in S$ for every $S \in \sigma_h(\mathcal{F})$.

Now assume Ψ is not valid, i.e., there is $Y' \subseteq Y$ such that for all $Z' \subseteq Z$, $Y' \cup Z'$ is not a model of φ. We will show that $\sigma_i(\mathcal{F}) \subset \sigma_h(\mathcal{F})$. Fix $Z' \subseteq Z$ and let $E = Y' \cup Z' \cup \{\bar{x} \mid x \notin Y' \cup Z'\}$. Moreover, let $E_1 = E \cup \mathcal{C}' \cup \{\varphi_1\}$ and $E_2 = E \cup \mathcal{C}' \cup \{\varphi_2\}$ where $\mathcal{C}' \subseteq \mathcal{C}$ contains all clauses cl_i such that $E \cap cl_i = \emptyset$. Clearly, $E_1, E_2 \in ad(F)$ ($E_1, E_2 \in cf(F)$) and thus $E_1 = cl(E_1), E_2 = cl(E_2) \in ad_i(\mathcal{F})$ ($E_1 = cl(E_1), E_2 = cl(E_2) \in cf_i(\mathcal{F})$). Observe that $(E_2)_F^{\oplus} \subset (E_1)_F^{\oplus}$ since d_1 is attacked by $\varphi_1 \in E_1$ but there is no $a \in E_2$ such that $(a, d_1) \in R$. It follows that $E_2 = cl(E_2) \notin \sigma_i(\mathcal{F})$. We show that $E_2 \in \sigma_h(\mathcal{F})$ for $\sigma \in \{ss, stg\}$, i.e., we show that $(E_2)_{\mathcal{F}}^{\circledast} = cl(A) \setminus (\{e, d\} \cup \{d_y \mid y \notin E\} \cup \{d_{\bar{y}} \mid \bar{y} \notin E\})$ is maximal among admissible (conflict-free) claim-sets: Towards a contradiction, assume there is $T \in ad_i(\mathcal{F})$ ($T \in cf_i(\mathcal{F})$) such that $T_{\mathcal{F}}^{\circledast} \supset (E_2)_{\mathcal{F}}^{\circledast}$. As $\{d_y \mid y \in Y'\} \cup \{d_{\bar{y}} \mid y \notin Y'\} \subseteq T_{\mathcal{F}}^+$ we have $Y' \cup \{\bar{y} \mid y \notin Y'\} \subseteq T$ and $T_{\mathcal{F}}^+$ does not contain any claim in $\{d_y \mid y \notin E\} \cup \{d_{\bar{y}} \mid \bar{y} \notin E\}$ since for every $y \in Y$, there is no conflict-free set attacking both d_y and $d_{\bar{y}}$. Moreover, $d \notin T_{\mathcal{F}}^+$ for every $T \in cf_i(\mathcal{F})$ since d_1 and d_2 are the only attackers of d_2 and d_1 is self-attacking. It follows that $e \in T_{\mathcal{F}}^+$ and thus $\varphi \in T$. Consider the unique cf_i-realization D of T. Since $\varphi \in D$ we have we have $cl_i \notin D$ for every $i \leq n$ and thus each cl_i is

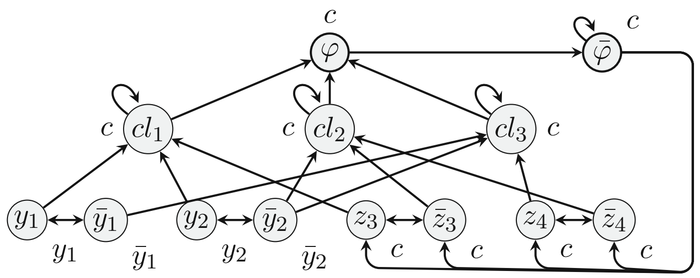


Figure 6.9: CAF from the proof of Proposition 6.3.19 for the QBF $\forall\{y_1, y_2\}\exists\{z_3, z_4\}$: $\{\{y_1, y_2, z_3\}, \{\bar{y}_2, \bar{z}_3, \bar{z}_4)\}, \{\bar{y}_1, \bar{y}_2, z_4\}\}$.

attacked by D. Thus we obtain that $M = D \cap X$ is a model of φ $Y' \subseteq M$, contradiction to our initial assumption $Y' \cup Z''$ is not a model of φ for every $Z'' \subseteq Z$. □

6.3.3 Concurrence of Stable Variants

We conclude this section by analyzing the concurrence problems for h-stable variants. That is, we ask ourselves how hard it is do decide whether the two variants of the claim-based stable semantics coincide. We write $Con^{\Delta}_{stb_h}$, $\Delta \in \{CAF, wf\}$ to denote this problem. Bearing in mind the complexity of the verification problem of the two semantics, the problem has to be contained in Π_2^{P}; however, as we show next, it also hard for this class for general CAFs. For well-formed CAFs recall that the two variants collapse anyway making this problem trivial for well-formed CAFs.

The hardness-proof relies on an appropriate claim-labelling of the standard reduction used for deciding skeptical acceptance for preferred semantics [89, Reduction 3.7].

Reduction 6.3.18. *Let $\Psi = \forall Y \exists Z \varphi(Y, Z)$ be an instance of $\mathsf{QSAT}_2^{\forall}$, where φ is given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $X = Y \cup Z$. Let (A, R) be the AF constructed from φ as in Reduction 6.2.11. We define $F = (A', R')$ with*

$$\begin{aligned} A' &= A \cup \{\bar{\varphi}\}, \text{ and} \\ R' &= R \cup \{(cl_i, cl_i) \mid i \leq n\} \cup \{(\varphi, \bar{\varphi}), (\bar{\varphi}, \bar{\varphi})\} \cup \{(\bar{\varphi}, z) \mid z \in Z\}. \end{aligned}$$

Proposition 6.3.19. *$Con^{CAF}_{stb_h}$ is Π_2^{P}-complete.*

Proof. To show hardness, let (A, R) denote the AF from Reduction 6.3.18. We define $\mathcal{F} = (A, R, cl)$ with $cl(y) = y$ for $y \in Y$, $cl(\bar{y}) = \bar{y}$ for $\bar{y} \in \bar{Y}$, and $cl(v) = c$ otherwise. See Figure 6.9 for an illustrative example of the reduction.

We show Ψ is valid iff $ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$. To do so, let us first prove

(i) for all $Y' \subseteq Y$, $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in cf\text{-}stb_h(\mathcal{F})$. Moreover, there is no other cf-h-stable claim-set in $\mathcal{F}$.

Proof of (i). Let $Y' \subseteq Y$ be arbitrary, let $z \in Z$ and let $E = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{z\}$. Clearly, E is conflict-free in F; moreover, E attacks every $a \in A$ such that $cl(a) \notin cl(E)$. It follows that $cl(E) = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in cf_i(\mathcal{F})$. Moreover, $cl(E)$ is maximal among all conflict-free claim-sets: Assume there is $T \in cf_i(\mathcal{F})$ such that $T \supset cl(E)$ for some $Y' \subseteq Y$. Then there is $y \in Y$ such that $y \in T$ and $\bar{y} \in T$, contradiction to cf-realizability of T since for every $y \in Y$, y and $\bar{y}$ mutually attack each other. We can furthermore conclude that no other h-stable claim-set exists since for every $y \in Y$, y and $\bar{y}$ mutually attack each other. Thus each cf-h-stable claim-set is of the form $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\}$ for some $Y' \subseteq Y$. $\Diamond$

First assume Ψ is valid. We show $stb_i(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$ ($ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$ follows since $stb_i(\mathcal{F}) \subseteq ad\text{-}stb_h(\mathcal{F}) \subseteq cf\text{-}stb_h(\mathcal{F})$). Let $Y' \subseteq Y$. Then there is $Z' \subseteq Z$ such that φ is satisfied by $M = Y' \cup Z'$. Let $E = M \cup \{\bar{x} \mid x \notin M\} \cup \{\varphi\}$. Since M satisfies each clause cl_i, there is either $x \in cl_i$ with $x \in M$ or there is $\bar{x} \in cl_i$ with $x \notin M$. It follows that each cl_i, $i \leq n$, is attacked by E; moreover, E attacks $\bar{\varphi}$ since $\varphi \in E$. Since E is also conflict-free we have shown that E is a stable extension of F and therefore $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in stb_i(\mathcal{F})$. As Y' was arbitrary, we have that $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in stb_i(\mathcal{F})$ for all $Y' \subseteq Y$. We conclude that $stb_i(\mathcal{F}) = ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$ by (i).

Now assume $ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$ and let $Y' \subseteq Y$. By (i) we have that $S = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \in ad\text{-}stb_h(\mathcal{F}) = cf\text{-}stb_h(\mathcal{F})$. Consider an ad-realization E of S and let $Z' = E \cap Z$. We show that $M = Y' \cup Z'$ satisfies φ. First observe that $\varphi \in E$: Since $c \in S$, there is some $a \in A$ with $cl(a) = c$ such that $a \in E$. Moreover, $a \in Z \cup \bar{Z} \cup \{\varphi\}$ since every other claim assigned with c is self-attacking. In case $a = \varphi$, we are done; in case $a = z$ or $a = \bar{z}$ for some $z \in Z$ we have $\varphi \in E$ since E defends a against $\bar{\varphi}$. Since $\varphi \in E$, we furthermore have that E attacks each clause cl_i since φ is defended by E against cl_i. We obtain that M is a model of φ. We have shown that for every $Y' \subseteq Y$, there is $Z' \subseteq Z$ such that $Y' \cup Z'$ satisfies φ. It follows that Ψ is valid. $\square$

6.4 Summary

In this chapter, we studied the computational complexity of hybrid semantics for CAFs. We want to highlight three observations here: (a) for both approaches the verification problem is harder than in the AF setting, which is in particular relevant when it comes to the enumeration of extensions; (b) however, when restricted to well-formed CAFs the complexity of verification drops to the complexity of AFs; and (c) the complexity of inherited and hybrid semantics differs for naive and preferred semantics.

Besides studying the standard reasoning tasks we also settled the complexity of the concurrence problem, which turns out to be surprisingly hard, ranging up to the third level of the polynomial hierarchy. The concurrence problem is in the tradition of the well-known coherence problem [81], which (a) for AFs is Π_2^P-complete; (b) remains Π_2^P-complete for inherited semantics [92]; and (c) also for hybrid semantics, despite the complexity increase for reasoning problems, remains Π_2^P-complete (Proposition 6.2.25).

CHAPTER 7

Dynamics Part I: Strong Equivalence

Equivalence is an important subject of research in knowledge representation and reasoning. Given a knowledge base $\mathcal{K}$, finding an equivalent one, say $\mathcal{K}'$, helps to obtain a better understanding as well as a more concise representation of $\mathcal{K}$. From a computational point of view, equivalence is particularly interesting whenever a certain subset of a collection of information can be replaced without changing the intended meaning. In propositional logic, for example, replacing a sub-formula ϕ of Φ with an equivalent one, say ϕ', yields a formula $\Phi[\phi/\phi']$ equivalent to Φ. That is, we may view ϕ as an independent module of Φ. Within the KR community it is folklore that this is usually not the case for non-monotonic logics (apart from folklore, we refer the reader to [21] for a rigorous study of this matter).

Driven by this observation, the notion of *strong* equivalence has been proposed, developed and investigated in various contexts [132, 142]. In a nutshell, strong equivalence requires the aforementioned property by design: $\mathcal{K}$ and $\mathcal{K}'$ are strongly equivalent if for any $\mathcal{H}$, the knowledge bases $\mathcal{K} \cup \mathcal{H}$ and $\mathcal{K}' \cup \mathcal{H}$ are equivalent. Hence, knowing that two frameworks are strongly equivalent to each other ensures that they yield the same outcome in a dynamic setting in which knowledge bases expand over time. Although a naive implementation would require to iterate over an infinite number of possible expansions $\mathcal{H}$, researchers discovered techniques to decide strong equivalence of two knowledge bases efficiently in the context of argumentation [142, 93].

In this chapter, we investigate the strong equivalence problem from a claim-centered point of view. First, we focus on ordinary equivalence and discuss several dependencies in Section 7.1 for general as well as for well-formed CAFs. In Section 7.2, we provide characterization results of strong equivalence between CAFs via semantics-dependent kernels for each CAF semantics under consideration. In Section 7.3, we introduce a novel equivalence concept based on argument renaming which is genuine for CAFs, and show that strong equivalence up to renaming can be characterized via kernel isomorphism. Finally, we present a complexity analysis of deciding equivalence for all of the aforementioned equivalence notions in Section 7.4. Knowing the computational complexity of deciding equivalence between two frameworks has several advantages. From a practical point of view, an exact complexity analysis is crucial for efficient algorithm design. Due to our kernel characterizations, we obtain that deciding strong equivalence for CAFs is tractable. Our novel notion strong equivalence up to renaming, on the other hand, has the same complexity as the graph isomorphism problem. We obtain a tractable fragment of this problem when restricting the notion to well-formed CAFs. Finally, we show that deciding ordinary equivalence can be computationally hard, up to the third level of the polynomial hierarchy. Here, our complexity analysis reveals that shortcuts that give insight about the equivalence of two frameworks (for instance, by exploiting the graph-structure of the CAFs) are unlikely to find. To decide whether two frameworks yield the same outcome, it is necessary to compute the claim-extensions explicitly.

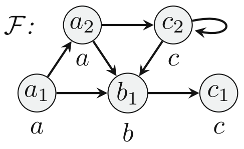
Background & Notation. *In this chapter, we make use of concepts and results from Chapter 4 where we introduced* hybrid semantics (h-semantics) *for CAFs. Background on complexity theory and corresponding definitions can be found in Section 6.1.*

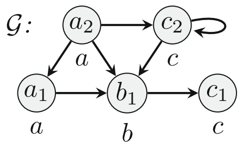
7.1 Ordinary Equivalence

The distinction between explicit and implicit information is essential in knowledge representation. The former is interpreted according to the underlying semantics of the considered formalism, i.e. the set of models in case of classical propositional logic or the set of extensions in case of classical AFs. In contrast, the implicit information of an knowledge base comes to light if it undergoes dynamic changes. Both concepts come along with an induced notion of equivalence, namely *ordinary* or *strong equivalence*, respectively. We start our analysis by investigating ordinary equivalence for CAFs.

Definition 7.1.1. *Two CAFs $\mathcal{F}$ and $\mathcal{G}$ are* ordinarily equivalent *w.r.t. semantics ρ, in symbols $\mathcal{F} \equiv_o^\rho \mathcal{G}$, if we have $\rho(\mathcal{F}) = \rho(\mathcal{G})$.*

Example 7.1.2. *Consider the following CAFs $\mathcal{F}$ and $\mathcal{G}$. Note that they disagree on the attack relation between a_1 and a_2 only.*





We have $stb(F_\mathcal{F}) = \emptyset$ and $stb(G_\mathcal{G}) = \{a_2, c_1\}$. Consequently, the inherited variants are $stb_i(\mathcal{F}) = \emptyset$ and $stb_i(\mathcal{G}) = \{a, c\}$ justifying $\mathcal{F} \not\equiv_o^{stb_i} \mathcal{G}$. If we consider instead the claim-based versions, we observe that the two CAFs agree on their outcome: More precisely, due to $stb_i(\mathcal{G}) \subseteq ad\text{-}stb_h(\mathcal{G}) \subseteq cf\text{-}stb_h(\mathcal{G})$ we obtain $\{a, c\} \in ad\text{-}stb_h(\mathcal{G}), cf\text{-}stb_h(\mathcal{G})$. Moreover, *we have that* $\{a, c\} \in ad\text{-}stb_h(\mathcal{F}), cf\text{-}stb_h(\mathcal{F})$ *since the set* $\{a_1, c_1\}$ *is admissible (thus, conflict-free) and defeats every remaining claim. As a side remark, we mention that the claim-set* $\{a, c\}$ *has two* cf_i*-realizations in* $\mathcal{F}$ *and* $\mathcal{G}$ *since both of the sets* $\{a_1, c_1\}$, $\{a_2, c_1\}$ *are conflict-free and have full claim-range. It can be checked that no other claim-set then* $\{a, c\}$ *satisfies the requirements of the claim-based stable versions. Consequently,* $\mathcal{F}$ *and* $\mathcal{G}$ *are ordinarily equivalent with respect to* $ad\text{-}stb_h$ *and* $cf\text{-}stb_h$ *semantics, in symbols:* $\mathcal{F} \equiv_o^{ad\text{-}stb_h} \mathcal{G}$ *and* $\mathcal{F} \equiv_o^{cf\text{-}stb_h} \mathcal{G}$.

In the following we consider (non-)relations between ordinary equivalences w.r.t. different semantics. We will see that the inherited variants behave differently in comparison to claim-based versions. Let us recap the case of Dung-style AFs. It was shown that sharing the same admissible/conflict-free sets guarantees no difference regarding preferred/naive extensions. Moreover, equivalence with respect to naive sets implies that the conflict-free sets coincide. Also, possessing the same complete extensions implies coinciding grounded and preferred extensions [142, Proposition 1].

Let us start with the relations between inherited semantics. We can transfer the following relations from the case for the respective AF semantics:

Proposition 7.1.3. *Consider two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$*. It holds that*

1. $\mathcal{F} \equiv_o^{co_i} \mathcal{G} \Rightarrow \mathcal{F} \equiv_o^{gr_i} \mathcal{G}$,
2. $\mathcal{F} \equiv_o^{na_i} \mathcal{G} \Rightarrow \mathcal{F} \equiv_o^{cf_i} \mathcal{G}$.

Interestingly, we observe that not all relations for AF semantics presented in [142] carry over to inherited semantics. This is due to the fact that i-preferred (i-naive) semantics are not necessarily $\subseteq$-maximal i-admissible (i-conflict-free) claim-sets. Let us consider the following example.

Example 7.1.4. *Assume we are given two CAFs as follows:*

We have $ad_i(\mathcal{F}) = ad_i(\mathcal{G}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$*. On the other hand,* $\{a, b\}$ *is the unique i-preferred claim-set of* $\mathcal{F}$ *while* $pr_i(\mathcal{G}) = \{\{a\}, \{a, b\}\}$ *witnessed by the extensions* $\{a_1, a_2\}$ *and* $\{a_1, b_1\}$*. Thus* $\mathcal{F} \equiv_o^{ad_i} \mathcal{G} \not\Rightarrow \mathcal{F} \equiv_o^{pr_i} \mathcal{G}$*. The example furthermore shows* $\mathcal{F} \equiv_o^{cf_i} \mathcal{G} \not\Rightarrow \mathcal{F} \equiv_o^{na_i} \mathcal{G}$ *since* cf_i *and* ad_i *as well as the respective variants of naive and preferred semantics coincide in* $\mathcal{F}$ *and* $\mathcal{G}$*.*

Let us next consider relations between inherited and claim-based semantics. Overall, we observe that equivalence with respect to cl-preferred semantics can be decided by looking either at i-admissible, i-complete, or i-preferred semantics. Moreover, coincidence of i-naive extension implies equivalence with respect to cl-naive semantics. Also, inherited conflict-free sets coincide if and only if cl-naive semantics yield the same claim-sets.

Proposition 7.1.5. *Consider two CAFs $\mathcal{F}$ and $\mathcal{G}$. It holds that*

1. $\mathcal{F} \equiv_o^{\rho} \mathcal{G} \Rightarrow \mathcal{F} \equiv_o^{pr_h} \mathcal{G}$,$\rho \in \{ad_i, pr_i, co_i\}$,
2. $\mathcal{F} \equiv_o^{cf_i} \mathcal{G} \Leftrightarrow \mathcal{F} \equiv_o^{na_h} \mathcal{G}$,
3. $\mathcal{F} \equiv_o^{na_i} \mathcal{G} \Rightarrow \mathcal{F} \equiv_o^{na_h} \mathcal{G}$.

Proof. First, assume $\mathcal{F} \equiv_o^{ad_i} \mathcal{G}$. By definition, cl-preferred extensions are the $\subseteq$-maximal i-admissible extensions, hence $\mathcal{F} \equiv_o^{pr_h} \mathcal{G}$ follows. Since cl-preferred extensions coincide with the $\subseteq$-maximal i-preferred and i-complete claim-sets for each CAF, we obtain that $\mathcal{F} \equiv_o^{pr_i} \mathcal{G}$ and $\mathcal{F} \equiv_o^{co_i} \mathcal{G}$ imply $\mathcal{F} \equiv_o^{pr_h} \mathcal{G}$.

Let us next consider the relation between i-conflict-free and cl-naive semantics. By definition, cl-naive extensions are the $\subseteq$-maximal i-conflict-free extensions, hence we obtain $\mathcal{F} \equiv_o^{cf_i} \mathcal{G}$ implies $\mathcal{F} \equiv_o^{na_h} \mathcal{G}$. For the other direction, note that each subset of a cl-naive extension has a conflict-free realization, hence the statement follows.

Finally, we note that cl-naive extensions are precisely the $\subseteq$-maximal i-naive extensions, which implies the equivalence in the last item. ☐

We furthermore obtain the following relations between ordinary equivalences when considering well-formed CAFs.

Proposition 7.1.6. *For any two well-formed CAFs $\mathcal{F}$ and $\mathcal{G}$, it holds that*

- $\mathcal{F} \equiv_o^{\rho} \mathcal{G} \Rightarrow \mathcal{F} \equiv_o^{pr_i} \mathcal{G}$, $\rho \in \{ad_i, co_i, pr_h\}$;
- $\mathcal{F} \equiv_o^{stb_i} \mathcal{G} \Leftrightarrow \mathcal{F} \equiv_o^{ad\text{-}stb_h} \mathcal{G} \Leftrightarrow \mathcal{F} \equiv_o^{cf\text{-}stb_h} \mathcal{G}$.

Proof. The relations follow since the variants of preferred as well as the variants of stable semantics collapse for well-formed CAFs. ☐

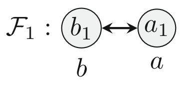
Let us now turn to the non-relations between the semantics. Negative results (i.e., counter-examples) generalize to CAFs from the corresponding AF semantics.

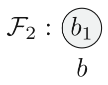
Lemma 7.1.7. *For two AF semantics σ and τ, if $\sigma(F) = \sigma(G) \not\Rightarrow \tau(F) = \tau(G)$ for some AFs F, G, then $\sigma_c(\mathcal{F}) = \sigma_c(\mathcal{G}) \not\Rightarrow \tau_c(\mathcal{F}) = \tau_c(\mathcal{G})$ for some CAFs $\mathcal{F}$, $\mathcal{G}$.*

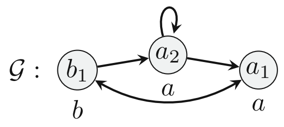
Indeed, when identifying AFs with CAFs where each claim is unique (i.e., taking $cl = id$), we obtain counter-examples from known results for AFs (cf. [142]). We furthermore recall that in this case, claim-level semantics coincide with their inherited counterparts. It remains to provide counter-examples for naive, semi-stable, and stage semantics as well as for stable semantics in the general case.

For naive semantics, we observe that in both CAFs from Example 7.1.4, preferred and naive semantics coincide in both variants. To separate the stable variants, we consider the following examples.

Example 7.1.8. *Consider the following CAFs $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{G}$:*

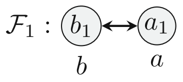


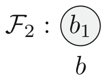


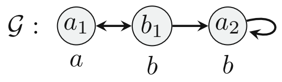


It holds that $stb_{hcf}(\mathcal{F}_1) = stb_{hcf}(\mathcal{G}) = \{\{a\}, \{b\}\}$ but $\rho(\mathcal{F}_1) \neq \rho(\mathcal{G})$ for $\rho \in \{stb_c, stb_{had}\}$; moreover, $\rho(\mathcal{F}_2) = \rho(\mathcal{G}) = \{\{a\}\}$ and $stb_{hcf}(\mathcal{F}_2) \neq stb_{hcf}(\mathcal{G})$.

Example 7.1.9. *Consider the following CAFs $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{G}$:*



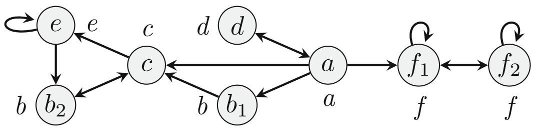




It holds that $stb_{had}(\mathcal{F}_1) = stb_{had}(\mathcal{G}) = \{\{a\}, \{b\}\}$ but $stb_c(\mathcal{F}_1) \neq stb_c(\mathcal{G})$; moreover, $stb_c(\mathcal{F}_2) = stb_c(\mathcal{G}) = \{\{a\}\}$ but $stb_{had}(\mathcal{F}_2) \neq stb_{had}(\mathcal{G})$.

It remains to consider semi-stable and stage semantics.

Example 7.1.10. *Consider the following (well-formed) CAF $\mathcal{F}$:*



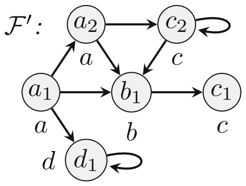
In $\mathcal{F}$, it holds that $ss_c(\mathcal{F}) = \{\{a\}\}$, $ss_h(\mathcal{F}) = \{\{b, d\}\}$, $stg_c(\mathcal{F}) = \{\{c\}, \{a\}\}$, and $stg_h(\mathcal{F}) = \{\{b, d\}, \{c\}\}$. To obtain counter-examples for the involved semantics, it suffices to construct a (well-formed) CAF $\mathcal{G}$ in which both variants agree on one of the aforementioned claim-sets of $\mathcal{F}$. First, let $\mathcal{G}_1 = (\{a\}, \emptyset, id)$, then all considered semantics return claim-set $\{a\}$. Thus $ss_c(\mathcal{F}) = ss_c(\mathcal{G})$ but $ss_h(\mathcal{F}) \neq ss_h(\mathcal{G})$. Likewise, we let $\mathcal{G}_2 = (\{b, d\}, \emptyset, id)$ to obtain a counter-example for the other direction. For stage semantics, we consider the CAFs $\mathcal{G}_3 = (\{a, c\}, \{(a, c), (c, a)\}, id)$ and $\mathcal{G}_4 = (\{b, c, d\}, \{(b, c), (c, d), (d, c), (c, d)\}, id)$ instead.

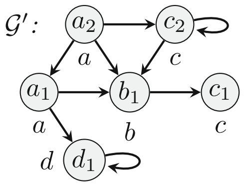
This concludes our study of relations between semantics with respect to ordinary equivalence. We considered both general and well-formed CAFs. Similar as for AFs, we observe that ordinary equivalence for CAF semantics are largely independent of each other.

7.2 Strong Equivalence

Strong equivalence has been introduced as a non-monotonic counter-part to classical equivalence in monotonic formalisms. In contrast to classical logic in which equivalent formulae are interchangeable, ordinary equivalence in argumentation is not robust when it comes to expansions of the frameworks, e.g., if an update in the knowledge base induces new arguments or attacks. Let us illustrate this at the following example:

Example 7.2.1. *Assume we are given an updated version of* $\mathcal{F}$ *and* $\mathcal{G}$ *from Example 7.1.2 where an additional argument has been introduced. Let* $\mathcal{F}'$ *and* $\mathcal{G}'$ *be given as follows:*





$\mathcal{F}'$ *and* $\mathcal{G}'$ *no longer agree on their h-ad-stable claim-sets: In* $\mathcal{G}'$*, the set* $\{a_2, c_1\}$ *does not defeat claim d, thus* $\textit{ad-stb}_h(\mathcal{G}') = \emptyset$ *while* $\{a, c\}$ *is h-ad-stable in* $\mathcal{F}'$*. Note that* $\textit{cf-stb}_h(\mathcal{F}') = \textit{cf-stb}_h(\mathcal{G}')$*, i.e., they remain equivalent w.r.t. h-cf-stable semantics.*

In the light of this issue, it is evident that a stronger notion is needed to handle equivalence between CAFs in a dynamical setting. In accordance with standard literature on *strong equivalence* in other non-monotonic formalisms [132, 142], we will call two CAFs strongly equivalent to each other if they possess the same extensions independently of any such (simultaneous) expansion of the frameworks. Before we can define this concept formally, we need to take care of the situation that the same argument has been assigned a different claim in the frameworks $\mathcal{F}$, $\mathcal{G}$ under consideration.

Definition 7.2.2. *Two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *are* compatible *to each other if* $cl_{\mathcal{F}}(a) = cl_{\mathcal{G}}(a)$ *for all* $a \in A_{\mathcal{F}} \cap A_{\mathcal{G}}$*. The union* $\mathcal{F} \cup \mathcal{G}$ *of two compatible CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *is defined componentwise, i.e.,* $\mathcal{F} \cup \mathcal{G} = (A_{\mathcal{F}} \cup A_{\mathcal{G}}, R_{\mathcal{F}} \cup R_{\mathcal{G}}, cl_{\mathcal{F}} \cup cl_{\mathcal{G}})$*.*

We are ready to introduce strong equivalence for CAFs.

Definition 7.2.3. *Two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *are* strongly equivalent *to each other w.r.t. a semantics* ρ*, in symbols* $\mathcal{F} \equiv^{\rho}_{s} \mathcal{G}$*, iff*

1. $\mathcal{F}$ *and* $\mathcal{G}$ *are compatible with each other; and*
2. $\rho(\mathcal{F} \cup \mathcal{H}) = \rho(\mathcal{G} \cup \mathcal{H})$ *for each CAF* $\mathcal{H}$ *which is compatible with* $\mathcal{F}$ *and* $\mathcal{G}$*.*

The definition extends strong equivalence for AFs. With a slight abuse of notation we also use $F \equiv^{\sigma}_{s} G$ to denote strong equivalence of two AFs F and G w.r.t. the semantics σ.

Strong equivalence for AFs has been characterized via syntactic equivalence of so-called (semantics-dependent) kernels, which are obtained by syntactical modifications (attack-removal or -addition) of the given frameworks. Let us recall the definitions of the stable, admissible, complete, grounded, and naive kernel [142, 23].

Definition 7.2.4. *For an AF* $F = (A, R)$*, we define the* stable kernel $F^{sk} = (A, R^{sk})$*;* admissible kernel $F^{ak} = (A, R^{ak})$*; the* complete kernel $F^{gk} = (A, R^{gk})$*;* grounded kernel $F^{gk} = (A, R^{gk})$*; and the* naive kernel $F^{nk} = (A, R^{nk})$ *with*

$$\begin{aligned}
R^{sk} &= R \setminus \{(a,b) \mid a \neq b, (a,a) \in R\}\\
R^{ak} &= R \setminus \{(a,b) \mid a \neq b, (a,a) \in R, \{(b,a),(b,b)\} \cap R \neq \emptyset\};\\
R^{ck} &= R \setminus \{(a,b) \mid a \neq b, (a,a), (b,b) \in R\};\\
R^{gk} &= R \setminus \{(a,b) \mid a \neq b, (b,b) \in R, \{(b,a),(a,a)\} \cap R \neq \emptyset\};\\
R^{nk} &= R \cup \{(a,b) \mid a \neq b, \{(a,a),(b,b),(b,a)\} \cap R \neq \emptyset\}.
\end{aligned}$$

For a CAF $\mathcal{F} = (F, cl)$*, we write* $\mathcal{F}^k$ *to denote* (F^k, cl) *for* $k \in \{sk, ak, ck, gk, nk\}$*.*

We recall the characterization results of strong equivalence for AF semantics.

Theorem 7.2.5 ([142, 23])**.** *For any two AFs* F *and* G*,*

$$\begin{aligned}
F &\equiv_s^{\sigma} G \textit{ iff } F^{sk} = G^{sk} \textit{ for } \sigma \in \{stb, stg\},\\
F &\equiv_s^{\sigma} G \textit{ iff } F^{ak} = G^{ak} \textit{ for } \sigma \in \{ad, pr, ss\}\\
F &\equiv_s^{co} G \textit{ iff } F^{ck} = G^{ck}\\
F &\equiv_s^{gr} G \textit{ iff } F^{gk} = G^{gk}\\
F &\equiv_s^{\sigma} G \textit{ iff } F^{nk} = G^{nk} \textit{ for } \sigma \in \{cf, na\}
\end{aligned}$$

For an AF F, we write $F^{k(\sigma)}$ to denote the kernel which characterizes strong equivalence for the semantics σ.

In the following subsections, we characterize strong equivalence for all considered CAF semantics by identifying appropriate kernels. In brief, our findings reveal that all semantics apart from h-*cf*-stable semantics can be characterized with the kernels of their AF semantics counterpart. We identify a novel kernel for h-*cf*-stable semantics, which exhibits interesting overlaps with the stable and the naive kernel for AF semantics.

To this end we will first discuss some general observations that turn out to be useful when providing our characterization results. We will show that (i) two CAFs are strongly equivalent to each other *only if* they agree on their arguments; and (ii) strongly equivalent CAFs have the same self-attacking arguments. We will make use of the following lemma.

Lemma 7.2.6. *For a CAF* $\mathcal{F}$ *and a set of claims* $S \subseteq cl(A_{\mathcal{F}})$*, it holds that* $S \subseteq S'$ *for some* $S' \in stb_i(\mathcal{F})$ *implies that there is some* $S'' \in \rho(\mathcal{F})$ *with* $S \subseteq S''$ *for all semantics* $\rho \neq gr_i$ *under consideration.*

Proof. For all except h-preferred and h-naive semantics, the statement follows directly from known relations between semantics using $stb_i(\mathcal{F}) \subseteq \rho(\mathcal{F})$. Let $\rho \in \{pr_h, na_h\}$ and consider some claim-set $S \subseteq cl(A)$ such that $S \subseteq S'$ for some $S' \in stb_i(\mathcal{F})$ ($\subseteq \tau(\mathcal{F})$ for $\tau \in \{pr_i, na_i\}$). Since h-preferred and h-naive claim-sets are precisely the $\subseteq$-maximal i-preferred resp. i-naive claim-sets, there is $T \in pr_h(\mathcal{F})$ ($T \in na_h(\mathcal{F})$) with $T \subseteq S'$. ☐

We will first show that two CAFs with different arguments are not strongly equivalent.

Lemma 7.2.7. *For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$, $A_\mathcal{F} \neq A_\mathcal{G}$ implies $\mathcal{F} \not\equiv_s^\rho \mathcal{G}$ for any considered semantics ρ.*

Proof. W.l.o.g., we may assume that there is $a \in A_\mathcal{F}$ with $a \notin A_\mathcal{G}$. Let $cl_\mathcal{F}(a) = c$. We distinguish the following cases: (a) $(a, a) \notin R_\mathcal{F}$ and (b) $(a, a) \in R_\mathcal{F}$.

- In case $(a, a) \notin R_\mathcal{F}$, we consider the following construction: For a fresh argument x and a fresh claim d, let $\mathcal{H} = (A_\mathcal{H}, R_\mathcal{H}, cl_\mathcal{H})$ with

  $$A_\mathcal{H} = (A_\mathcal{F} \cup A_\mathcal{G} \cup \{x\}) \setminus \{a\};$$
  $$R_\mathcal{H} = \{(x, b) \mid b \in (A_\mathcal{F} \cup A_\mathcal{G}) \setminus \{a\}\};$$

  and $cl_\mathcal{H}(b) = cl_\mathcal{F}(b)$ for $b \in A_\mathcal{F} \cup A_\mathcal{G}$ and $cl_\mathcal{H}(x) = d$; that is, we introduce a new argument having a fresh claim d which attacks every argument except a. Observe that $\{c, d\} \in gr_i(\mathcal{F} \cup \mathcal{H})$ and $\{c, d\} \in stb_i(\mathcal{F} \cup \mathcal{H})$ since $\{a, x\}$ is conflict-free, and x is unattacked and attacks all remaining arguments except a in $F \cup H$; thus there is $S \in \rho(\mathcal{F} \cup \mathcal{H})$ with $\{c, d\} \subseteq S$ for every semantics ρ under consideration by Lemma 7.2.6. On the other hand, $\{c, d\} \notin cf(\mathcal{G} \cup \mathcal{H})$ since x attacks every occurrence of $cl_\mathcal{H}(a)$ in $\mathcal{G}$; therefore, $\{c, d\} \notin \rho(\mathcal{G} \cup \mathcal{H})$.

- Now, let $(a, a) \in R_\mathcal{F}$. We construct our counter-example as follows: For a fresh argument x and a fresh claim d, let $\mathcal{H} = (A_\mathcal{H}, R_\mathcal{H}, cl_\mathcal{H})$ with

  $$A_\mathcal{H} = A_\mathcal{F} \cup A_\mathcal{G} \cup \{x\};$$
  $$R_\mathcal{H} = \{(x, b) \mid b \in (A_\mathcal{F} \cup A_\mathcal{G}) \setminus \{a\}\};$$

  and $cl_\mathcal{H}(b) = cl_\mathcal{F}(b)$ for $b \in A_\mathcal{F}$, $cl_\mathcal{H}(b) = cl_\mathcal{G}(b)$ for $b \in A_\mathcal{G}$; and $cl_\mathcal{H}(x) = d$; i.e., argument x attacks every argument in $A_\mathcal{F} \cup A_\mathcal{G}$ except a. Observe that a is unattacked in $\mathcal{G} \cup \mathcal{H}$ since a is a newly introduced argument in $\mathcal{G} \cup \mathcal{H}_1$ by assumption $a \notin A_\mathcal{G}$. Therefore $\{c, d\} \in gr_i(\mathcal{G} \cup \mathcal{H})$ since $\{a, x\}$ is conflict-free and unattacked; moreover, $\{c, d\} \in stb_i(\mathcal{G} \cup \mathcal{H})$ since $\{a, x\}$ is conflict-free and attacks all remaining arguments in $G \cup H$. By Lemma 7.2.6, $\{c, d\}$ is thus contained in some ρ-claim-set for every semantics ρ under consideration. On the other hand, $\{c, d\} \notin cf(\mathcal{F} \cup \mathcal{H})$ since every realisation of $\{c, d\}$ is conflicting: a is self-attacking and x attacks every other occurrence of c. Thus $\{c, d\} \notin \rho(\mathcal{F} \cup \mathcal{H}_\mathcal{H})$ for each considered semantics ρ.

In both cases, we found a witness $\mathcal{H}$ showing that $\rho(\mathcal{F} \cup \mathcal{H}) \neq \rho(\mathcal{G} \cup \mathcal{H})$. ☐

Next we show that two strongly equivalent CAFs $\mathcal{F}$ and $\mathcal{G}$ possess the same self-attackers.

Lemma 7.2.8. *For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$, $(a,a) \in R_{\mathcal{F}} \Delta R_{\mathcal{G}}$ implies $\mathcal{F} \not\equiv_s^\rho \mathcal{G}$ for any semantics ρ under consideration.*

Proof. By Lemma 7.2.7, we may assume that $A_{\mathcal{F}} = A_{\mathcal{G}} (= A)$, i.e., a is contained in both CAFs $\mathcal{F}$ and $\mathcal{G}$. W.l.o.g., let $(a,a) \in R_{\mathcal{F}}$ and $(a,a) \notin R_{\mathcal{G}}$. Let $cl_{\mathcal{F}}(a) = cl_C G(a) = c$. Now, for a fresh argument x and fresh claim d, consider the CAF $\mathcal{H} = (A, R_{\mathcal{H}}, cl_{\mathcal{H}})$ with

$$R_{\mathcal{H}} = \{(x,b) \mid b \in A \setminus \{a\}\}$$

and $cl_{\mathcal{H}}(b) = cl_{\mathcal{F}}(b)$ for $b \in A$ and $cl_{\mathcal{H}}(x) = d$. Then $\{c,d\}$ has no cf-realisation in $\mathcal{F} \cup \mathcal{H}$ since a is self-attacking and x attacks every remaining occurrence of c in $\mathcal{F} \cup \mathcal{H}$. On the other hand, $\{c,d\} \in gr_i(\mathcal{G} \cup \mathcal{H})$ and $\{c,d\} \in stb_i(\mathcal{G} \cup \mathcal{H})$ since $\{a,x\}$ is conflict-free and attacks every other argument, moreover, x is unattacked. By Lemma 7.2.6, for all semantics ρ, there is $S \in \rho(\mathcal{G} \cup \mathcal{H})$ which contains $\{c,d\}$. Thus $\mathcal{F} \not\equiv_s^\rho \mathcal{G}$. ☐

Remark 7.2.9. *Let us remark that we do not discuss strong equivalence for for well-formed CAFs separately since our general results also apply for the special case when restricting the problem to well-formed CAFs $\mathcal{F}$ and $\mathcal{G}$.*

7.2.1 Hybrid-*cf*-stable Semantics

Let us start with h-*cf*-stable semantics. First, we observe that outgoing attacks from self-attacking arguments can be removed (apart from the self-attack itself) since such an argument cannot be part of a *cf*-stb_h-realization E, and moreover, it is not necessary that E defends itself against such attacks.

While the removal of outgoing attacks from self-attacking arguments has been already observed in the context of Dung AFs as integal part of many kernels (and defines the stable kernel, cf. Definition 7.2.4), we observe a specific behavior regarding arguments with the same claims: Coming back to our CAFs $\mathcal{F}'$ and $\mathcal{G}'$ from Example 7.2.1, we recall that they yield the same h-*cf*-stable claim-sets even after the argument d_1 has been added. The reason is that the direction of the attack between the arguments a_1 and a_2 is irrelevant since both arguments possess the same claim a. Thus it suffices to include one of them in a h-*cf*-stable claim-set in case not both of them are attacked.

Inspired by these observations, we introduce the *cf*-stable kernel for CAFs where we

- *remove* all outgoing attacks $(a,b) \neq (a,a)$ from each self-attacking argument a, and
- *add* attacks between arguments a, b, $a \neq b$, if they both carry the same claim.

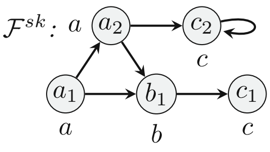
Definition 7.2.10. *For a CAF $\mathcal{F} = (A, R, cl)$, we define the* cf-stable kernel *$\mathcal{F}^{csk} = (A, R^{csk}, cl)$ with*

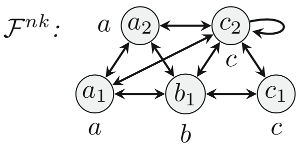
$$R^{csk} = R \cup \{(a,b) \mid a \neq b, (a,a) \in R \vee (cl(a) = cl(b) \wedge \{(b,a),(b,b)\} \cap R \neq \emptyset)\}.$$

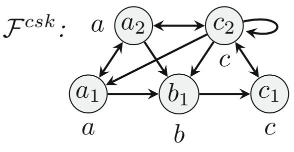
We denote the underlying AF (A, R^{csk}) by F^{csk}.

Remark 7.2.11. *The cf-stable kernel is a combination of the stable and naive kernel for AFs, where the claim-independent part stems from the stable kernel and the case where two arguments have the same claim relates to the naive kernel. In a nutshell, it is safe to introduce attacks* (a, b), $a \neq b$ *where* a *is self-attacking without changing stable semantics because attacks of this form neither interfere with the conflict-free extensions of an AF nor change the range of a conflict-free set. In case two arguments have the same claim, it is irrelevant which of these arguments is included in an extension. It is thus safe to introduce attacks between two arguments in case their union is conflicting.*

Example 7.2.12. *Consider again our previous CAF* $\mathcal{F}$*. Below we depict the stable kernel* $\mathcal{F}^{sk}$*, the naive kernel* $\mathcal{F}^{nk}$*, and the cf-stable kernel* $\mathcal{F}^{csk}$ *of* $\mathcal{F}$ *as follows:*







In what follows, we will prove that the *cf*-stable-kernel characterizes strong equivalence for hybrid *cf*-stable and stage semantics. For this, we will first show that (i) a CAF admits the same h-*cf*-stable (h-stage) claim-sets as its *cf*-stable kernel and (ii) syntactic equivalence of the kernels implies that the kernels coincide under any possible expansion.

Lemma 7.2.13. *For any CAF* $\mathcal{F}$, $\rho(\mathcal{F}) = \rho(\mathcal{F}^{csk})$ *for the semantics* $\rho \in \{cf\text{-}stb_h, stg_h\}$.

Proof. We show (a) $cf(F) = cf(F^{csk})$ and (b) for all $E \in cf(F)$, $E^*_{\mathcal{F}} = E^*_{\mathcal{F}^{csk}}$.

To show (a), first observe that $cf(F^{csk}) \subseteq cf(F)$ since no new attacks between two unconflicting arguments are introduced. Moreover, we remove only attacks (a, b) where either a or b is self-attacking, thus we obtain $cf(F) \subseteq cf(F^{csk})$.

To show (b), let $E \in cf(F)$. It holds that $E^*_{\mathcal{F}} \subseteq E^*_{\mathcal{F}^{csk}}$. Now, let $c \in E^*_{\mathcal{F}^{csk}}$ and assume $c \notin E^*_{\mathcal{F}}$, i.e., there is $b \in A$ with $cl(b) = c$ which is not attacked by E in $\mathcal{F}$ but there is $a \in E$ such that $(a, b) \in R^{csk}$. Hence either $(a, a) \in R$ or $cl(a) = cl(b)$ and $(b, a) \in R$ or $(b, b) \in R$, contradiction to E being conflict-free in F^{csk}. □

Next, we show that syntactic equivalence of *cf*-stable kernels of two CAFs $\mathcal{F}$ and $\mathcal{G}$ implies that the kernels of $\mathcal{F} \cup \mathcal{H}$ and $\mathcal{G} \cup \mathcal{H}$ coincide for any compatible $\mathcal{H}$. This means that attack removal can performed iteratively.

Lemma 7.2.14. *For any two compatible CAFs* $\mathcal{F}$ *and* $\mathcal{G}$, $\mathcal{F}^{csk} = \mathcal{G}^{csk}$ *implies* $(\mathcal{F} \cup \mathcal{H})^{csk} = (\mathcal{G} \cup \mathcal{H})^{csk}$ *for any CAF* $\mathcal{H}$ *compatible with* $\mathcal{F}$ *and* $\mathcal{G}$.

Proof. First observe that (i) $\mathcal{F} \cup \mathcal{H} \subseteq \mathcal{F}^{csk} \cup \mathcal{H}^{csk} \subseteq (\mathcal{F} \cup \mathcal{H})^{csk}$ holds for every two CAFs $\mathcal{F}$ and $\mathcal{H}$. Moreover, (ii) $\mathcal{F}^{csk} = \mathcal{G}^{csk}$ implies that $\mathcal{F}$, $\mathcal{G}$ contain the same self-attacks by definition of the *cf*-stable kernel.

Now, suppose $\mathcal{F}^{csk} = \mathcal{G}^{csk}$ and let $(a,b) \in (\mathcal{F} \cup \mathcal{H})^{csk}$. We show that $(a,b) \in (\mathcal{G} \cup \mathcal{H})^{csk}$ (the other direction is analogous): In case $(a,b) \in \mathcal{F} \cup \mathcal{H}$, we have $(a,b) \in \mathcal{F}^{csk} \cup \mathcal{H}^{csk}$ by (i). Since $\mathcal{F}^{csk} \cup \mathcal{H}^{csk} = \mathcal{G}^{csk} \cup \mathcal{H}^{csk}$ we conclude $(a,b) \in (\mathcal{G} \cup \mathcal{H})^{csk}$. In case $(a,b) \notin \mathcal{F} \cup \mathcal{H}$, either $(a,a) \in \mathcal{F} \cup \mathcal{H}$ or $cl(a) = cl(b)$ and $\{(b,b),(b,a)\} \cap (\mathcal{F} \cup \mathcal{H}) \neq \emptyset$. In case $(a,a) \in \mathcal{F} \cup \mathcal{H}$ ($(b,b) \in \mathcal{F} \cup \mathcal{H}$), we are done since $(a,a) \in \mathcal{G} \cup \mathcal{H}$ ($(b,b) \in \mathcal{G} \cup \mathcal{H}$) by (ii). Now, suppose $cl(a) = cl(b)$ and $(b,a) \in \mathcal{F} \cup \mathcal{H}$, then $(b,a) \in \mathcal{F}^{csk} \cup \mathcal{H}^{csk}$ by (i), thus also $(b,a) \in \mathcal{G}^{csk} \cup \mathcal{H}^{csk}$ by assumption $\mathcal{F}^{csk} = \mathcal{G}^{csk}$. In case $(b,a) \in \mathcal{G} \cup \mathcal{H}$, we get $(a,b) \in (\mathcal{G} \cup \mathcal{H})^{csk}$; else we have $cl(a) = cl(b)$ and $\{(a,a),(b,b),(a,b)\} \cap (\mathcal{G} \cup \mathcal{H}) \neq \emptyset$. By definition of the cf-stable kernel we obtain $(a,b) \in (\mathcal{G} \cup \mathcal{H})^{csk}$. ☐

We are now ready to prove our first main result stating that two CAFs $\mathcal{F}$ and $\mathcal{G}$ are strongly equivalent to each other w.r.t. h-cf-stable and h-stage semantics if and only if their h-stable kernels coincide. Let us sketch the idea.

First note that we obtain the 'if'-direction from Lemma 7.2.13 and 7.2.14: indeed, in case $\mathcal{F}^{csk} = \mathcal{G}^{csk}$ holds for two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$, it holds that $(\mathcal{F} \cup \mathcal{H})^{csk} = (\mathcal{G} \cup \mathcal{H})^{csk}$ for any compatible CAF $\mathcal{H}$ by Lemma 7.2.14. From Lemma 7.2.13, we infer $\rho(\mathcal{F} \cup \mathcal{H}) = \rho((\mathcal{F} \cup \mathcal{H})^{csk})$ as well as $\rho((\mathcal{G} \cup \mathcal{H})^{csk}) = \rho(\mathcal{G} \cup \mathcal{H})$, hence we obtain $\rho(\mathcal{F} \cup \mathcal{H}) = \rho(\mathcal{G} \cup \mathcal{H})$.

For the other direction, we will assume that the kernels disagree. By Lemma 7.2.7 and 7.2.8, it holds that $\mathcal{F}$ and $\mathcal{G}$ have the same arguments and in particular the same self-attackers. It thus remains to provide counter-examples for the case that the kernels of $\mathcal{F}$ and $\mathcal{G}$ disagree on an attack (a,b) for $a \neq b$. Figure 7.1 illustrates the counter-example for the case $cl(a) = cl(b)$ (case (b) in the proof of Theorem 7.2.15).

Theorem 7.2.15. *For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$,*

$$\mathcal{F}^{csk} = \mathcal{G}^{csk} \text{ iff } \mathcal{F} \equiv_s^\rho \mathcal{G} \text{ for } \rho \in \{cf\text{-}stb_h, stg_h\}.$$

Proof. We obtain $\mathcal{F}^{csk} = \mathcal{G}^{csk}$ implies $\mathcal{F} \equiv_s^\rho \mathcal{G}$ from Lemma 7.2.13 and 7.2.14 as outlined above. It remains to prove the other direction. To do so, we suppose $\mathcal{F}^{csk} \neq \mathcal{G}^{csk}$. By Lemma 7.2.13 we may assume $\rho(\mathcal{F}^{csk}) = \rho(\mathcal{G}^{csk})$; and $A_\mathcal{F} = A_\mathcal{G} (= A)$ by Lemma 7.2.7. Thus it holds that $R_{\mathcal{F}^{csk}} \neq R_{\mathcal{G}^{csk}}$. W.l.o.g., let $(a,b) \in R_{\mathcal{F}^{csk}} \setminus R_{\mathcal{G}^{csk}}$; we have $a \neq b$ by Lemma 7.2.8. Moreover, $(a,a) \notin R_\mathcal{G}^{csk}$ (and thus, $(a,a) \notin R_\mathcal{F}^{csk}$), otherwise, $(a,b) \in R_{\mathcal{G}^{csk}}$ by definition. We distinguish the following cases: (a) $cl(a) \neq cl(b)$, and (b) $cl(a) = cl(b)$.

(a) In case $cl(a) \neq cl(b)$, consider two newly introduced arguments x, y and fresh claims c, d. We consider the AF $\mathcal{H}_1 = (A \cup \{x,y\}, R_1, cl_1)$ where

$$R_1 = \{(x,y)\} \cup \{(y,h) \mid h \in A \cup \{x\}\} \cup \{(x,h) \mid h \in A \setminus \{a,b\}\},$$

and the function cl_1 is given as follows: $cl_1(x) = c$, $cl_1(y) = d$, and the other claims coincide with the given ones, i.e., $cl_1(h) = cl_\mathcal{F}(h)$ if $h \in A$. First observe that $\{d\}$ is i-stable in both $\mathcal{F}^{csk} \cup \mathcal{H}_1$ and $\mathcal{G}^{csk} \cup \mathcal{H}_1$ and thus guarantees that $\rho(\mathcal{F}^{csk} \cup \mathcal{H}_1)$ and $\rho(\mathcal{G}^{csk} \cup \mathcal{H}_1)$ are non-empty. It can be checked that $S = \{cl(a), c\}$ is h-cf-stable and

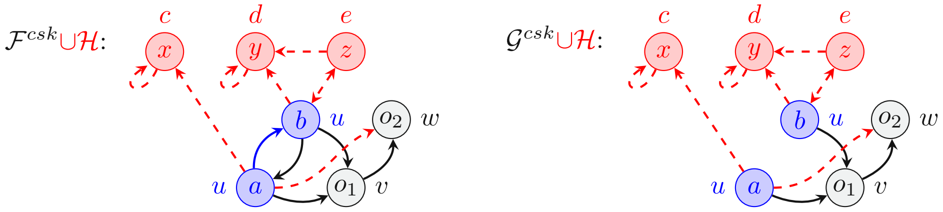


Figure 7.1: Counter-example for the case $(a, b) \in R_{\mathcal{F}^{csk}} \setminus R_{\mathcal{G}^{csk}}$ (case (b) in the proof of Theorem 7.2.15). New arguments introduced by $\mathcal{H}$ are $\{x, y, z\}$ (in red), new attacks are dashed (and red). The claim-set $\{u\}$ is h-*cf*-stable (h-stage) in $\mathcal{G}^{csk} \cup \mathcal{H}$ (since the set $\{a, b\}$ is stable in the underlying AF) but not in $\mathcal{F}^{csk} \cup \mathcal{H}$.

h-stage in $\mathcal{F}^{csk} \cup \mathcal{H}_1$ (since $\{a, x\}$ is stable); on the other hand, $S \notin \rho(\mathcal{G}^{csk} \cup \mathcal{H}_1)$ since b is not defeated by $\{a, x\}$. However, this is our only candidate since S has no other *cf*-realization in $\mathcal{G}^{csk} \cup \mathcal{H}_1$.

(b) Now consider the case $cl(a) = cl(b)$ and observe that $(a, a), (b, b), (b, a) \notin R^{csk}_{\mathcal{G}}$ (otherwise $(a, b) \in R_{\mathcal{G}^{csk}}$). Since $\mathcal{F}$ and $\mathcal{G}$ contain the same self-attacks, we furthermore have $(a, a), (b, b) \notin R_{\mathcal{F}^{csk}}$. Having established this situation let us construct $\mathcal{H}_2$ as follows: For fresh arguments x, y, z and fresh claims c, d, e, we consider $\mathcal{H}_2 = (A \cup \{x, y, z\}, R_2, cl_2)$ where

$$R_2 = \{(a, h) \mid h \in (A \cup \{x\}) \setminus \{a, b\}\} \cup \{(x, x), (b, y), (y, y), (z, b), (b, z), (z, y)\}$$

and as before we let $cl_2(h) = cl_{\mathcal{F}}(h)$ for $h \in A$; for the fresh arguments let $cl_2(x) = c$, $cl_2(y) = d$, as well as $cl_2(z) = e$. It can be checked that each CAF admits a stable extension; thus it suffices to show that the h-*cf*-stable claim-sets disagree. First observe that we now have $\{cl_2(a)\} \in \rho(\mathcal{G}^{csk} \cup \mathcal{H}_2)$ since $\{a, b\}$ is a stable extension in $G^{csk} \cup H_2$. On the other hand, we have that $\{cl_2(a)\}$ is neither *cf*-stb_h-realizable nor stg_h-realizable in $\mathcal{F}^{csk} \cup \mathcal{H}_2$. Figure 7.1 illustrates the construction.

In every case, we have found some $\mathcal{H}$ showing $\rho(\mathcal{F}^{csk} \cup \mathcal{H}) \neq \rho(\mathcal{G}^{csk} \cup \mathcal{H})$. By Lemma 7.2.13, we get $\rho(\mathcal{F} \cup \mathcal{H}) = \rho((\mathcal{F} \cup \mathcal{H})^{csk}) = \rho(\mathcal{F}^{csk} \cup \mathcal{H}) \neq \rho(\mathcal{G}^{csk} \cup \mathcal{H}) = \rho((\mathcal{G} \cup \mathcal{H})^{csk}) = \rho(\mathcal{G} \cup \mathcal{H})$. Thus it holds that $\mathcal{F} \not\equiv^s_\rho \mathcal{G}$. □

7.2.2 Inherited Semantics

Next we discuss strong equivalence w.r.t. inherited semantics. We show that inherited semantics can be characterized via known AF kernels. We prove the following theorem.

Theorem 7.2.16. *For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$,*

$$\mathcal{F} \equiv^{\sigma_i}_s \mathcal{G} \textit{ iff } F \equiv^{\sigma}_s G \textit{ iff } F^{k(\sigma)} = G^{k(\sigma)} \textit{ for each AF semantics } \sigma.$$

Recall that $F \equiv^{\sigma}_s G$ iff $F^{k(\sigma)} = G^{k(\sigma)}$ holds by known results for AF semantics [142, 23]. We moreover obtain that syntactic coincidence of the kernels implies strong equivalence of the CAFs: it holds that $F \equiv_s^\sigma G$ implies $\mathcal{F} \equiv_s^{\sigma_i} \mathcal{G}$ since $\sigma(F \cup H) = \sigma(G \cup H)$ implies $\sigma_i(\mathcal{F} \cup \mathcal{H}) = \sigma_i(\mathcal{G} \cup \mathcal{H})$ for any CAF $\mathcal{H}$ which is compatible with $\mathcal{F}$ and $\mathcal{G}$.

It remains to prove the other direction. To do so, we assume $F^{k(\sigma)} \neq G^{k(\sigma)}$ for the respective kernel which characterizes semantics σ. Moreover, we assume $A_\mathcal{F} = A_\mathcal{G} (= A)$ by Lemma 7.2.7. Therefore, there must be some attack $(a, b) \in R_\mathcal{F}^{k(\sigma)} \Delta R_\mathcal{G}^{k(\sigma)}$. W.l.o.g., let $(a, b) \in R_\mathcal{F}^{k(\sigma)}$. By Lemma 7.2.8, we furthermore may assume that $a \neq b$.

We discuss each kernel separately in the following propositions.

Proposition 7.2.17. *Given two CAFs $\mathcal{F}$ and $\mathcal{G}$ satisfying $(a, a) \in R_\mathcal{F}$ iff $(a, a) \in R_\mathcal{G}$ and $A_\mathcal{F} = A_\mathcal{G} (= A)$. Then $(a, b) \in R_\mathcal{F}^{sk} \setminus R_\mathcal{G}^{sk}$ implies $\mathcal{F} \not\equiv_s^{\sigma_i} \mathcal{G}$ for $\sigma \in \{stb, stg\}$.*

Proof. Since $(a, b) \in R_\mathcal{F}^{sk}$, we conclude that a is not self-attacking in F (which implies $(a, a) \notin R_\mathcal{G}$ by Lemma 7.2.8). We construct our counter-example as follows: for fresh arguments x, y, z and fresh claims c, d, e, let $\mathcal{H} = (A \cup \{x, y, z\}, R, cl)$ with

$$R = \{(b, z)\} \cup \{(x, h) \mid h \in (A \cup \{y\}) \setminus \{a, b\}\} \cup \{(y, h) \mid h \in A \cup \{x, z\}\}$$

and $cl(h) = cl_\mathcal{F}(h)$ for $h \in A$, $cl(x) = c$, $cl(y) = d$, and $cl(z) = e$. First observe that $\{y\}$ is stable in both $stb(F^{sk} \cup H)$ and $stb(G^{sk} \cup H)$, thus $stb(F^{sk} \cup H) = stg(F^{sk} \cup H)$ and $stb(G^{sk} \cup H) = stg(G^{sk} \cup H)$. Moreover, $\{a, x, z\} \in stb(F^{sk} \cup H)$ since x attacks each remaining argument; thus $\{cl(a), c, e\} \in stb_i(\mathcal{F}^{sk} \cup \mathcal{H})$. On the other hand, $\{cl(a), c, e\}$ has no stb_i-realisation in $\mathcal{G}^{sk} \cup \mathcal{H}$ since $\{a, x, z\}$ does not attack b; every other realisation of $\{cl(a), c, e\}$ in $\mathcal{G}^{sk} \cup \mathcal{H}$ is conflicting. □

Proposition 7.2.18. *Given two CAFs $\mathcal{F}$ and $\mathcal{G}$ satisfying $(a, a) \in R_\mathcal{F}$ iff $(a, a) \in R_\mathcal{G}$ and $A_\mathcal{F} = A_\mathcal{G} (= A)$. Then $(a, b) \in R_\mathcal{F}^{ak} \setminus R_\mathcal{G}^{ak}$ implies $\mathcal{F} \not\equiv_s^{\sigma_i} \mathcal{G}$ for $\sigma \in \{ad, pr, ss\}$.*

Proof. Since $(a, b) \in R_\mathcal{F}^{ak}$, it holds that either (a) $(a, a) \notin R_\mathcal{F}^{ak}$; or (b) $(a, a) \in R_\mathcal{F}^{ak}$ and $\{(b, a), (b, b)\} \notin R_\mathcal{F}^{ak}$.

(a) In case $(a, a) \notin R_\mathcal{F}$, consider construction $\mathcal{H}$ from the proof of Proposition 7.2.17. Then $\{cl(a), c, e\} \in \sigma_i(\mathcal{F}^{ak} \cup \mathcal{H})$ since $\{cl(a), c, e\} \in stb_i(\mathcal{F}^{ak} \cup \mathcal{H})$; on the other hand, $\{cl(a), c, e\}$ has no ad-realisation in $\mathcal{G}^{ak} \cup \mathcal{H}_1$ since z is not defended against b; every other realisation of $\{cl(a), c, e\}$ in $\mathcal{G}^{ak} \cup \mathcal{H}_1$ is conflicting since z is attacked by b and x attacks every remaining argument.

(b) For a fresh argument x and a fresh claim c, let

$$\mathcal{H}_2 = (A \cup \{x\}, \{(x, h) \mid h \in A \setminus \{a, b\}\}, cl_2)$$

with $cl_2(h) = cl_\mathcal{F}(h)$ for $h \in A$ and $cl_2(x) = c$. Then $\{b, x\} \in ad(G^{ak} \cup H_2)$ since b is not attacked by a in G^{ak} and defended against any other potential attack by x; moreover, $\{b, x\}$ semi-stable in $G^{ak} \cup H_2$ since there is no other set $D \subseteq A \cup \{x\}$ with $x \in D^{\oplus}_{\mathcal{G}^{ak} \cup \mathcal{H}_2}$ (besides $\{x\}$ which is a proper subset of $\{b, x\}$). Thus $\{cl_2(b), c\} \in \sigma_i(\mathcal{G}^{ak} \cup \mathcal{H}_1)$. On the other hand, $\{b, x\} \notin ad(F^{ak} \cup H_2)$ since b is not defended against a in $F^{ak} \cup H_2$. Thus $\{cl_2(b), c\} \notin \sigma_i(\mathcal{F}^{ak} \cup \mathcal{H}_2)$. □

Proposition 7.2.19. *Given two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *satisfying* $(a,a) \in R_\mathcal{F}$ *iff* $(a,a) \in R_\mathcal{G}$ *and* $A_\mathcal{F} = A_\mathcal{G}(= A)$*. Then* $(a,b) \in R^{ck}_\mathcal{F} \setminus R^{ck}_\mathcal{G}$ *implies* $\mathcal{F} \not\equiv^{co_i}_s \mathcal{G}$*.*

Proof. We have either $(a,a) \notin R^{ck}_\mathcal{F}$ or $(b,b) \notin R^{ck}_\mathcal{F}$. The case $(a,a) \notin R^{ck}_\mathcal{F}$ is analogous to the case (a) in the proof of Proposition 7.2.18. It remains to discuss the case $(b,b) \notin R^{ck}_\mathcal{F}$. For fresh arguments x, y and fresh claims c, d, let

$$\mathcal{H}_3 = (A \cup \{x,y\}, \{(y,a),(y,y)\} \cup \{(x,h) \mid h \in A \setminus \{a,b\}\}, cl_3)$$

with $cl_3(h) = cl_\mathcal{F}(h)$ for $h \in A$, $cl_3(x) = c$, $cl_3(y) = d$. Then $\{cl_3(b), c\} \in co_i(\mathcal{G}^{ck} \cup \mathcal{H}_3)$ since $\{b,x\}$ is conflict-free and x defends b against each attack; moreover, a is not defended by $\{b,x\}$ against y. On the other hand, $\{cl_3(b), c\} \notin co_i(\mathcal{F}^{ck} \cup \mathcal{H}_3)$ since the only conflict-free sets containing x are $\{b,x\}$, which is not defended against a; $\{x\}$, which does not realise $cl_3(b)$; and $\{a,x\}$, which is not defended against y (and a has potentially a different claim than b). ☐

Proposition 7.2.20. *Given two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *satisfying* $(a,a) \in R_\mathcal{F}$ *iff* $(a,a) \in R_\mathcal{G}$ *and* $A_\mathcal{F} = A_\mathcal{G}(= A)$*. Then* $(a,b) \in R^{gk}_\mathcal{F} \setminus R^{gk}_\mathcal{G}$ *implies* $\mathcal{F} \not\equiv^{gr_i}_s \mathcal{G}$*.*

Proof. Either (a) $(b,b) \in R^{gk}_\mathcal{F}$ and $\{(b,a),(a,a)\} \notin R^{gk}_\mathcal{F}$; or (b) $(b,b) \notin R^{gk}_\mathcal{F}$. The case (b) is analogous to the case $(b,b) \notin R^{ck}_\mathcal{F}$ considered in the proof in Proposition 7.2.19. It remains to discuss case $(b,b) \in R^{gk}_\mathcal{F}$. For fresh arguments x, y and fresh claims c, d, let

$$\mathcal{H}_4 = (A \cup \{x,y\}, \{(b,y)\} \cup \{(x,h) \mid h \in A \setminus \{a,b\}\}, cl_4)$$

with $cl_4(h) = cl_\mathcal{F}(h)$ for $h \in A$, $cl_4(x) = c$, $cl_4(y) = d$. Then x is unattacked and defends a in $\mathcal{F}^{gk} \cup \mathcal{H}_4$, which in turn defends y. Thus $\{cl_4(a), c, d\} \in gr_i(\mathcal{F}^{gk} \cup \mathcal{H}_4)$. On the other hand, we have $\{cl_4(a), c, e\} \notin gr_i(\mathcal{G}^{gk} \cup \mathcal{H}_4)$ since y is not defended against b. ☐

This concludes the proof for the semantics $\sigma \in \{stb, stg, ad, pr, ss, gr, co\}$: in every case, we found a witness $\mathcal{H}$ showing $\sigma_i(\mathcal{F}^{k(\sigma)} \cup \mathcal{H}) \neq \sigma_i(\mathcal{G}^{k(\sigma)} \cup \mathcal{H})$. By Lemma 7.2.22, we get

$$\sigma_i(\mathcal{F} \cup \mathcal{H}) = \sigma_i((\mathcal{F} \cup \mathcal{H})^{k(\sigma)}) = \sigma_i(\mathcal{F}^{k(\sigma)} \cup \mathcal{H}) \neq \sigma_i(\mathcal{G}^{k(\sigma)} \cup \mathcal{H}) = \sigma_i((\mathcal{G} \cup \mathcal{H})^{k(\sigma)}) = \sigma_i(\mathcal{G} \cup \mathcal{H}).$$

Hence it follows that $\mathcal{F} \not\equiv^s_{\sigma_i} \mathcal{G}$.

It remains to discuss conflict-free and naive semantics.

Proposition 7.2.21. *Given two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$ *satisfying* $(a,a) \in R_\mathcal{F}$ *iff* $(a,a) \in R_\mathcal{G}$ *and* $A_\mathcal{F} = A_\mathcal{G}(= A)$*. Then* $F^{nk} \neq G^{nk}$ *implies* $\mathcal{F} \not\equiv^{\sigma_i}_s \mathcal{G}$ *for* $\sigma \in \{cf, na\}$*.*

Proof. For $\sigma \in \{cf, na\}$, first notice that we can assume $\sigma_i(\mathcal{F}) = \sigma_i(\mathcal{G})$ otherwise let $\mathcal{H} = (\emptyset, \emptyset, \emptyset)$; furthermore, we can assume $\sigma(F) \neq \sigma(G)$; otherwise consider instead $\mathcal{F} \cup \mathcal{H}$ and $\mathcal{G} \cup \mathcal{H}$ for a compatible CAF $\mathcal{H}$ with $\sigma_i(\mathcal{F} \cup \mathcal{H}) \neq \sigma_i(\mathcal{G} \cup \mathcal{H})$.

First consider the case that there is some $E \in \sigma(F)\Delta\sigma(G)$ such that E is not conflict-free in F (or G, respectively). W.l.o.g., let $E \in \sigma(F)$ such that E is subset-minimal among $\sigma(F)\Delta\sigma(G)$, i.e., there is no $E' \in \sigma(F)\Delta\sigma(G)$ with $E' \subset E$; otherwise, exchange the roles of F and G. For a fresh argument x and a fresh claim c, let $\mathcal{H}_5 = (A \cup \{x\}, \{(x, b) \mid b \in A \setminus E, cl_5)$ with $cl_5(b) = cl_{\mathcal{F}}(b)$ for $b \in A$ and $cl_5(x) = c$. Then $cl_5(E) \cup \{c\} \in na(\mathcal{F} \cup \mathcal{H}_5)$ but $\{cl_5(E) \cup \{c\}$ has no cf-realisation in $\mathcal{G} \cup \mathcal{H}_5$ since every subset of E is conflicting and x attacks all remaining arguments, thus $cl_5(E) \cup \{c\} \notin \sigma_i(\mathcal{G} \cup \mathcal{H}_5)$. Observe that this suffices to conclude the proof for conflict-free semantics.

For naive semantics, assume that for all $E \in \sigma(F)\Delta\sigma(G)$, $E \in cf(F) \cap cf(G)$. We derive a contradiction: W.l.o.g., let $E \in \sigma(F)$ such that E is subset-minimal among $\sigma(F)\Delta\sigma(G)$. Since E is conflict-free in G, there is some $E' \in na(G)$ with $E \subseteq E'$. But then $E' \in cf(G)$ and thus $E \in cf(F)$ by assumption, contradiction to E being a subset-maximal conflict-free extension in F.

We have shown $\mathcal{F} \not\equiv_s^{\sigma_i} \mathcal{G}$ for $\sigma \in \{cf, na\}$. □

7.2.3 Hybrid Semantics and AF Kernels

In this section, we discuss h-*ad*-stable, h-semi-stable, h-preferred, and h-naive semantics. We will show that strong equivalence with respect to these semantics can be characterized via AF kernels: for deciding strong equivalence for h-*ad*-stable, h-semi-stable, and h-preferred semantics it suffices to compute the admissible kernel while h-naive semantics are characterized by the naive kernel for AFs.

First, we will prove the statement for h-*ad*-stable and h-semi-stable semantics. For this, we observe that the claim-extensions of a CAF are preserved under these semantics when constructing the admissible kernel. This follows from known results for AFs [142] together with the observation that the range of every admissible set of a CAF F remains unchanged in F^{ak}.

Lemma 7.2.22. *For any CAF $\mathcal{F}$, it holds that $\rho(\mathcal{F}) = \rho(\mathcal{F}^{ak})$ for $\rho \in \{ad\text{-}stb_h, ss_h\}$.*

Next we will prove that two CAFs are strongly equivalent under h-*ad*-stable and h-semi-stable semantics iff their admissible kernels coincide.

Theorem 7.2.23. *For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$,*

$$\mathcal{F} \equiv_s^{\rho} \mathcal{G} \text{ iff } F^{ak} = G^{ak} \text{ for } \rho \in \{ad\text{-}stb_h, ss_h\}.$$

Proof. First suppose $F^{ak} = G^{ak}$ and let $\mathcal{H}$ be a CAF compatible with $\mathcal{F}$, $\mathcal{G}$. By Lemma 7.2.22, and since $F \cup H = (F \cup H)^{ak}$ by known results for AF [142, Lemma 5], we obtain $\rho(\mathcal{F} \cup \mathcal{H}) = \rho((\mathcal{F} \cup \mathcal{H})^{ak}) = \rho((\mathcal{G} \cup \mathcal{H})^{ak}) = \rho(\mathcal{G} \cup \mathcal{H})$. Therefore, $\mathcal{F} \equiv_s^{\rho} \mathcal{G}$.

Now assume $\mathcal{F}^{ak} \neq \mathcal{G}^{ak}$. We may assume $\rho(\mathcal{F}^{ak}) = \rho(\mathcal{G}^{ak})$ by Lemma 7.2.22; also, $A_{\mathcal{F}} = A_{\mathcal{G}}(= A)$ by Lemma 7.2.7 and $\mathcal{F}$ and $\mathcal{G}$ contain the same self-attacks by Lemma 7.2.8.

Thus there is $(a,b) \in R^{ak}_{\mathcal{F}} \Delta R^{ak}_{\mathcal{G}}$; w.l.o.g., let $(a,b) \in R^{ak}_{\mathcal{F}}$. We distinguish three cases: (a) $(a,a) \notin R_{\mathcal{F}^{ak}}$; (b) $(a,a) \in R_{\mathcal{F}^{ak}}$ and $cl(a) \neq cl(b)$; and (c) $(a,a) \in R_{\mathcal{F}^{ak}}$ and $cl(a) = cl(b)$.

(a) In case $(a,a) \notin R_{\mathcal{F}^{ak}}$, let $\mathcal{H}1 = (A \cup \{x,y\}, R_1, cl_1)$ with

$$R_1 = \{(b,y)\} \cup \{(x,h) \mid h \in A \setminus \{a,b\}\}$$

and $cl_1(h) = cl_{\mathcal{F}}(h)$ if $h \in A$ and $cl_1(x) = c$, $cl_1(y) = d$ for newly introduced arguments x, y and fresh claims c, d. Note that $\{a,x,y\} \in stb(F^{ak} \cup H_1)$ since a defends y against b and x attacks every remaining argument. Consequently, $\{cl_1(a), c, d\} \in stb_i(\mathcal{F}^{ak} \cup \mathcal{H}_1) \subseteq \rho(\mathcal{F}^{ak} \cup \mathcal{H}_1)$.

On the other hand, we have that $\{cl_1(a), c, d\}$ is not admissible in $\mathcal{G}^{ak} \cup \mathcal{H}_1$ since it has no ad-realisation in $G^{ak} \cup H_1$: Clearly, every candidate set must contain x, y, which are the only arguments having claims c, d. The only cf-realisation of $\{cl_1(a), c, d\}$ is $\{a, x, y\}$ since every other argument is attacked by x. Observe that y is not defended against b by $\{a,x,y\}$ in $G^{ak} \cup H_1$, thus $\{cl_1(a), c, d\} \notin \rho(\mathcal{G}^{ak} \cup \mathcal{H}_1)$.

(b) In case $(a,a) \in R_{\mathcal{F}^{ak}}$, $cl(a) \neq cl(b)$, let $\mathcal{H}_2 = (A \cup \{x\}, R_2, cl_2)$ with

$$R_2 = \{(x,h) \mid h \in A \setminus \{a,b\}\}$$

for a fresh argument x with $cl_2(h) = cl_{\mathcal{F}}(h)$ for $h \in A$ and $cl_2(x) = cl_{\mathcal{F}}(a)$. First observe that $(b,b) \notin R^{ak}_{\mathcal{F}}$ (and thus also not in $R^{ak}_{\mathcal{G}}$), otherwise $(a,b) \notin R^{ak}_{\mathcal{F}}$ by definition. Hence $E = \{b, x\}$ is admissible in $G^{ak} \cup H_2$ since a does not attack b and x attacks each remaining argument. Let $S = cl_2(E)$ and observe that $S \cup E^{*}_{\mathcal{G}^{ak} \cup \mathcal{H}_2} = S \cup cl_2(A \setminus \{a\}) = cl_2(A)$ since $cl_2(a) \in S$. Thus $S \in \rho(\mathcal{G}^{ak} \cup \mathcal{H}_2)$.

On the other hand, $S \notin ad_i(\mathcal{F}^{ak} \cup \mathcal{H}_2)$: Consider a cf-realisation D of S. In case $x \notin D$, we have that D is not defended against x in $F^{ak} \cup H_2$ since x attacks any potential realization of $cl_2(a)$ in F which is not self-attacking. Now assume $x \in D$, then also $b \in D$, since x attacks any other possible choice of $cl_2(b)$ in F. In this case we have that D is not defended against a in $G^{ak} \cup H_2$ and thus $S \notin ad_i(\mathcal{F}^{ak} \cup \mathcal{H}_2)$. It follows that $\rho(\mathcal{F}^{ak} \cup \mathcal{H}_2) \neq \rho(\mathcal{G}^{ak} \cup \mathcal{H}_2)$.

(c) Now assume $(a,a) \in R_{\mathcal{F}^{ak}}$ and $cl(a) = cl(b)$. Let $\mathcal{H}_3 = (A \cup \{x,y\}, R_3, cl_3)$ with

$$R_3 = \{(x,y),(y,x)\} \cup \{(y,h \mid h \in A \cup \{x\}\} \cup \{(x,h) \mid h \in A \setminus \{a,b\}\}$$

and $cl_3(h) = cl_{\mathcal{F}}(h)$ if $h \in A$ and $cl_3(x) = c$, $cl_3(y) = d$ for newly introduced arguments x, y and fresh claims c, d, that is, $\mathcal{H}_3$ coincides with the construction $\mathcal{H}_1$ from case (a) in the Proof of Theorem 7.2.15. The argument y guarantees that $ad\text{-}stb_h(\mathcal{F}^{ak} \cup \mathcal{H}_3) \neq \emptyset$ and $ad\text{-}stb_h(\mathcal{G}^{ak} \cup \mathcal{H}_3) \neq \emptyset$ since in both $\mathcal{F}^{ak} \cup \mathcal{H}_3$ and $\mathcal{G}^{ak} \cup \mathcal{H}_3$, the claim-set $\{d\}$ is i-stable. Moreover, we have that $\{cl_3(b), c\} \in ad\text{-}stb_h(G^{ak} \cup H_3)$ (and thus $\{cl_3(b), c\} \in ss_h(G^{ak} \cup H_3)$) since $\{b, x\}$ is conflict-free and defends itself in $G^{ak} \cup H_3$—recall that $(b,b), (a,b) \notin R^{ak}_{\mathcal{G}}$ and x attacks every remaining argument except a. Since $cl_3(a) = cl_3(b)$ it follows that $\{b, x\}$ has full claim-range. On the other hand, we have that $\{cl_3(b), c\}$ has no ad-realisation in $F^{ak} \cup H_3$: Clearly, each candidate must contain x which is the only argument having claim c. Thus $\{b, x\}$ is the only cf-realisation of $\{cl_3(b), c\}$ in $F^{ak} \cup H_3$. Observe that $\{b, x\}$ is not admissible since b is not defended against the attack from a. We obtain $\rho(\mathcal{F}^{ak} \cup \mathcal{H}_3) \neq \rho(\mathcal{G}^{ak} \cup \mathcal{H}_3)$.

In every case, we have found a witness $\mathcal{H}$ showing $\rho(\mathcal{F}^{ak} \cup \mathcal{H}) \neq \rho(\mathcal{G}^{ak} \cup \mathcal{H})$. By Lemma 7.2.22, we get $\rho(\mathcal{F} \cup \mathcal{H}) = \rho((\mathcal{F} \cup \mathcal{H})^{ak}) = \rho(\mathcal{F}^{ak} \cup \mathcal{H}) \neq \rho(\mathcal{G}^{ak} \cup \mathcal{H}) = \rho((\mathcal{G} \cup \mathcal{H})^{ak}) = \rho(\mathcal{G} \cup \mathcal{H})$. It follows that $\mathcal{F} \not\equiv^{s}_{\rho} \mathcal{G}$. □

We show that deciding strong equivalence w.r.t. h-naive and h-preferred semantics coincides with deciding strong equivalence w.r.t. their inherited counterparts.

Theorem 7.2.24. *For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$,*

$$\mathcal{F} \equiv^{\sigma_h}_{s} \mathcal{G} \text{ iff } \mathcal{F} \equiv^{\sigma_i}_{s} \mathcal{G} \text{ for } \sigma \in \{na, pr\}.$$

Proof. If $\mathcal{F} \equiv^{\sigma_i}_{s} \mathcal{G}$, then $\sigma_i(\mathcal{F} \cup \mathcal{H}) = \sigma_i(\mathcal{G} \cup \mathcal{H})$ for every compatible CAF $\mathcal{H}$. $\mathcal{F} \equiv^{\sigma_h}_{s} \mathcal{G}$ follows since $\sigma_h(\mathcal{F} \cup \mathcal{H})$ are the subset-maximal i-naive claim-sets of $\mathcal{F} \cup \mathcal{H}$ and, analogously, $\sigma_h(\mathcal{G} \cup \mathcal{H})$ are the subset-maximal i-naive claim-sets of $\mathcal{G} \cup \mathcal{H}$.

Now assume $\mathcal{F} \not\equiv^{\sigma_i}_{s} \mathcal{G}$ and let $\sigma = pr$ (the proof for $\sigma = na$ is analogous). We may assume $A_{\mathcal{F}} = A_{\mathcal{G}} (= A)$ (by Lemma 7.2.7); also, $pr_i(\mathcal{F}) \neq pr_i(\mathcal{G})$ (otherwise consider instead $\mathcal{F} \cup \mathcal{H}$ and $\mathcal{G} \cup \mathcal{H}$ for a compatible CAF $\mathcal{H}$ with $pr_i(\mathcal{F} \cup \mathcal{H}) \neq pr_i(\mathcal{G} \cup \mathcal{H})$). Hence $ad(F) \neq ad(G)$. Consider a $\subseteq$-minimal set $E \in ad(F) \Delta ad(G)$. W.l.o.g., let $E \in ad(F)$.

In case there is no $D \in ad(F) \cap ad(G)$ with $D \subset E$, we consider the following construction: For a fresh argument x and a fresh claim c, let

$$\mathcal{H}_1 = ((A \cup \{x\}, \{(x, b) \mid b \in (A \setminus E\}, cl_1)$$

with $cl_1(b) = cl_{\mathcal{F}}(b)$ for $b \in A$ and $cl_1(x) = c$. Then $E \cup \{x\} \in ad(F \cup H)$ since $E \cup \{x\}$ is conflict-free and defends itself, thus $cl(E) \cup \{c\} \in ad_i(\mathcal{F} \cup \mathcal{H}_1)$. Also observe that there is no other admissible set D with $D \not\subseteq E \cup \{x\}$ which contains x, thus $cl(E) \cup \{x\}$ is a subset-maximal i-admissible set in $\mathcal{F} \cup \mathcal{H}_1$. On the other hand, $cl(E) \cup \{x\}$ has no ad-realisation in $\mathcal{G} \cup \mathcal{H}_1$ since no subset of E is admissible in G by minimality of E and x attacks every remaining argument. Thus $cl(E) \cup \{c\} \notin pr_h(\mathcal{G} \cup \mathcal{H}_1)$.

Observe that for naive semantics, this concludes the proof since by minimality of E, we can always find a conflict-free set E such that there is no $D \in cf(F) \cap cf(G)$ with $D \subset E$.

In case of preferred semantics, we now assume that the assumption is not satisfied, i.e., there is $D \in ad(F) \cap ad(G)$ with $D \subset E$. There is some $a \in E$ such that $a \notin D$ for any $D \in ad(F) \cap ad(G)$ with $D \subset E$: Otherwise every argument $a \in E$ is contained in some admissible set $D \subset E$, and thus $\bigcup\{D \in ad(G) \cap ad(F) \mid D \subset E\} = E$, i.e., the union of all admissible sets contained in E coincides with E, which implies E is admissible in

G, contradiction to the assumption. We consider the following construction: For fresh arguments x, y and fresh claims c, d, let

$$\mathcal{H}_2 = (A \cup \{x, y\}, \{(a, y)\} \cup \{(y, b) \mid b \in E\} \cup \{(x, b) \mid b \in (A \setminus E)\}, cl_2)$$

with $cl_1(b) = cl_{\mathcal{F}}(b)$ for $b \in A$, $cl_2(y) = d$ and $cl_2(x) = c$. First observe that there is no $D \subset E$ such that $D \in ad(F \cup H_2)$ (or $D \in ad(G \cup H_2)$ by the choice of a: y attacks every argument $b \in E$ and a is the only argument which defends E against y. Similar as above, we conclude that $cl(E) \cup \{c\} \in pr_h(\mathcal{F} \cup \mathcal{H}_2)$ since E is admissible in $F \cup H_2$ and x attacks every remaining argument; on the other hand, $cl(E) \cup \{c\} \notin pr_h(\mathcal{G} \cup \mathcal{H}_2)$ since no subset D of E is admissible in G.

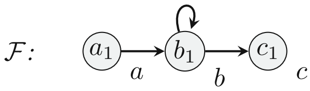
In every case, we have found a witness $\mathcal{H}$ such that $\sigma_h(\mathcal{F} \cup \mathcal{H}) \neq \sigma_h(\mathcal{G} \cup \mathcal{H})$. ☐

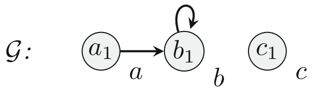
7.2.4 Strong Equivalence for Well-formed CAFs

We end this section with a brief discussion on strong equivalence for well-formed CAFs. Observe that our kernel characterizations also apply for the special case when restricting the problem to well-formed CAFs (note that we do not restrict our expansions $\mathcal{H}$).

Although the variants of stable and preferred semantics coincide for well-formed CAFs, we observe that this is in general not the case when we consider strong equivalence.

Example 7.2.25. *Consider the following two well-formed CAFs $\mathcal{F}$ and $\mathcal{G}$ depicted below:*





The set $\{a, c\}$ is stable in both CAFs; also, $\mathcal{F}^{sk} = \mathcal{G}^{sk}$ hence $\mathcal{F} \equiv_s^{stb_i} \mathcal{G}$ by Theorem 7.2.16. However, if we add a novel argument x with claim b that attacks a_1, we have $\{a, c\}$ is h-ad-stable in the expansion of $\mathcal{F}$ (witnessed by $\{x, c_1\}$) but $\{a, c\}$ is not even admissible in $\mathcal{G} \cup \{x\}$ (we already used this construction in the proof of Theorem 7.2.23).

Interestingly, we observe a close correspondence of i-stable and h-*cf*-stable semantics.

Proposition 7.2.26. *$F^{csk} = G^{csk}$ iff $F^{sk} = G^{sk}$ for every two well-formed CAFs $\mathcal{F}$, $\mathcal{G}$.*

Proof. First note that, for each well-formed CAF, the set $\{a, b\}$ with $cl(a) = cl(b)$ is conflicting iff $(a, a) \in R$ or $(b, b) \in R$. By Lemma 7.2.7 and 7.2.8, $\mathcal{F}$ and $\mathcal{G}$ have the same (self-attacking) arguments. Hence $(a, b) \in R_{\mathcal{F}}$ iff $(b, b) \in R_{\mathcal{F}}$ iff $(b, b) \in R_{\mathcal{G}}$ iff $(a, b) \in R_{\mathcal{G}}$ for all $a, b \in A_C F$ with $cl(a) = cl(b)$. Hence if $\mathcal{F}$ and $\mathcal{G}$ agree on their *cf*-stable kernels then the restriction to arguments with the same claims yields identical graphs. Since the stable and the *cf*-stable kernel both delete the same attacks between arguments not having the same claim, the statement follows. ☐

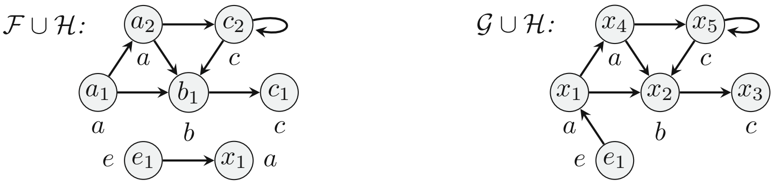
It follows that two well-formed CAFs are strongly equivalent w.r.t. inherited stable semantics iff they are strongly equivalent w.r.t. h-*cf*-stable semantics.

7.3 Renaming and Strong Equivalence

The equivalence notions we investigated so far were operating on the given arguments together with their claims. However, as we have seen in Chapter 3, the particular argument names are often not of importance. Consider the following illustrative example.

Example 7.3.1. *Assume we are given two CAFs $\mathcal{F}$ (cf. Example 7.1.2) and $\mathcal{G}$ which both stem from instantiating the same knowledge base using different argument naming schemes – the CAF $\mathcal{F}$ relates argument names with the corresponding claim (e.g., arguments with claim a are named a_i) while $\mathcal{G}$ uses a consecutive numbering for all arguments:*

It is evident that $\mathcal{F}$ and $\mathcal{G}$ are ordinary equivalent w.r.t. all considered semantics despite the mismatch in argument names because they represent the same knowledge base. However, when we consider equivalence in a dynamic setting, we observe that different argument naming patterns can cause unwanted effects. To illustrate this let us suppose we are given $\mathcal{H}$ in a way that a novel argument e_1 with claim e is given which attacks x_1:



This is fine when insisting on the specific names of the arguments. On claim-level, however, $\mathcal{H}$ disrupts the similarity between $\mathcal{F}$ and $\mathcal{G}$ in an unintended way.

The example suggests that the usual notion of strong equivalence does not handle situations where we are interested in claims only very well. Our goal is hence to develop notions of equivalence which handle such scenarios in a more intuitive way. The first step to formalize the underlying idea is the following notion of a renaming.

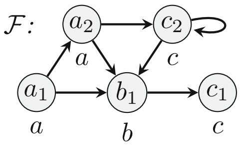
Definition 7.3.2. *For a CAF $\mathcal{F}$ and a set A' of arguments we call a bijective mapping $f : A_{\mathcal{F}} \rightarrow A'$ a* renaming for $\mathcal{F}$*. By $f(\mathcal{F})$ we denote the induced CAF (A_f, R_f, cl_f) where*

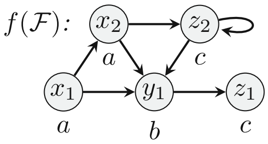
- $A_f = A'$,
- $R_f = \{(a', b') \in A' \times A' \mid (f^{-1}(a'), f^{-1}(b')) \in R_{\mathcal{F}}\}$
- $cl_f(a') = cl_{\mathcal{F}}(f^{-1}(a'))$

Since f is bijective we can reformulate the latter two conditions as follows:

- $(a, b) \in R_{\mathcal{F}}$ iff $(f(a), f(b)) \in R_f$ and
- $cl_{\mathcal{F}}(a) = cl_f(f(a))$.

Example 7.3.3. *Consider again our previous CAF $\mathcal{F}$ and let $A' = \{x_1, x_2, y_1, z_1, z_2\}$. The renaming f with $a_i \mapsto x_i$, $b_1 \mapsto y_1$ and $c_i \mapsto z_i$ induces the following CAF $f(\mathcal{F})$:*





We observe that f does not change the structure of $\mathcal{F}$ on claim-level. In particular, we observe that $\rho(\mathcal{F}) = \rho(f(\mathcal{F}))$ for all considered semantics ρ.

The last observation we made was no coincidence in the specific situation. More precisely, renaming does not change the meaning of our CAF for any considered semantics. Recall that all considered semantics satisfy argument-name independence (cf. Principle 5.1.5).

We utilize the notion of a renaming to define an appropriated notion of strong equivalence in order to handle situations like the one described in Example 7.3.1 in a satisfying way.

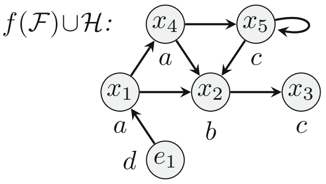
Definition 7.3.4. *Two CAFs $\mathcal{F}$ and $\mathcal{G}$ are* strongly equivalent up to renaming *w.r.t. a semantics ρ, in symbols $\mathcal{F} \equiv^{\rho}_{sr} \mathcal{G}$, iff there are renamings f and g for $\mathcal{F}$ and $\mathcal{G}$, respectively, s.t. $f(\mathcal{F}) \equiv^{\rho}_{s} g(\mathcal{G})$.*

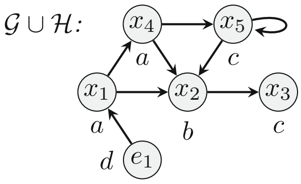
Replacing the strong equivalence requirement with its definition yields:

1. $f(\mathcal{F})$ and $g(\mathcal{G})$ are compatible with each other; and
2. $\rho(f(\mathcal{F}) \cup \mathcal{H}) = \rho(g(\mathcal{G}) \cup \mathcal{H})$ for each CAF $\mathcal{H}$ compatible with $f(\mathcal{F})$ and $g(\mathcal{G})$.

Let us reconsider our motivating Example 7.3.1.

Example 7.3.5. *Recall the CAFs $\mathcal{F}$ and $\mathcal{G}$ and consider the renamings $g = id$ and f with $f(a_1){=}x_1$, $f(b_1){=}x_2$, $f(c_1){=}x_3$, $f(a_2){=}x_4$, and $f(c_2){=}x_5$. Augmenting both $f(\mathcal{F})$ and $\mathcal{G}$ with the CAF $\mathcal{H}$, we obtain the following desired situation:*





Strong equivalence up to renaming implies the strong equivalence between two frameworks. This can be obtained by setting $f = g = id$.

Proposition 7.3.6. *For any two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$*, if* $\mathcal{F} \equiv_s^\rho \mathcal{G}$*, then* $\mathcal{F} \equiv_{sr}^\rho \mathcal{G}$ *for all semantics* ρ *satisfying argument-name independence.*

Moreover, strong equivalence survives moving to renamed versions of $\mathcal{F}$ and $\mathcal{G}$ as well.

Proposition 7.3.7. *For any two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$*, if* $\mathcal{F} \equiv_{sr}^\rho \mathcal{G}$*, then* $f(\mathcal{F}) \equiv_{sr}^\rho g(\mathcal{G})$ *for any renamings* f *and* g *for* $\mathcal{F}$ *and* $\mathcal{G}$*, respectively, for all semantics* ρ *under consideration.*

Proof. We have $\rho(g(\mathcal{F}) \cup \mathcal{H}) = \rho(\mathcal{G} \cup \mathcal{H})$ for each $\mathcal{H}$ for some renaming g because we assume $\mathcal{F} \equiv_{sr}^\rho \mathcal{G}$. Since f is a bijection we find $\rho(g(f^{-1}(f(\mathcal{F}))) \cup \mathcal{H}) = \rho(\mathcal{G} \cup \mathcal{H})$, thus $g \circ f^{-1}$ is our witnessing renaming for $f(\mathcal{F}) \equiv_{sr}^\rho \mathcal{G}$. □

Let us now come to the kernels. Since our notion of strong equivalence up to renaming allows for changing the names of the arguments, we expect our kernels to behave similarly. More specifically, we also need to consider renamed versions of the CAFs before evaluating the kernels. Hence deciding renaming strong equivalence will surely require to take the structure of the CAFs into consideration. In particular, we will see that *isomorphisms* (cf. Definition 2.2.8) play a crucial role for characterizing our novel equivalence notion.

The following proposition collects basic properties of CAF isomorphisms.

Proposition 7.3.8. *For any two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$*,*

(a) if $\mathcal{F}$ *and* $\mathcal{G}$ *are isomorphic, then* $\rho(\mathcal{F}) = \rho(\mathcal{G})$ *for any semantics* ρ *satisfying argument-name independence; and*

(b) if f *is a renaming for* $\mathcal{F}$*, then* $\mathcal{F}$ *and* $f(\mathcal{F})$ *are isomorphic.*

As it turns out, we obtain *exactly* the result we desire to: We check strong equivalence up to renaming by choosing the appropriate kernel for ρ, computing the kernels of $\mathcal{F}$ and $\mathcal{G}$ and then checking whether those are isomorphic to each other. Informally speaking, our tailored notion of equivalence which does not take the names of arguments into account yields the exact same kernels after relabeling the arguments in a suitable way.

Theorem 7.3.9. *For any two CAFs* $\mathcal{F}$ *and* $\mathcal{G}$*, for any semantics* ρ *under consideration,*

$$\mathcal{F} \equiv_{sr}^\rho \mathcal{G} \text{ iff } \mathcal{F}^{k(\rho)} \text{ and } \mathcal{G}^{k(\rho)} \text{ are isomorphic.}$$

Proof. (⇐) Let $\mathcal{F}^{k(\rho)}$ and $\mathcal{G}^{k(\rho)}$ be isomorphic, witnessed by the isomorphism f. We have $f(\mathcal{F}^{k(\rho)}) = \mathcal{G}^{k(\rho)}$; with the same mapping f we obtain $f(\mathcal{F})^{k(\rho)} = \mathcal{G}^{k(\rho)}$. By the results from Section 7.2 we are done.

(⇒) Now assume the kernels $\mathcal{F}^{k(\rho)}$ and $\mathcal{G}^{k(\rho)}$ are not isomorphic, i.e., for any two renamings f and g, $f(\mathcal{F}^{k(\rho)}) \neq g(\mathcal{G}^{k(\rho)})$. Hence we find $f(\mathcal{F})^{k(\rho)} \neq g(\mathcal{G})^{k(\rho)}$ for each such f, g. Again by the results from Section 7.2 we are done. □

Example 7.3.10. *For our CAFs $\mathcal{F}$ and $\mathcal{G}$ from Example 7.3.1 we see that their kernels are isomorphic. Hence $\mathcal{F}$ and $\mathcal{G}$ are strongly equivalent up to renaming w.r.t. all semantics considered in this paper.*

7.4 Computational Complexity

In this section we examine the computational complexity of deciding equivalence between two CAFs $\mathcal{F}$ and $\mathcal{G}$ for every equivalence notion which has been established in this paper. First, we will discuss ordinary equivalence for both general and well-formed CAFs. Our results reveal that in general, ordinary equivalence can be computationally hard, up to the third level of the polynomial hierarchy for both variants of semi-stable and stage semantics as well as for i-preferred semantics. For the remaining semantics under consideration, the problem is Π_2^{P}-complete; the only exception is i-grounded semantics for which deciding ordinary equivalence is in P. Restricting the problem to well-formed CAFs causes a drop by one level in the polynomial hierarchy for all considered semantics. The computational complexity of deciding strong equivalence, on the other hand, is tractable, as our kernel characterizations demonstrate. Moreover, we show that deciding strong equivalence up to renaming extends the list of problems which lie in NP but are not known to be NP-complete.

7.4.1 Ordinary equivalence for general CAFs

First we present our complexity results for ordinary equivalence regarding general CAFs.

Ver-OE$_\rho$
Input: Two CAFs $\mathcal{F}$, $\mathcal{G}$.
Output: TRUE iff $\mathcal{F}$, $\mathcal{G}$ are ordinary equivalent w.r.t. ρ.

The complexity of deciding ordinary equivalence is summarized as follows.

Theorem 7.4.1. Ver-OE$_\rho$ *is*

- *in* P *for* $\rho = gr_i$;
- Π_2^{P}*-complete for* $\rho \in \{cf_i, ad_i, co_i, na_i, pr_h, na_h, stb_i, cf\text{-}stb_h, ad\text{-}stb_h, \}$; *and*
- Π_3^{P}*-complete for* $\rho \in \{pr_i, ss_i, stg_i, stg_h, ss_h\}$.

Let us note that deciding Ver-OE$_{gr_i}$ is in P since computing the unique grounded extensions of F and G and comparing the claims can be done in polynomial time (cf. Table 6.2). In the following we will provide proofs for the remaining results from Theorem 7.4.1. To begin with, we present membership proofs.

Proposition 7.4.2. Ver-OE$_\rho$ *is in* Π_2^{P} *for* $\rho \in \{cf_i, ad_i, co_i, na_i, pr_h, na_h, stb_i, cf\text{-}stb_h, ad\text{-}stb_h, \}$; *and in* Π_3^{P} *for* $\rho \in \{pr_i, ss_i, stg_i, stg_h, ss_h\}$.

Proof. Membership proofs are by standard guess-and-check procedures for the complementary problems: Given two CAFs $\mathcal{F}$ and $\mathcal{G}$. First, we guess a set of claims S and check whether it holds that $S \in \mathcal{F}$ as well as $S \notin \mathcal{G}$. For the semantics $\rho \in \{cf_i, ad_i, co_i, na_i, stb_i, cf\text{-}stb_h, ad\text{-}stb_h\}$, the latter requires two NP-oracle calls; for $\rho \in \{pr_h, na_h\}$ we require four NP-oracle calls (recall that verification for h-preferred and h-naive semantics is in DP), which shows that VER-OE_ρ is in Π_2^{P}. For the semantics $\rho \in \{pr_i, ss_i, stg_i, ss_h, stg_h\}$, we require two Σ_2^{P}-oracle calls to check $S \in \mathcal{F}$ and $S \notin \mathcal{G}$; yielding Π_3^{P}-procedures for the decision problem VER-OE_ρ. □

To show hardness of VER-OE_ρ for $\rho \neq gr_i$, we present reductions from $\mathsf{QSAT}_2^\forall$ or $\mathsf{QSAT}_3^\exists$, respectively. The overall idea is to construct two CAFs $\mathcal{F}$ and $\mathcal{G}$ where $\rho(\mathcal{F})$ depends on the particular instance of the source problem while $\mathcal{G}$ serves as controlling entity. Let us outline the idea for our Π_2^{P}-hardness proofs.

For a given instance $\Psi = \forall Y \exists Z \varphi$ of $\mathsf{QSAT}_2^\forall$, we construct two CAFs $\mathcal{F}$ and $\mathcal{G}$ as follows:

- First, the claim-extensions (under some given semantics ρ) of both CAFs $\mathcal{F}$ and $\mathcal{G}$ should be of the form $Y' \cup \bar{Y}' \cup Z$ for some subset $Y' \subseteq Y$ and its complement $\bar{Y}' = \{\bar{y} \mid y \notin Y'\}$ (note that $\bar{y}$ represents $\neg y$, as usual).
- Second, we construct $\mathcal{F}$ such that the models of φ determine the claim-extensions of $\mathcal{F}$. That is, given an arbitrary subset $Y' \subseteq Y$ and its complement $\bar{Y}'$, we want that $Y' \cup \bar{Y}' \cup Z$ is a claim-extension of $\mathcal{F}$ if and only if there exists a subset $Z' \subseteq Z$ such that $Y' \cup Z'$ is a model of φ.

  Then it holds that $Y' \cup \bar{Y}' \cup Z$ is a claim-extension of $\mathcal{F}$ for all $Y' \subseteq Y$ if and only if the formula Ψ is valid.
- Finally, we construct our controlling CAF $\mathcal{G}$. This CAF is independent of the validity of Ψ. It realizes *all* claim-extensions $Y' \cup \bar{Y}' \cup Z$ for each subset $Y' \subseteq Y$ by default.

  Thus it holds that $\mathcal{F}$ and $\mathcal{G}$ yield the same claim-extensions if and only if Ψ is valid.

For our constructions, we make use of our complexity results that we have already established in this work. In particular, we utilize our results for deciding concurrence (cf. Section 6.3) for constructing the CAF $\mathcal{F}$. For our constructions in the concurrence section, we have shown that $\sigma_h(\mathcal{F}) = \sigma_i(\mathcal{F})$ iff the considered instance Ψ of $\mathsf{QSAT}_2^\forall$ (or $\mathsf{QSAT}_3^\exists$) is valid. It remains to construct the CAF $\mathcal{G}$ in such a way such that $\sigma_h(\mathcal{F}) = \sigma_i(\mathcal{G})$.

Let us demonstrate this procedure for inherited stable semantics.

Proposition 7.4.3. *Deciding* VER-OE_{stb_i} *is* Π_2^{P}*-hard.*

Proof. Consider an instance $\Psi = \forall Y \exists Z \varphi(Y, Z)$ of $\mathsf{QSAT}_2^\forall$ where φ is given by a set of clauses C over atoms in $V = Y \cup Z$. We may assume that Z is not empty (otherwise,

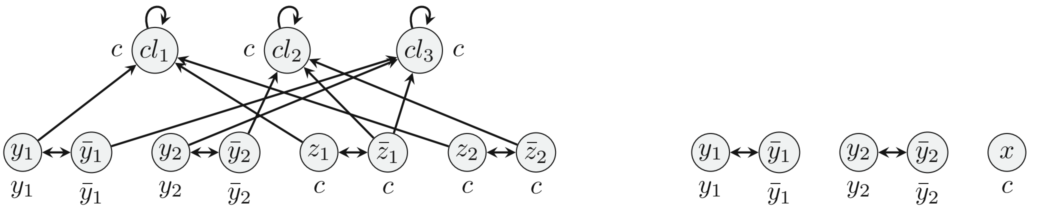


Figure 7.2: CAFs $\mathcal{F}$ (left) and $\mathcal{G}$ (right) illustrating the reduction from the Proof of Proposition 7.4.3 for the formula $\Psi = \forall Y \exists Z \varphi(Y,Z)$ where $\varphi(Y,Z)$ is given by the clauses $\{\{y_1, z_1, z_2\}, \{\bar{z}_1, \bar{z}_2, \bar{y}_2)\}, \{\bar{y}_1, \bar{z}_1, y_2\}\}$.

extend φ with a clause containing a single atom z). We construct two CAFs $\mathcal{F}$ and $\mathcal{G}$. For $\mathcal{F}$, we consider the CAF from Reduction 6.3.6. The CAF $\mathcal{G}$ is given by $A_\mathcal{G} = Y \cup \bar{Y} \cup \{c\}$, $R_\mathcal{G} = \{(y,\bar{y}), (\bar{y},y) \mid y \in Y\}$, and $cl_\mathcal{G} = id$. We observe that $stb(G) = stb_i(\mathcal{G}) = \{Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \mid Y' \subseteq Y\}$. An example of both CAFs is given in Figure 7.2.

As shown in the proof of Proposition 6.3.7, Ψ is valid iff $stb_i(\mathcal{F}) = \tau\text{-}stb_h(\mathcal{F})$ for $\tau \in \{ad, cf\}$, and $\tau\text{-}stb_h(\mathcal{F}) = \{Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{c\} \mid Y' \subseteq Y\}$, i.e., $\tau\text{-}stb_h(\mathcal{F}) = stb_i(\mathcal{G})$. Thus we obtain Ψ is valid iff $stb_i(\mathcal{F}) = stb_i(\mathcal{G})$. □

By modifying the constructions from the proof of Proposition 7.4.3 we obtain Π_2^P-hardness of VER-OE$_{na_i}$.

Proposition 7.4.4. *Deciding* VER-OE$_{na_i}$ *is* Π_2^P*-hard.*

Proof. Consider an instance $\Psi = \forall Y \exists Z \varphi(Y,Z)$ of $\mathsf{QSAT}_2^\forall$, where φ is a 3-CNF, given by a set of clauses $C = \{cl_1, \dots, cl_n\}$ over atoms in $V = Y \cup Z$. We construct two CAFs $\mathcal{F} = (A_\mathcal{F}, R_\mathcal{F}, cl_\mathcal{F})$, $\mathcal{G} = (A_\mathcal{G}, R_\mathcal{G}, id)$, where $\mathcal{F}$ modifies the standard construction (A, R) (cf. Reduction 6.2.11) as follows:

$$\begin{aligned} A_\mathcal{F} &= (A \setminus \{\varphi\}) \cup Y_2 \cup \bar{Y}_2 \cup Z_2; \\ R_\mathcal{F} &= (R \cap (A_\mathcal{F}^2)) \cup \{(y_2, \bar{y}_2), (y, \bar{y}_2), (y_2, \bar{y}) \mid y \in Y\}; \end{aligned}$$

and $cl_\mathcal{F}(y) = cl_\mathcal{F}(y_2) = y$, $cl_\mathcal{F}(\bar{y}) = cl_\mathcal{F}(\bar{y}_2) = \bar{y}$ for $y \in Y$, $cl_\mathcal{F}(z_2) = z$ for $z \in Z$, and $cl(a) = c$ otherwise (cf. Figure 7.3, left).

Observe that $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z \cup \{c\}$ is i-naive for every $Y' \subseteq Y$: Let $E = Y_2' \cup \{\bar{y}_2 \mid y_2 \notin Y_2'\} \cup Z_2 \cup C \cup E'$ with $Y_2' \subseteq Y_2$ and $E' \subseteq V \cup \bar{V}$ is a non-conflicting subset-maximal set of arguments which do not attack any $cl \in C$. E is conflict-free and subset-maximal by the choice of E'; moreover, $cl_\mathcal{F}(E) = Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z \cup \{c\}$.

We construct $\mathcal{G} = (A_\mathcal{G}, R_\mathcal{G}, cl_\mathcal{G})$ such that $na_i(\mathcal{G})$ contains all sets of the form $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z \cup \{c\}$ and $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z$ for each $Y' \subseteq Y$. Let

$$\begin{aligned} A_\mathcal{G} &= Y_1 \cup \bar{Y}_1 \cup Y_2 \cup \bar{Y}_2 \cup Z \cup \{x\}; \\ R_\mathcal{G} &= \{(y_i, \bar{y}_i) \mid y_i \in Y_i, i \leq 2\} \cup \{(a,b) \mid a \in Y_1 \cup \bar{Y}_1, b \in Y_2 \cup \bar{Y}_2 \cup \{x\}\}; \end{aligned}$$

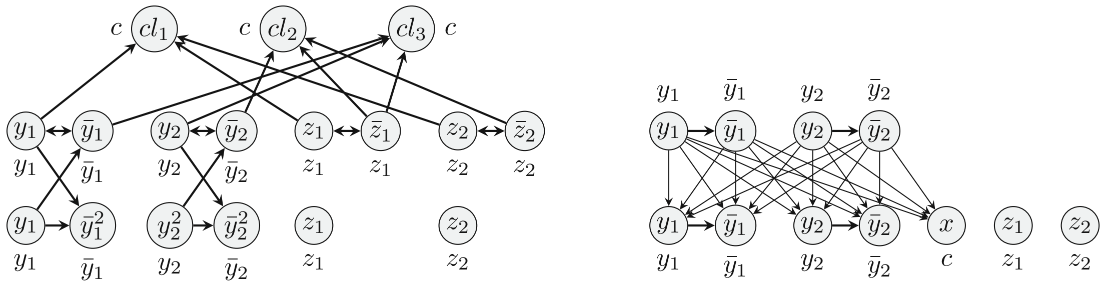


Figure 7.3: CAFs $\mathcal{F}$ (left) and $\mathcal{G}$ (right) illustrating the reduction from the Proof of Proposition 7.4.4 for the formula $\Psi = \forall Y \exists Z \varphi(Y, Z)$ where $\varphi(Y, Z)$ is given by the clauses $\{\{y_1, z_1, z_2\}, \{\bar{z}_1, \bar{z}_2, \bar{y}_2)\}, \{\bar{y}_1, \bar{z}_1, y_2\}\}$.

and $cl_{\mathcal{G}}(y_i) = y$, $cl_{\mathcal{G}}(\bar{y}_i) = \bar{y}$ for $y_i \in Y_i$; $cl_{\mathcal{G}}(z) = z$ for $z \in Z$; $cl_{\mathcal{G}}(x) = c$. See Figure 7.3 for an illustrative example of $\mathcal{F}$ and $\mathcal{G}$. It can be checked that $\mathcal{G}$ has precisely the desired i-naive extensions.

We show that Ψ is valid iff $na_i(\mathcal{F}) = na_i(\mathcal{G})$. First, assume Ψ is valid and fix some $Y' \subseteq Y$. There is $Z' \subseteq Z$ such that $M = Y' \cup Z'$ is a model of φ. Let $E = M \cup \{\bar{v} \mid v \notin M\} \cup Y_2' \cup \{\bar{y}_2 \mid y_2 \notin Y_2'\} \cup Z_2$. E is conflict-free; moreover, E is subset-maximal among conflict-free sets since any other argument $a \in A_{\mathcal{F}} \setminus E$ is in conflict with E. On the one hand, E attacks every $cl \in C$ since M is a model of φ. Also, E contains either v or $\bar{v}$ for any atom $v \in Y \cup Z \cup Y_2$, thus any argument representing a literal in $\mathcal{F}$ which is not a member of E is attacked by E. It follows that $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z \in na_i(\mathcal{F})$ for every $Y' \subseteq Y$. Each i-naive claim-set is thus either of the form $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z \cup \{\bar{\varphi}\}$ or $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z$. Consequently, $na_i(\mathcal{F}) = na_i(\mathcal{G})$ in case Ψ is valid.

Now assume $na_i(\mathcal{F}) = na_i(\mathcal{G})$ and fix $Y' \subseteq Y$. Consider a na_i-realisation E of $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup Z$ and let $Z' = E \cap Z$. Then $M = Y' \cup Z'$ is a model of φ (since $E \cup \{cl\}$ is conflicting for each $cl \in C$), as desired. □

For the construction of $\mathcal{F}$ in the Π_2^{P}-hardness proof of VER-OE$_\rho$, $\rho = \{cf_i, ad_i, na_h, pr_h\}$, we choose a slightly different approach: For an instance $\Psi = \forall Y \exists Z \varphi(Y, Z)$ of $\mathsf{QSAT}_2^{\forall}$, we construct $\mathcal{F}$ such that each literal in a clause cl is represented by an argument having claim cl; we furthermore introduce arguments for each atom $y \in Y$ and its negation; finally, every two arguments representing negated literals attack each other. We construct $\mathcal{G}$ in a way such that $\rho(\mathcal{G})$ contains precisely the claim-sets $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup C$. Similar as above, it can be shown that Ψ is valid iff $\rho(\mathcal{F}) = \rho(\mathcal{G})$.

Proposition 7.4.5. *Deciding* VER-OE$_\rho$ *is* Π_2^{P}*-hard,* $\rho \in \{cf_i, ad_i na_h, pr_h\}$.

Proof. We will first show the statement for h-naive semantics: Consider an instance $\Psi = \forall Y \exists Z \varphi(Y, Z)$ of $\mathsf{QSAT}_2^{\forall}$, where φ is a 3-CNF, given by a set of clauses $C = \{cl_1, \ldots, cl_n\}$ over atoms in $V = Y \cup Z$. We construct two CAFs $\mathcal{F} = (A_{\mathcal{F}}, R_{\mathcal{F}}, cl_{\mathcal{F}})$, $\mathcal{G} = (A_{\mathcal{G}}, R_{\mathcal{G}}, id)$.

For $\mathcal{F}$, we use parts of Reduction 6.2.7. Let (A, R, cl) be given as in Reduction 6.2.7, then we let CAF $\mathcal{F}$ be defined as $A_\mathcal{F} = A \setminus \{a_1, \dots, a_n, \varphi\}$, and $R_\mathcal{F} = R \cap A^2_\mathcal{F}$. Similar as in Lemma 6.2.8, it can be shown that $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n\} \in na_i(\mathcal{F})$ iff there is $Z' \subseteq Z$ such that $Z' \cup Y'$ is a model of φ.

We construct a CAF $\mathcal{G}$ having the h-naive claim-sets $Y' \cup \{\bar{y} \mid y \notin Y'\} \cup \{1, \dots, n\}$ for every $Y' \subseteq Y$ by setting $A_\mathcal{G} = Y \cup \bar{Y} \cup \{1, \dots, n\}$ and $R_\mathcal{G} = \{(y, \bar{y}), (\bar{y}, y) \mid y \in Y\}$. Thus Ψ is valid iff the h-naive claim-sets of $\mathcal{F}$ and $\mathcal{G}$ coincide.

Since conflict-free semantics satisfy downward closure (each subset of a conflict-free set is conflict-free), we have $cf_i(\mathcal{F}) = cf_i(\mathcal{G})$ iff $na_h(\mathcal{F}) = na_h(\mathcal{G})$ and thus the statement follows for i-conflict-free semantics. By symmetry of $\mathcal{F}$ and $\mathcal{G}$ we furthermore have $ad(F) = cf(F)$ and $ad(G) = cf(G)$ which implies $ad_i(\mathcal{F}) = cf_i(\mathcal{F})$, $ad_i(\mathcal{G}) = cf_i(\mathcal{G})$, $pr_h(\mathcal{F}) = na_h(\mathcal{F})$, and $pr_h(\mathcal{G}) = na_h(\mathcal{G})$. ☐

It remains to provide Π_2^P-hardness proofs for complete and h-stable semantics. For this, we make use of intertranslatability-results between semantics. In Section 6.2.1, we introduced a translation that maps i-stable semantics to h-stable semantics. Using this result we can reduce Ver-OE$_{stb_i}$ to Ver-OE$_{\tau\text{-}stb_h}$ as follows: Given two CAFs $\mathcal{F}$ and $\mathcal{G}$, we compute $Tr_2(\mathcal{F})$ and $Tr_2(\mathcal{G})$ in polynomial time and check whether $\tau\text{-}stb_h(Tr_2(\mathcal{F})) = \tau\text{-}stb_h(Tr_2(\mathcal{G}))$. From a translation presented in [91] that maps admissible to complete semantics we obtain an analogous result for complete semantics. In this way, we obtain lower bounds for the remaining decision problems which are Π_2^P-complete.

Proposition 7.4.6. Ver-OE$_{stb_i}$ $\leq_p$ Ver-OE$_{\tau\text{-}stb_h}$*,* $\tau \in \{ad, cf\}$*, and* Ver-OE$_{ad_i}$ $\leq_p$ Ver-OE$_{co_i}$*.*

Turning now to Π_3^P-hardness results, we adapt our strategy slightly. Similarly as for our Π_2^P-hardness proofs, we construct two CAFs $\mathcal{F}$ and $\mathcal{G}$ such that the claim-extensions of $\mathcal{F}$ depend on the validity of an instance $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$ of $\mathsf{QSAT}_3^\exists$ while the claim-extensions of $\mathcal{G}$ are independent of Ψ. Now, our target is to construct $\mathcal{F}$ in a way such that $\rho(\mathcal{F}) \neq \rho(\mathcal{G})$ iff Ψ is valid.

Let us start with inherited semi-stable and stage semantics. To show Π_3^P-hardness of Ver-OE$_{ss_i}$ and Ver-OE$_{stg_i}$, we will make use of Reduction 6.3.8. As shown in Section 6.2, $ss_i(\mathcal{F}) = stg_i(\mathcal{F})$ (likewise, $ss_h(\mathcal{F}) = stg_h(\mathcal{F})$) for the CAF $\mathcal{F}$ obtained from the reduction; moreover, each claim-set of the form $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y \cup Z \cup \{e\}$ for $X' \subseteq X$, $e \in \{\varphi, \bar{\varphi}\}$ is h-semi-stable (h-stage) in $\mathcal{F}$. It holds that Ψ is not valid iff the inherited and hybrid semi-stable (stage) variants coincide. Thus is suffices to construct a CAF $\mathcal{G}$ which realizes each claim-set $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y \cup Z \cup \{e\}$ for $X' \subseteq X$, $e \in \{\varphi, \bar{\varphi}\}$ under i-semi-stable (stage) semantics.

Proposition 7.4.7. *Deciding* Ver-OE$_\rho$ *is* Π_3^P*-hard,* $\rho \in \{ss_i, stg_i\}$*.*

Proof. Consider an instance $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$ of $\mathsf{QSAT}_3^\exists$, where φ is given by a set of clauses C over atoms in $V = X \cup Y \cup Z$. Let $\mathcal{F}$ be given as in Reduction 6.3.8.

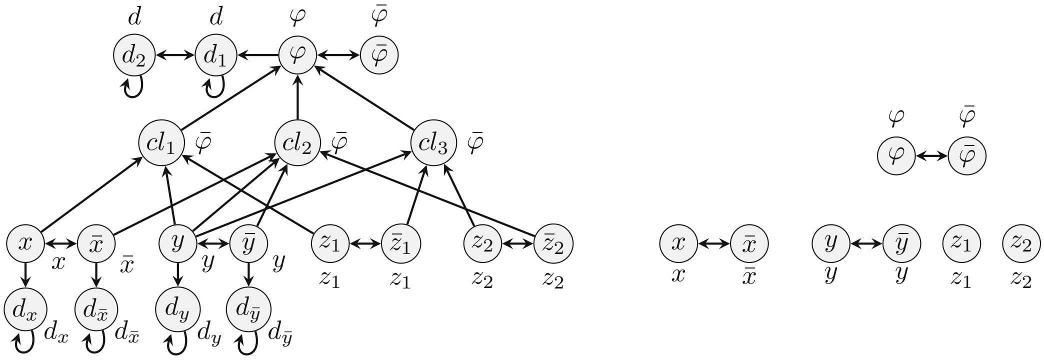


Figure 7.4: CAF $\mathcal{F}$ ((left) and CAF $\mathcal{G}$ (right) from the proof of Proposition 7.4.7 for the formula $\exists X \forall Y \exists Z \varphi(X, Y, Z)$ with clauses $\{\{z_1, x, y\}, \{\neg x, \neg y, \neg z_2, y\}, \{\neg z_1, z_2, y\}\}$.

We have shown that $ss_h(\mathcal{F}) = stg_h(\mathcal{F})$, moreover, Ψ is not valid iff $ss_i(\mathcal{F}) = ss_h(\mathcal{F}) = \{X' \cup \{\bar{x} \mid x \notin X'\} \cup Y \cup Z \cup \{e\} \mid X' \subseteq X,\ e \in \{\varphi, \bar{\varphi}\}\}$. It suffices to construct $\mathcal{G}$ in such a way that $\rho(\mathcal{G}) = ss_h(\mathcal{F})$: Then Ψ is not valid iff $ss_i(\mathcal{G}) = ss_h(\mathcal{F}) = ss_i(\mathcal{F})$. We construct such a CAF $\mathcal{G} = (A_\mathcal{G}, R_\mathcal{G}, id)$ by setting

$$A_\mathcal{G} = X \cup \bar{X} \cup Y \cup Z \cup \{\varphi, \bar{\varphi}\}, \text{ and}$$
$$R_\mathcal{G} = \{(x, \bar{x}), (\bar{x}, x) \mid x \in X\} \cup \{(\varphi, \bar{\varphi}), (\bar{\varphi}, \varphi)\}.$$

$\mathcal{G}$ possesses exactly the desired i-semi-stable claim-sets. This concludes the proof for i-semi-stable semantics. Π_3^P-hardness of VER-OE$_{stg_i}$ follows from the fact that $ss_i(\mathcal{F}) = stg_i(\mathcal{F})$ and $ss_i(\mathcal{G}) = stg_i(\mathcal{G})$. Figure 7.4 provides an illustrative example of $\mathcal{F}$ and $\mathcal{G}$. ☐

To show Π_3^P-hardness of i-preferred semantics, we adapt Reduction 6.3.18.

Proposition 7.4.8. *Deciding* VER-OE$_{pr_i}$ *is* Π_3^P*-hard.*

Proof. We show hardness via a reduction from $\mathsf{QSAT}_3^\exists$. Consider an instance $\Psi = \exists X \forall Y \exists Z \varphi(X, Y, Z)$ of $\mathsf{QSAT}_3^\exists$, where φ is given by a set of clauses C over atoms in $V = X \cup Y \cup Z$. W.l.o.g., we can assume there is $y_0 \in Y$ which is contained in each clause $cl \in C$. First, we construct a CAF $\mathcal{F} = (A_\mathcal{F}, R_\mathcal{F}, cl_\mathcal{F})$ where $(A_\mathcal{F}, R_\mathcal{F})$ is given as in Reduction 6.3.18 and $cl_\mathcal{F}(y) = cl_\mathcal{F}(\bar{y}) = y$ for $y \in Y$ and $cl_\mathcal{F}(v) = v$ otherwise.

Second, we construct a CAF $\mathcal{G} = (A_\mathcal{G}, R_\mathcal{G}, cl_\mathcal{G})$ such that $pr_i(\mathcal{G}) = \{V' \cup \{\bar{v} \mid v \notin V'\} \cup Y \cup \{\varphi\} \mid V' \subseteq X \cup Z\} \cup \{X' \cup \{\bar{x} \mid x \notin X'\} \cup Y \mid X' \subseteq X\}$ by setting

$$\begin{aligned} A_\mathcal{G} &= X_i \cup \bar{X}_i \cup Y \cup Z \cup \bar{Z} \cup \{\varphi\}; \\ R_\mathcal{G} &= \{(v_i, \bar{v}_j), (\bar{v}_i, v_j) \mid v_i, v_j \in X_1 \cup X_2\} \cup \{(v, \bar{v}), (\bar{v}, v) \mid v \in Z\} \cup \\ &\quad \{(a, b), (b, a) \mid a \in A' \cup \{\varphi\}, b \in X_2 \cup \bar{X}_2\}; \end{aligned}$$

for two copies X_i, $\bar{X}_i$, $i \leq 2$, of X and $\bar{X}$, respectively; moreover, $cl_\mathcal{G}(x_i) = x$, $cl_\mathcal{G}(\bar{x}_i) = \bar{x}$, and $cl_\mathcal{G}(a) = a$ for all remaining $a \in A_\mathcal{G}$.

First observe that $\{V' \cup \{\bar{v} \mid v \notin V'\} \cup Y \cup \{\varphi\} \mid V' \subseteq X \cup Z\} \subseteq pr_i(\mathcal{F})$ since $y_0 \in cl$ for every clause cl, that is, for every atom $v \in V \setminus \{y_0\}$, we can choose either v or $\bar{v}$ as long as y_0 is contained in $E \subseteq A_{\mathcal{F}}$, we have that E defends φ against each attack.

In case Ψ is not valid, consider some $X' \subseteq X$. Since Ψ is not valid, there is some $Y' \subseteq Y$ such that for all $Z' \subseteq Z$, some clause $cl \in C$ is not satisfied. It follows that $E = X' \cup \{\bar{x} \mid x \notin X'\} \cup Y' \cup \{\bar{y} \mid y \notin Y'\}$ is preferred in F: Clearly, E is conflict-free and defends itself. Now assume there is $a \in A \setminus E$ such that $E \cup \{a\} \in ad(F)$. In case $a = \varphi$ we have that each $cl \in C$ is attacked, that is, for every clause $cl \in C$ there is $v \in X' \cup Y'$ such that either $v \in X' \cup Y'$ with $v \in cl$ or $v \notin X' \cup Y'$ with $\neq v \in cl$. Thus $X' \cup Y'$ is a model of φ, contradiction to Ψ being not valid. Observe that the case $a \in Z \cup \bar{Z}$ requires $\varphi \in E$, otherwise a is not defended against $\bar{\varphi}$. We have thus shown that $cl(E) = X' \cup \{\bar{x} \mid x \notin X'\} \cup Y \in pr_i(\mathcal{F})$ for every $X' \subseteq X$.

We show that every i-preferred set of $\mathcal{F}$ is either of the form (a) $V' \cup \{\bar{v} \mid v \notin V'\} \cup Y \cup \{\varphi\}$ for some $V' \subseteq X \cup Z$ or (b) $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y$ for some $X' \subseteq X$. As outlined above, any such set is i-preferred in $\mathcal{F}$, thus it remains to show that there is no other i-preferred set in $\mathcal{F}$. First notice that each i-preferred claim-set of $\mathcal{F}$ contains $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y$ for some $X' \subseteq X$ since every preferred set E of F contains $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y' \cup \{\bar{y} \mid y \notin Y'\}$ for some $X' \subseteq X$, $Y' \subseteq Y$ by construction. Now assume there is $S \subseteq cl(A_{\mathcal{F}})$ such that $S \in pr_i(\mathcal{F})$ and S is not of the form (a) or (b). Let E be a pr_i-realisation of S. First assume $\varphi \notin E$. Then $z, \bar{z} \notin E$ for any $z \in Z$ since φ is the only argument which defends $z, \bar{z}$ against $\bar{\varphi}$. By the above consideration there are $X' \subseteq X$, $Y' \subseteq Y$ such that $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y' \cup \{\bar{y} \mid y \notin Y'\} \subseteq E$. Observe that $a \notin E$ for any $a \in (X \setminus X') \cup \{\bar{x} \mid x \in X'\} \cup (Y \setminus Y') \cup \{\bar{y} \mid y \in Y'\}$ since $v, \bar{v}$ are mutually attacking for any $v \in X \cup Y$. Since every remaining argument is either attacked by E or self-attacking it follows that $S = X' \cup \{\bar{x} \mid x \notin X'\} \cup Y$. In case $\varphi \in E$, we have that every $z, \bar{z}$ is defended against $\bar{\varphi}$. Thus E contains either z or $\bar{z}$ for every $z \in Z$ by subset-maximality of E. Thus there is $Z' \subseteq Z$ such that $E = V' \cup \{\bar{v} \mid v \notin V'\} \cup \{\varphi\}$. Since every remaining argument is either attacked by E or self-attacking, we have $S = V' \cup \{\bar{v} \mid v \notin V'\} \cup Y \cup \{\varphi\}$ for some $V' \subseteq X \cup Z$. It follows that $pr_i(\mathcal{F}) = pr_i(\mathcal{G})$.

Now assume $pr_i(\mathcal{F}) = pr_i(\mathcal{G})$ and consider some $X' \subseteq X$. Let E be a pr_i-realisation of $X' \cup \{\bar{x} \mid x \notin X'\} \cup Y$ and let $Y' = E \cap Y$. We show that for all $Z' \subseteq Z$, $X' \cup Y' \cup Z'$ is not a model of φ. Fix some $Z' \subseteq Z$ and let $M = X' \cup Y' \cup Z'$. Since E is preferred in $\mathcal{F}$ we have that φ is not defended by $E \cup Z' \cup \{\bar{z} \mid z \notin Z'\}$; i.e., there is some $cl \in C$ such that $E \cup Z' \cup \{\bar{z} \mid z \notin Z'\}$ does not attack cl. Consequently, for all $v \in V$, in case $v \in cl$ we have $v \notin M$, and in case $\neq v \in cl$ we have $v \in M$. Hence M is not a model of φ. □

We obtain lower bounds for the remaining semantics from translations Tr_3 and Tr_1 (cf. Section 6.2.1). In this way, we obtain lower bounds for the remaining decision problems.

Proposition 7.4.9. VER-OE$_{pr_i}$ $\leq_p$ VER-OE$_{ss_h}$ *and* VER-OE$_{stg_i}$ $\leq_p$ VER-OE$_{stg_h}$.

This concludes our complexity analysis of ordinary equivalence for general CAFs.

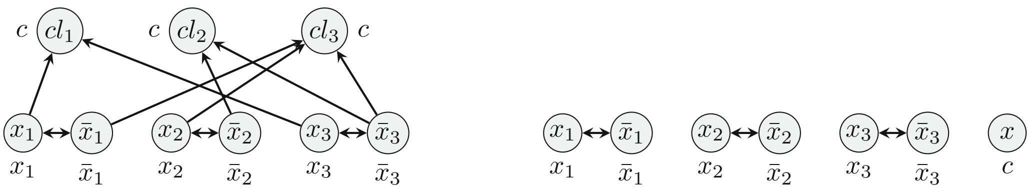


Figure 7.5: CAFs $\mathcal{F}$ (left) and $\mathcal{G}$ (right) illustrating the reduction from the Proof of Proposition 7.4.11 for the formula φ given by the clauses $\{\{x_1, x_3\}, \{\bar{x}_2, \bar{x}_3)\}, \{\bar{x}_1, \bar{x}_3, x_2\}\}$.

7.4.2 Ordinary equivalence for well-formed CAFs

In this section, we discuss ordinary equivalence for well-formed CAFs. By $\textsc{Ver-OE}_\rho^{wf}$ we denote the problem of deciding ordinary equivalence restricted to well-formed CAFs. In general, we observe that the computational complexity of deciding ordinary equivalence drops one level in the polynomial hierarchy for all considered semantics (except for grounded semantics) when considering well-formed CAFs only. Our results can be summarized as follows.

Theorem 7.4.10. $\textsc{Ver-OE}_\rho^{wf}$ *is*

- *in* P *for* $\rho = gr_i$*;*
- coNP*-complete for* $\rho \in \{cf_i, ad_i, co_i, na_i, na_h, stb_i\}$*; and*
- Π_2^{P}*-complete for* $\rho \in \{pr_i, ss_i, stg_i, stg_h, ss_h\}$*.*

Membership results are obtained in the same way as for general CAFs. We obtain lower bounds for deciding ordinary equivalence with respect to admissible, complete, stable, and preferred semantics for (well-formed) CAFs from the corresponding results for AFs [24].

For h-naive and conflict-free semantics, we utilize the standard construction once again.

Proposition 7.4.11. $\textsc{Ver-OE}_\rho^{wf}$ *is* coNP*-hard for* $\rho \in \{cf_i, na_i\}$*.*

Proof. Consider a SAT instance φ given by a set of clauses C over atoms in X. We may assume that there is no clause $cl \in C$ such that $x, \bar{x} \in cl$ for any atom $x \in X$. We construct two CAFs: For $\mathcal{F}$, construct (A, R, cl) from Reduction 6.3.6 and let $A_\mathcal{F} = A$, $R_\mathcal{F} = R \setminus \{(cl, cl) \mid cl \in C\}$, and $cl_\mathcal{F}(cl) = c$ for all $cl \in C$ and $cl(x) = x$ otherwise. The CAF $\mathcal{G}$ is given by $A_\mathcal{G} = X \cup \bar{X} \cup \{c\}$, $R_\mathcal{G} = \{(x, \bar{x}), (\bar{x}, x) \mid x \in X\}$, and $cl_\mathcal{G} = id$. Then $stb(G) = stb_i(\mathcal{G}) = \{X' \cup \bar{X}' \cup \{c\} \mid X' \subseteq X\}$. An example is given in Figure 7.5.

For $\mathcal{F}$, it holds that each h-naive claim-set contains either x or $\bar{x}$ for each literal x (recall that we excluded clauses containing both x and $\bar{x}$). Each set $X' \cup \bar{X}'$, $X' \subseteq X$, is conflict-free in F; moreover, $X' \cup \bar{X}'$ is naive in F iff all clause-arguments are attacked iff φ is satisfiable. Regarding h-stable semantics, it thus holds that φ is unsatisfiable iff $na_h(\mathcal{F}) = na_h(\mathcal{G})$. This concludes the proof for h-naive semantics. Since each subset of a naive extension is conflict-free, the statement for conflict-free semantics follows. $\square$

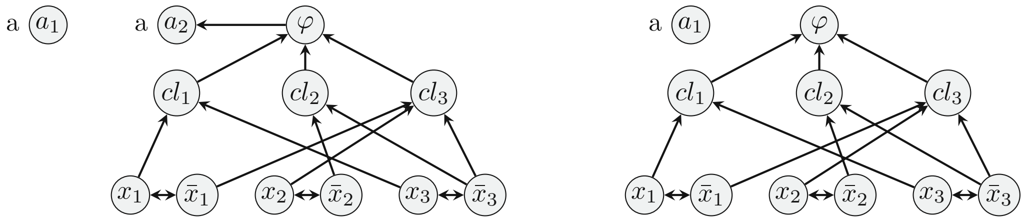


Figure 7.6: CAF $\mathcal{F}$ and $\mathcal{G}$ from the proof of Proposition 7.4.13 for a formula φ which is given by the clauses $\{\{x_1, x_3\}, \{\bar{x}_3, \bar{x}_2)\}, \{\bar{x}_1, \bar{x}_3, x_2\}\}$.

We obtain coNP-hardness of deciding ordinary equivalence with respect to i-naive semantics and Π_2^{P}-hardness for i-semi-stable and i-stage semantics for well-formed CAFs analogous to the proofs of Propositions 7.4.3 or 7.4.7 for the general case. We utilize the constructions for concurrence proofs for the CAF $\mathcal{F}$, satisfying that a given formula is unsatisfiable (invalid, respectively) iff $\sigma_i(\mathcal{F}) = \sigma_h(\mathcal{F})$ holds, and construct $\mathcal{G}$ appropriately such that $\sigma_i(\mathcal{G}) = \sigma_h(\mathcal{F})$ is satisfied.

For i-semi-stable and i-stage semantics, the CAF $\mathcal{G}$ is similar to the one constructed in Proposition 7.4.7. Moreover, for h-semi-stable and h-stage semantics, we utilize translations Tr_1 and Tr_3 to obtain Π_2^{P}-hardness of deciding ordinary equivalence.

Proposition 7.4.12. Ver-OE$_{\rho}^{wf}$ *is* Π_2^{P}*-hard for* $\rho \in \{ss_i, stg_i, ss_h, stg_h\}$.

For i-naive semantics, we modify Reduction 6.3.3 by removing argument a_2 and the attack (φ, a_2). The resulting CAF $\mathcal{G}$ has a unique claim per argument.

Proposition 7.4.13. Ver-OE$_{na_i}^{wf}$ *is* coNP*-hard.*

Proof. Consider an UNSAT instance φ given by clauses C over variables in X. We let $\mathcal{F}$ be defined as in Reduction 6.3.3. We obtain $\mathcal{G}$ from $\mathcal{F}$ by setting $A_{\mathcal{G}} = A_{\mathcal{F}} \setminus \{a_2\}$ and $R_{\mathcal{G}} = R_{\mathcal{F}} \setminus \{(\varphi, a_2)\}$. An example of this construction is given in Figure 7.6.

We can compute the naive extensions of F from G as follows: We obtain $na(F)$ from $na(G)$ by (1) taking all extensions containing φ, i.e., $E \in na(F)$ for all $E \in na(G)$ with $\varphi \in E$; (2) replacing φ in each extension by a_2, i.e., $(E \setminus \{\varphi\}) \cup \{a_2\} \in na(F)$ for all $E \in na(G)$ with $\varphi \in E$; and (3) adding a_2 to all naive extensions of E not containing φ, i.e., $E \cup \{a_2\} \in na(F)$ for all $E \in na(G)$ with $\varphi \notin E$. Hence $na(F) = \{E \mid E \in na(G), \varphi \in E\} \cup \{(E \setminus \{\varphi\}) \cup \{a_2\} \mid E \in na(G), \varphi \in E\} \cup \{E \cup \{a_2\} \mid E \in na(G), \varphi \notin E\}$.

In case (1) and (3), the set E (and its modified version) has the same claims in $\mathcal{F}$ and $\mathcal{G}$ (for the latter, observe that a_1 is contained in each extension thus each set contains claim a). Now, consider a set $E \in na(G)$ with $\varphi \in E$, and let $E' = (E \setminus \{\varphi\}) \cup \{a_2\}$. It holds that $cl(E') = X' \cup \bar{X}' \cup \{a\}$ for some $X' \subseteq X$. Observe that E' is not naive in G since E is a proper superset of it. On the other hand, the set E' is naive in $\mathcal{F}$ iff it is in conflict with each clause-argument, i.e., iff E' attacks each $cl_i \in C$. This is the case iff X' is a model of φ, i.e., iff φ is satisfiable. Hence $na_i(\mathcal{F}) = na_i(\mathcal{G})$ iff φ is unsatisfiable. □

7.4.3 Strong Equivalence and Renaming Strong Equivalence

Having established complexity results for ordinary equivalence it remains to discuss the computational complexity of strong equivalence and strong equivalence up to renaming.

Ver-SE$_\rho$
Input: Two CAFs $\mathcal{F}$, $\mathcal{G}$.
Output: TRUE iff $\mathcal{F}$, $\mathcal{G}$ are strongly equivalent w.r.t. ρ.

Recall that in Section 7.2, we have shown that strong equivalence of two CAFs $\mathcal{F}$ and $\mathcal{G}$ can be characterized via syntactic equivalence of their kernels. Since the computation and comparison of the kernels of $\mathcal{F}$ and $\mathcal{G}$ can be done in polynomial time, we obtain tractability of strong equivalence for every semantics under consideration.

Theorem 7.4.14. *The problem* Ver-SE$_\rho$ *can be solved in polynomial time for any semantics ρ considered in this work.*

Finally, we consider strong equivalence up to renaming.

Ver-SER$_\rho$
Input: Two CAFs $\mathcal{F}$, $\mathcal{G}$.
Output: TRUE iff $\mathcal{F}$, $\mathcal{G}$ are strongly equivalent up to renaming w.r.t. ρ.

As outlined above, the computation of the kernels lies in P and is therefore negligible; the complexity of verifying strong equivalence up to renaming thus stems entirely from deciding whether two labelled graphs (i.e., the kernels of the given CAFs) are isomorphic. As a consequence we obtain that the complexity of Ver-SER$_\rho$ coincides with the complexity of the well-known graph isomorphism problem. It is well-known that the graph isomorphism problem lies in NP but is not known to be NP-complete (although the latter is considered unlikely [167]).

Theorem 7.4.15. *The problem* Ver-SER$_\rho$ *is exactly as hard as the graph isomorphism problem for any semantics ρ considered in this work.*

Proof. For a reduction of the graph isomorphism problem to Ver-SER$_\rho$, consider two undirected, unlabelled graphs $F = (V, E)$ and $G = (V', E')$. We define the CAFs $\mathcal{F}$ and $\mathcal{G}$ by replacing each undirected edge by a symmetric one, moreover, each argument is labelled with the same claim. Formally, $\mathcal{F} = (V, \{(v, v'), (v', v) \mid \{v, v'\} \in E\}, cl)$ and $\mathcal{G} = (V', \{(v, v'), (v', v) \mid \{v, v'\} \in E'\}, cl)$ with $cl(v) = c$ for a fixed claim c. For any considered semantics ρ, the ρ-kernel of $\mathcal{F}$ ($\mathcal{G}$) coincides with $\mathcal{F}$ ($\mathcal{G}$, respectively): the CAFs do not contain self-attacking arguments; moreover, each conflict between arguments with the same claim is already symmetric (i.e., $(a, b) \in R$ iff $(b, a) \in R$), thus no new attacks are introduced by computing the stb_h-kernel. Hence F is isomorphic to G iff $\mathcal{F}$ and $\mathcal{G}$ are isomorphic iff $\mathcal{F}$ and $\mathcal{G}$ are strongly equivalent up to renaming w.r.t. ρ. For the other direction, we note that CAF isomorphism corresponds to the labelled variant of the graph isomorphism problem that is both edge- and label-preserving. ☐

Interestingly, when focusing on well-formed CAFs, we obtain tractability of deciding strong equivalence up to renaming. More generally speaking, it holds that the graph isomorphism problem is tractable for the class of well-formed CAFs, as we show next.

Theorem 7.4.16. *Deciding whether $\mathcal{F} \cong \mathcal{G}$ for two well-formed CAFs $\mathcal{F}$ and $\mathcal{G}$ is in* P.

Proof. We present a poly-time algorithm for deciding graph isomorphism. First, we check whether $\mathcal{F}$ and $\mathcal{G}$ have the same claims, and if, so, whether they have the same number of arguments carrying a claim. That is, we check whether $cl(A_{\mathcal{F}}) = cl(A_{\mathcal{G}})$ and then for each claim c occurring in $\mathcal{F}$, check if

$$|\{x \in A_{\mathcal{F}} \mid cl(x) = c\}| = |\{y \in A_{\mathcal{G}} \mid cl(y) = c\}|;$$

if not, stop (then $\mathcal{F}$ and $\mathcal{G}$ are not isomorphic to each other).

Otherwise, we proceed as follows: for each claim c, we first choose an unmarked argument $x \in A_{\mathcal{F}}$ with $cl(x) = c$ and compute $cl(x^-_F)$; second, we loop through all arguments $y \in A_{\mathcal{G}}$ and check whether (i) $cl(y) = c$ and (ii) $cl(y^-_G) = cl(x^-_F)$. If the search is successful, mark x and y as mapped to each other; otherwise, if such y does not exist, we stop and return 'no' (then $\mathcal{F}$ and $\mathcal{G}$ are not isomorphic to each other).

If the algorithm successfully maps each x with $cl(x) = c$ to some y with $cl(y) = c$ for each claim c occurring in both CAFs, the mapping suggested by the algorithm is an isomorphism. If not, then there is some claim c and some set C of claims s.t.

$$|\{x \in A_{\mathcal{F}} \mid cl(x) = c, cl(x^-) = C\}| \neq |\{y \in A_{\mathcal{G}} \mid cl(y) = c, cl(y^-) = C\}|,$$

i.e., no isomorphism exists. ☐

Let $\textsc{Ver-SER}_{\rho}^{wf}$ denote the problem of deciding strong equivalence up to renaming w.r.t. semantics ρ restricted to the class of well-formed CAFs. We obtain the following corollary from the above theorem.

Corollary 7.4.17. *The problem* $\textsc{Ver-SER}_{\rho}^{wf}$ *can be solved in polynomial time for any semantics ρ under consideration.*

7.5 Summary & Outlook

Summary. In this chapter, we considered ordinary and strong equivalence as well as a novel equivalence notion based on argument renaming for CAFs and well-formed CAFs w.r.t. inherited as well as hybrid semantics. We characterized strong equivalence via semantics-dependent kernels w.r.t. to all semantics considered in this work and provided a complexity analysis of all considered equivalence notions.

Our characterization results for strong equivalence are in line with existing studies for related argumentation formalisms [142, 93]. In addition, we adapt an argument-independent view by considering equivalence under renaming which models strong equivalence in situations in which the particular name of the arguments does not matter.

Our complexity analysis yields the following picture: due to our characterizations of strong equivalence via kernels, we obtain tractability of strong equivalence w.r.t. all considered semantics. In contrast, ordinary equivalence can be computationally expensive, ranging up to the third level of the polynomial hierarchy. We furthermore show that strong equivalence up to renaming has the same complexity as the graph isomorphism problem and is thus presumably of higher complexity than classical strong equivalence. When restricting the problem to the class of well-formed CAFs, we can exploit the structure of the graphs sufficiently to compute an isomorphism in polynomial time. Hence we identified a tractable fragment of the graph isomorphism problem in the course of our complexity analysis of renaming strong equivalence.

Consequences for related non-monotonic reasoning formalisms. We advance research on equivalence regarding the broad family of formalisms that can be identified as CAFs via instantiation procedures in several ways. First, our ordinary equivalence results give insights into the connection of static comparison between semantics; moreover, our complexity analysis indicates that there is no alternative to computing each extension in order to decide ordinary equivalence between two knowledge bases. Second, our kernel characterizations of strong equivalence give rise to concise representations of instantiated knowledge bases. Let us point out that the deletion of attacks might lead to frameworks that violate well-formedness—recall that well-formedness is an important property which is satisfied by many instantiation procedures as outlined in Chapter 3, so one might be afraid that such changes might lead to unwanted results. However, our results guarantee that the intended meaning of the original instance is not violated when performing such syntactic operations in the abstract representation. Similar as in propositional logics, the notion of strong equivalence allows for viewing (sub-)frameworks as independent modules that can be replaced within larger frameworks when they are strongly equivalent to each other. Consequently, our strong equivalence investigations successfully adapt desired properties of the classical equivalence notion to claim-based reasoning. Third, strong equivalence up to renaming acknowledges the important observation that the names of the arguments in instantiation procedures are often secondary. Furthermore, our abstract representation as CAF is independent of the original formalism of the considered instances. Hence it is even possible to test equivalence between argumentation systems stemming from entirely different base formalisms (assuming a common formal language or appropriate associations between claims that occur in the frameworks).

Criticisms (or: Outlook to next chapter). Our investigations regarding strong equivalence deal with the situation in which two instantiated knowledge bases are expanded with the same arguments and attacks on the abstract level. However, when expanding two knowledge bases $\mathcal{K}$ and $\mathcal{K}'$ by inserting novel elements in the original instances we might obtain different changes in the abstract level. Indeed, the addition of a new rule might add several novel attacks in $\mathcal{F}_{\mathcal{K}}$ (e.g., if we add a rule with conclusion p that contradicts the support of other arguments constructed from $\mathcal{K}$) and might result in a single isolated argument in $\mathcal{F}_{\mathcal{K}'}$ (e.g., if no argument constructed from $\mathcal{K}'$ makes use of $\neg p$). Moreover, the instantiated knowledge base often satisfies specific structural restrictions; most prominently well-formedness of the attack relation. Hence it can be reasonable to assume that the expansions of the abstract frameworks should confirm to this restriction as well. The underlying observation is that attacks between arguments that are instantiated from a knowledge base confirm to a specific structure that permit only certain attacks but can enforce others. We payed, however, little attention to restrictions imposed by the underlying formalism. While the general perspective we considered in this chapter tackles the problem of deciding strong equivalence on a very general level—and is indeed relevant for comparisons of *partial* instantiations of knowledge bases—we dedicate the next chapter to an abstract representation of dynamics in instantiated knowledge bases which takes such structural restrictions into account. In particular, we will discuss impacts of well-formedness.

Conclusion. In this chapter, we tackled the problem of (strong) equivalence in claim-based reasoning from an abstract perspective. Although we observe certain obstacles when applying our strong equivalence results to decide strong equivalence between the original instances we encounter several benefits that significantly advance the study on equivalence in claim-based argumentation. In particular, our strong equivalence characterizations give rise to concise representations of instantiated frameworks.

CHAPTER 8

Dynamics Part II: Shaping CAFs for Instantiations

Argumentation is a dynamic process. New information can enter the stage, and knowledge bases may change over time [4, 147]. In recent years, researchers in the field of formal argumentation have taken up this topic in various ways [118, 144, 155, 72]. Among the most prominent problems in this line of research is the *enforcement problem* [18, 180, 46] which deals with the manipulation of a knowledge base to ensure a certain outcome. Research concerned with this issue contributes to predict conceivable future scenarios and possible outcomes of a debate and can serve as a guidance when trying to defend a certain point of view. While manipulations of knowledge bases encompass many different operations such as addition and deletion of certain elements of the knowledge base, it is often assumed that existing knowledge persist, meaning that only the addition of new information to the knowledge base is permitted. Here, we encounter a close relation to a problem which we have already considered in this work: *strong equivalence* is concerned with the similarity of knowledge bases which expand over time.

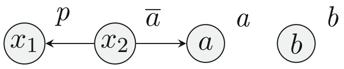
In this chapter, we study these problems for the broad class of formalisms that satisfy *well-formedness* when they are instantiated as argumentation framework. That is, formalisms satisfying that arguments with the same claim attack the same arguments in the resulting CAF. We encounter this behavior in several settings; e.g., when instantiating logic programs [77, 61] or instances of structured argumentation [116, 62] (we refer the reader to Chapter 3 for an overview of such instantiation procedures). We study enforcement and strong equivalence from a complexity-theoretic perspective with main focus on *tractable fragments*. As we have encountered in the previous chapter, the problem of strong equivalence is tractable for well-formed CAFs, even when abstracting away from the particular names of the arguments (we refer to tractability of *strong equivalence up to renaming*, cf. Definition 7.3.4, Theorem 7.4.16). Bearing in mind that strong equivalence is intractable for logic programs [146, 133] although they instantiate into well-formed

CAFs, we anticipate similar complexity-theoretic gaps in the abstract representation of structured argumentation formalisms. On the search for tractable fragments, we aim to exploit the abstract representation to obtain tractability results for a broad class of non-monotonic reasoning formalisms.

A closer inspection of the aforementioned instantiation procedures, however, reveals a certain drawback that becomes apparent when moving from static to dynamic scenarios.

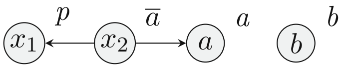
Example 8.0.1. *We consider an instantiation of an assumption-based argumentation (ABA) framework* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \overline{\ \ })$ *with assumptions* $\mathcal{A} = \{a, b\}$, *their contraries* $\overline{a}$ *and* $\overline{b}$, *resp., and rules* r_1 : '$p \leftarrow a$' *and* r_2 : '$\overline{a} \leftarrow b$'*. We obtain the associated CAF* $\mathcal{F}_D$ *as follows (cf. Section 3.3.2): each assumption* a, b *yields an argument with claims* a *and* b, *respectively; each rule* r_i *yields an argument* x_i *with claim* $head(r_i)$*. Attacks depend on the claim of the attacking argument, e.g.,* x_2 *attacks* x_1 *because* $\overline{a}$ *is the contrary of* a.

$\mathcal{F}_D$:



It turns out that we have abstracted away critical information: The rule r_2 *can be disabled by adding a rule with conclusion* $\overline{b}$, *e.g., the fact* '$\overline{b} \leftarrow$'*; this is, however, not reflected in* F_D*. To illustrate this, let us consider an adjusted version* D' *of* D *by replacing* r_2 *with rule* r_2' : '$\overline{a} \leftarrow$', *i.e.,* $\overline{a}$ *can be considered as fact. The instantiation yields the same CAF:*

$\mathcal{F}_{D'}$:



Although D *and* D' *encode different information we obtain* $\mathcal{F}_D = \mathcal{F}_{D'}$*. The CAFs do not carry sufficient information to investigate dynamics. Consider the following questions:*

- *Is it possible to accept assumption* a *by adding suitable rules? The answer is "yes" in* D, *but "no" in* D'*. This information cannot be extracted from* $\mathcal{F}_D = \mathcal{F}_{D'}$.
- *What are the stable models after adding* '$\overline{b} \leftarrow$'*? In* D, $\{a\}$ *is stable while in* D', *we obtain* $\{b\}$*. We cannot judge the situation correctly by comparing* $\mathcal{F}_D$ *and* $\mathcal{F}_{D'}$.
- *More generally, are* D *and* D' *strongly equivalent? The answer is clearly "no" when inspecting* D *and* D' *but again we cannot tell by comparing their associated CAFs.*

In all of these questions, the missing peace of information is that x_2 *has a* hidden weakness $\overline{b}$ *in* $\mathcal{F}_D$ *but not in* $\mathcal{F}_{D'}$*. It is thus impossible to attack* x_2 *in* $\mathcal{F}_{D'}$ *whereas in* $\mathcal{F}_D$, x_2 *can be attacked by an argument with conclusion* $\overline{b}$.

As this example shows, the minimal generalization to tailor CAFs suitable for dynamic settings is to extend the abstract representation with the *vulnerabilities* of an argument which describes all possibilities to attack an argument, i.e., it contains conclusions of all potential attackers. This means that for an argument $S \vdash_R p$ in the spirit of ABA [43], (i.e., atom p is derivable from assumptions S via rules R) the vulnerabilities are the contraries of the assumptions in S while p is the argument's conclusion. A potential weakness of the logic-based argument $(\{\alpha, \alpha \rightarrow \beta\}, \beta)$ is the sentence $\neg\alpha$; its conclusion is β. Considering ASPIC [137], also a rule can be a vulnerability: an argument $B : q \Rightarrow p$ with defeasible rule $d_1 : q \Rightarrow p$ can be attacked by an argument with conclusion $\neg d_1$.

In this chapter, we study enforcement and strong equivalence with main focus on ABA. We show that, as anticipated, both problems are intractable, in contrast to their counterparts in (claim-augmented) abstract argumentation. On the search for tractable fragments, we present a generalization of CAFs by augmenting arguments with *vulnerabilities* (cvAFs). This allows us to identify a fragment of ABA for which deciding enforcement and strong equivalence becomes tractable. We present cvAF characterization results for argument and conclusion enforcement and show that strong equivalence can be characterized by semantics-dependent kernels. Our results show that both problems are tractable for cvAFs. To demonstrate the flexibility of our approach, we also transfer our results to LPs and identify a fragment for which enforcement and strong equivalence is tractable.

Background & Notation. *We refer to Section 3.3.2 for background on ABA. We focus on flat, finite ABA frameworks. Moreover, we focus on* complete, preferred, grounded, *and* stable semantics. *To compare the dynamical behavior of ABA frameworks with those of (well-formed) CAFs, we also make use of results from Chapter 7.*

8.1 Dynamics in Assumption-based Argumentation

In this section, we discuss enforcement and strong equivalence notions for ABA. We show that in contrast to analogous settings in abstract argumentation, deciding enforceability as well as strong equivalence is intractable. We compare our findings with our results about claim-augmented argumentation frameworks (CAFs) established in Chapter 7.

The *expansion* of a framework is a central concept to both of our problems: naturally, expansions are an integral part of strong equivalence; moreover, since we assume that existing knowledge cannot be deleted, we study claim enforcement under the assumption that we can only add novel elements to our knowledge representation formalism. Below, we settle the notion of framework expansions for ABA frameworks. We fix $\mathcal{L}$ and a countably infinite set of assumptions $\mathcal{L}_A \subseteq \mathcal{L}$ and the contrary function $^{-} : \mathcal{L}_A \rightarrow \mathcal{L}$.

Definition 8.1.1. *For two ABA frameworks* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ *and* $D' = (\mathcal{L}, \mathcal{R}', \mathcal{A}', ^{-})$, *we call* $D \cup D' := (\mathcal{L}, \mathcal{R} \cup \mathcal{R}', \mathcal{A} \cup \mathcal{A}', ^{-})$ *the* expansion *of* D *by* D'.

For a rule $r = p \leftarrow S$, we write $D \cup \{r\}$ short for $D \cup D'$ with $D' = (\mathcal{L}, \{r\}, \emptyset, ^{-})$. By fixing $\mathcal{L}$ and the contrary function, we ensure that all expansions are compatible.

8.1.1 Conclusion Enforcement

We require that a conclusion p cannot be enforced by simply adding conclusion p or elements that introduce a novel argument with conclusion p since this would trivialize the problem. Formally, we consider the following problem:

Definition 8.1.2. *Given an ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$, a conclusion $p \in \mathcal{L}$, and a semantics σ, we say that p is* enforceable with respect to σ *iff there is some expansion D' of D (and p does not appear as conclusion in $\mathcal{H}$) such that there is $S \in \sigma_{Th}(D)$ with $p \in S$ (we say, p is credulously accepted with respect to σ in $D \cup D'$).*

We observe an interesting discrepancy between structured and abstract formalisms: While it is possible to credulously enforce any claim in a given CAF as long as a it is not self-attacking, the problem of claim enforceability is NP-hard in ABA, as we show next.

Reduction 8.1.3. *For a CNF formula φ with clauses $C = \{c_1, \dots, c_n\}$ over variables in X, we define the corresponding ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ with*

$$\mathcal{A} = \{x_a^T, x_p^F, x_a^F, x_p^T \mid x \in X\} \cup \{c, e\}$$

where $\overline{x_p^F} = x_a^T$, $\overline{x_p^T} = x_a^F$, and $\overline{c}, \overline{e}, \overline{x_a^T}, \overline{x_a^F} \in \mathcal{L} \setminus \mathcal{A}$. Also, $\mathcal{R}$ contains the following rules:

- *$\varphi \leftarrow c, e$,*
- *for all $x \in X$, $\mathcal{R}$ contains a rule $\overline{e} \leftarrow x_p^T, x_p^F$;*
- *for each $i \leq n$, $\mathcal{R}$ contains a rule of the form $\overline{c} \leftarrow \{x_a^T \mid x \in c_i\} \cup \{x_a^F \mid \neg x \in c_i\}$.*

For each variable, we introduce four assumptions, associated to different truth values on the one hand, and to 'active' (x_a^T, x_a^F) and 'passive' (x_p^T, x_p^F) assumptions on the other hand, meaning that the 'passive' assumptions cannot be defeated by newly introduced rules because their contrary is itself an assumption (recall that we are operating in flat frameworks). Figure 8.1 depicts the resulting AF for the formula $(x \vee y) \wedge (\neg x) \wedge (\neg y)$.

Theorem 8.1.4. *Deciding whether a conclusion p (assumption a) is enforceable in a given ABA framework D w.r.t. a semantics $\sigma \in \{gr, co, pr, stb\}$ is* NP*-hard.*

Proof. We present a reduction from SAT which shows hardness for grounded, complete, preferred, and stable semantics. Given a CNF formula φ with clauses $C = \{c_1, \dots, c_n\}$ over variables in X, we let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ be defined as in Reduction 8.1.3. We show φ is enforceable w.r.t. σ iff φ is satisfiable.

First assume φ is satisfiable. Let $M \subseteq X$ be a model of φ. For each $x \in M$, we introduce rules of the form $\overline{x_a^T} \leftarrow$, for each $x \notin M$, we add rules $\overline{x_a^F} \leftarrow$. Each of these conclusions is contained in the grounded extension (is derivable by the empty set of assumptions $\mathcal{E}$). Moreover, for each $x \in X$, if $\mathcal{E} \vdash \overline{x_a^T}$ then x_a^F is unattacked and thus contained in the

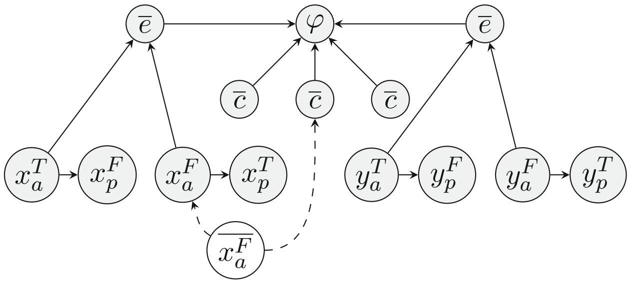


Figure 8.1: Reduction from the Proof of Theorem 8.1.4 for the formula φ given by clauses $\{x, y\}, \{\neg x\}, \{\neg y\}$; depicted with the argument arising from the additional rule $\overline{x_a^F} \leftarrow$ (fact), in white, with dashed attacks.

grounded extension G (since we have introduced a fact for each atom). G contains the assumptions c and e: Since M is a satisfying assignment of φ, each clause-rule with head $\overline{c}$ is attacked by the newly introduced rules, thus we have $c \in G$. Moreover, for every $x \in X$, either x_p^T or x_p^F is attacked by G, thus $e \in G$. We obtain $G \vdash \varphi$.

We observe that the AF arising from D is acyclic (clearly, also after adding facts to D), thus $gr(D) = co(D) = pr(D) = stb(D)$. Consequently, φ is satisfiable implies the conclusion φ is enforceable under all considered semantics.

Now assume φ is unsatisfiable. Towards a contradiction, assume φ is enforceable w.r.t. σ. That is, there is a set of rules $\mathcal{R}'$, there is a σ-assumption-set $A \subseteq \mathcal{A}$, such that φ is derivable by A in $D' = (\mathcal{L}, \mathcal{R} \cup \mathcal{R}', \mathcal{A}, ^-)$. This is the case if A defends φ against all attacks. Consequently, (a) for each $x \in X$, $\mathcal{R}'$ contains either rules with conclusion $\overline{x_a^T}$ or $\overline{x_a^F}$ but not both, otherwise both x_a^T, x_a^F are not contained in G and thus the attack on e from $\{x_p^T, x_p^F\}$ stays undefeated; also, (b) for each $i \leq n$, $\mathcal{R}'$ contains some rule with conclusion $\overline{a}$ for some $a \in A_i$, that is, either $\overline{x_a^T}$ or $\overline{x_a^F}$ for some $x \in X$. Thus for all c_i, either $G \vdash \overline{x_a^T}$ in case $x \in c_i$ or $G \vdash \overline{x_a^F}$ in case $\neg x \in c_i$. We obtain that $M = \{x \mid G \vdash x_a^T\}$ is a satisfying assignment of φ, contradiction to the assumption φ is unsatisfiable.

To show NP-hardness of assumption-enforcement, we adapt Reduction 8.1.3 as follows: we define the ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ corresponding to φ with

$$\mathcal{A} = \{x_a^T, x_p^F, x_a^F, x_p^T \mid x \in X\} \cup \{\varphi\}$$

where $\overline{x_p^F} = x_a^T$, $\overline{x_p^T} = x_a^F$, and $\overline{x_a^T}, \overline{x_a^F}, \overline{\varphi} \in \mathcal{L} \setminus \mathcal{A}$. Also, $\mathcal{R}$ contains the following rules:

- for all $x \in X$, $\mathcal{R}$ contains a rule $\overline{\varphi} \leftarrow x_p^T, x_p^F$;
- for each $i \leq n$, $\mathcal{R}$ contains a rule of the form $\overline{\varphi} \leftarrow \{x_a^T \mid x \in c_i\} \cup \{x_a^F \mid \neg x \in c_i\}$.

Considering the example in Figure 8.1, we have replaced all arguments with conclusions $\overline{e}$ or $\overline{c}$ with arguments having conclusion $\overline{\varphi}$ without changing the incoming attacks. The remaining part of the proof is analogous to the case of conclusion-enforcement. □

For CAFs on the other hand, the problem is tractable, as we show next.

Proposition 8.1.5. *A claim c is enforceable in a (well-formed) CAF $\mathcal{F}$ with respect to a semantics ρ if $\mathcal{F}$ contains a non-self-attacking argument $x \in A$ with $cl(x) = c$.*

Proof. Consider a CAF $\mathcal{F}$ and assume the existence of a non-self-attacking argument $x \in A$ with $cl(x) = c$. We enforce c by adding an argument y with a fresh claim d that attacks all other arguments in $\mathcal{F}$. Formally, we let $\mathcal{H} = (A_{\mathcal{F}} \cup \{y\}, R_{\mathcal{H}}, cl)$ with $R_{\mathcal{H}} = \{(y, z) \mid z \in A \setminus \{x\}\}$. Then the set $\{x, y\}$ is grounded as well as stable in F; thus, the argument x is credulously accepted with respect to all considered semantics. □

8.1.2 Strong Equivalence Revisited

In this section, we discuss strong equivalence for ABA. Notice that we consider strong equivalence relative to different fragments of ABA.

Definition 8.1.6. *Consider a fragment $\mathfrak{C}$ of ABA frameworks. Two ABA frameworks $D, D' \in \mathfrak{C}$ are* strongly equivalent *to each other with respect to a semantics σ iff*

1. *$\sigma(D \cup H) = \sigma(D' \cup H)$ for each $H \in \mathfrak{C}$; and*
2. *$D \cup \mathcal{H}$ and $D' \cup \mathcal{H}$ are instances of $\mathfrak{C}$.*

By adapting the proof of Theorem 8.1.4 we obtain the following result.

Theorem 8.1.7. *Deciding whether two ABA frameworks are strongly equivalent w.r.t. a given semantics $\sigma \in \{gr, co, pr, stb\}$ is* coNP*-hard.*

Proof. We present a reduction from UNSAT: Given a CNF formula φ with clauses $C = \{c_1, \dots, c_n\}$ over variables in X, we let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ be defined as in Reduction 8.1.3, and $D' = (\mathcal{L}, \mathcal{R}', \mathcal{A}, ^-)$ with $\mathcal{R}' = \mathcal{R} \setminus \{\varphi \leftarrow c, e\}$, that is, we consider two independent frameworks that differ in a single rule: D' has no argument for φ. If some expansion of D' has a σ-assumption-extension concluding φ then only because an argument with conclusion φ has been added when expanding D'. By our results from Theorem 8.1.4, we have that φ is satisfiable iff there is a set of rules $\mathcal{R}''$ such that $(\mathcal{L}, \mathcal{R} \cup \mathcal{R}'', \mathcal{A}, ^-)$ admits a σ-assumption-extension that concludes φ. Consequently, φ is satisfiable iff there is some expansion D'' of D and D' such that $\sigma(D \cup D'') \neq \sigma(D' \cup D'')$, i.e., D and D' are not strongly equivalent to each other. □

For CAFs on the other hand, the problem is tractable as we have seen in the previous chapter. To decide strong equivalence for CAFs, one only needs to compute the kernels of both frameworks and check their syntactical coincidence. As we have seen, even the problem of deciding renaming strong equivalence is tractable for well-formed CAFs.

Generally speaking, we observe a discrepancy between knowledge bases and their abstract representation complexity-wise. While status enforcement and strong equivalence is tractable for CAFs, the problems are NP-hard resp. coNP-hard with respect to ABA. Interestingly, when inspecting the proofs of Theorem 8.1.4 and 8.1.7 we observe that the intractability even holds for atomic ABA frameworks. On the search for tractable fragments, we go one step beyond well-formedness and consider a small adaption of CAFs which turns out to handle dynamic situations on abstract level very well.

8.2 An Instantiation for Dynamics

In this section we will augment the standard instantiation procedure with some additional information in order to make it better suitable for dynamic scenarios. Thereby, we will obtain so-called cvAFs ("claim and vulnerability augmented AFs") which extend CAFs with additional information concerning the occurring arguments. It is, however, clear that an exact correspondence of the reasoning problems in ABA and cvAFs would again yield an intractable notion of enforcement and strong equivalence. This is why our cvAFs will be developed in a way that they carry just enough information in order to correspond to a meaningful fragment of ABA, while the aforementioned tasks stay tractable. This way, we obtain tractable fragments for Theorems 8.1.4 and 8.1.7.

Instantiated Arguments. Our cvAFs incorporate a crucial observation regarding the instantiations of knowledge bases which adhere well-formedness: arguments are typically characterized by their *claim* and their potential weaknesses (*vulnerabilities*) on which they can be attacked. While CAFs assign each argument a claim via a function, we go one step further and *identify* arguments in cvAFs with their claims and vulnerabilities.

Definition 8.2.1. *Given a set $\mathcal{L}$ of sentences. An $\mathcal{L}$-*instantiated argument *is a tuple $x = (vul(x), cl(x))$ where $vul(x) \subseteq \mathcal{L}$ are the* vulnerabilities *and $cl(x) \in \mathcal{L}$ is the* conclusion *of x.*

$\mathcal{L}$-instantiated arguments are a flexible tool and may stem from an arbitrary instantiation procedure which makes use of conclusions and vulnerabilities in a certain sense. For instance, in the context of ABA frameworks, we obtain instantiated arguments as follows:

for an ABA argument $S \vdash p$, we obtain the instantiated argument $(\overline{S}, p)$.

That is, the claim of the $\mathcal{L}$-instantiated argument is the conclusion of the ABA argument (as usual) and the vulnerabilities correspond to the contraries of the assumptions used in the ABA argument. Note that this representation is not restricted to assumption-based argumentation. The vulnerabilities of arguments obtained from logic programs correspond to the negated atoms of the rules used in the construction. For logic-based argumentation, the vulnerabilities of an argument are obtained by negating the premises; for ASPIC+, we furthermore consider the negation of defeasible rules as part of the vulnerabilities.

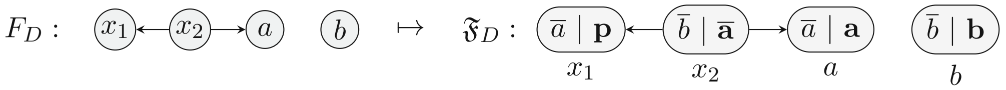
$\mathcal{L}$-instantiated arguments thus provide a uniform representation for arguments with claims and defeasible elements.

We are ready to formally introduce cvAFs as generalization of AFs by replacing abstract arguments with $\mathcal{L}$-instantiated arguments.

Definition 8.2.2. *A cvAF is a tuple* $\mathfrak{F} = (A, R)$ *(in* $\mathcal{L}$*) where* A *is a set of* $\mathcal{L}$*-instantiated arguments and* $R \subseteq A \times A$*.*

Notation 8.2.3. *In the remaining part of the chapter, we drop* $\mathcal{L}$ *and simply say 'instantiated arguments' whenever no ambiguity arises.*

An example of a cvAF is given by the representation of our running example as cvAF (cf. $\mathfrak{F}_D$ below). Here, each argument contains its vulnerabilities (left) and its conclusion (right, in boldface), e.g., argument x_1 has a single vulnerability $\bar{a}$ and conclusion p.



Each cvAF $\mathfrak{F} = (A, R)$ corresponds to a CAF $\mathcal{F} = (A, R, cl)$ (where cl corresponds to the claim-function as in Definition 8.2.1). Hence our cvAFs are a proper generalization of CAFs. This means that *all* results regarding CAFs established in this work carry over to cvAFs. In a similar fashion, we can make use of results established for AFs.

Notation 8.2.4. *For a cvAF* $\mathfrak{F} = (A, R)$*, we write* $\mathcal{F}$ *to denote the corresponding CAF* (A, R, cl) *and* F *to denote the corresponding AF* (A, R)*.*

We make use of functions and notations for AFs and CAFs. For a cvAF $\mathfrak{F}$ and a set of arguments E we write $E^+_{\mathfrak{F}}(= E^+_F)$ to denote all arguments attacked by E, $E^*_{\mathfrak{F}}(= E^*_{\mathcal{F}})$ to denote all claims defeated by E, and $\Gamma_{\mathfrak{F}}(E)(= \Gamma_F(E))$ to denote the set of arguments which are defended. Other notations are transferred to cvAFs accordingly. Semantics for cvAFs can be defined in terms of arguments or of claims.

Definition 8.2.5. *Given an cvAF* $\mathfrak{F}$*, an AF semantics* σ*, and a CAF semantics* ρ*. We let* $\sigma(\mathfrak{F}) = \sigma(F)$ *denote the* σ*-argument-extensions and* $\rho(\mathfrak{F}) = \rho(\mathcal{F})$ *the* ρ*-conclusion-extensions of* $\mathfrak{F}$*.*

Adapting our standard notation for CAFs, we write $\sigma_i(\mathfrak{F})$ for the inherited variant of σ and $\sigma_h(\mathfrak{F})$ when evaluating an cvAF $\mathfrak{F}$ with respect to the hybrid variant of σ.

Well-formedness in cvAFs. With cvAFs, we can define well-formedness in dependency of the claims and the vulnerabilities of the arguments as follows.

Definition 8.2.6. *A cvAF* $\mathfrak{F} = (A, R)$ *is called* well-formed *iff it satisfies:* $(x, y) \in R$ *iff* $cl(x) \in vul(y)$ *for each* $x, y \in A$*.*

As we have seen by now, the property of well-formedness entails many useful properties. For instance, inherited and hybrid preferred and stable variants of the semantics coincide (cf. Proposition 4.1.5 and 4.1.16), moreover, each σ-argument-extension corresponds to a unique σ-conclusion-extension for $\sigma \in \{gr, co, pr, stb\}$ (cf. Proposition 3.1.2).

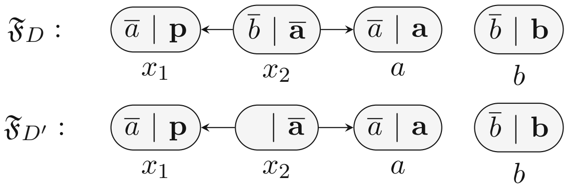
Let us note that in contrast to well-formed CAFs, a well-formed cvAF does not possess copies of arguments (cf. Definition 3.1.16) since the pair $(\textsc{Vul}(x), cl(x))$ can appear in a cvAF only once.

ABA and cvAFs. Let us now see or formalism at work when applied to ABA frameworks. We adapt the standard instantiation of ABA (cf. Translation 3.3.15, 3.3.18) as follows.

Definition 8.2.7. *For an ABA framework D, $\mathfrak{F}_D = (A, R)$ is the cvAF with instantiated arguments $A = \{(\overline{S}, p) \mid S \vdash p\}$ and $(x, y) \in R$ iff $cl(x) \in vul(y)$.*

Our cvAF instantiation is a faithful generalization of the usual one; the instantiation preserves the semantics of the original instance. moreover, each instantiated cvAF is well-formed. This follows directly from results established for CAFs in Section 3.3.2.

Example 8.2.8. *When instantiating our ABA frameworks D and D' from Example 8.0.1 as cvAFs, we obtain the following picture:*



Comparing these instantiations with our CAF instantiations from Example 8.0.1, we observe a crucial difference: while the CAFs corresponding to D and D' are identical we observe that $\mathfrak{F}_D$ and $\mathfrak{F}_{D'}$ differ. Indeed, the argument x_2 has vulnerability $\textsc{Vul}_D(x_2) = \{\bar{b}\}$ in $\mathfrak{F}_D$ but no vulnerabilities in $\mathfrak{F}_{D'}$.

Since our formalism of interest yields well-formed cvAFs, we restrict our studies to *well-formed cvAFs only.*

Assumption 8.2.9. *In the remaining part of this chapter, we assume that each cvAF is well-formed.*

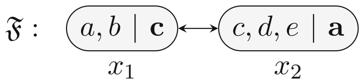
cvAFs and Dynamics. We are ready to investigate dynamics in structured argumentation by means of cvAFs. Suppose we are given a knowledge base $\mathcal{K}$ and the instantiated cvAF $\mathfrak{F}_{\mathcal{K}}$. If we want to move to a superset $\mathcal{K} \cup \mathcal{H}$ we can construct $\mathfrak{F}_{\mathcal{K} \cup \mathcal{H}}$ immediately by inspecting the relevant conclusions and vulnerabilities.

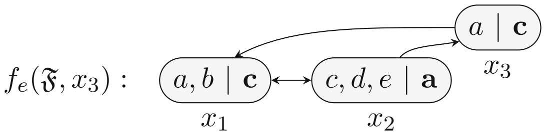
Definition 8.2.10. *Given a cvAF $\mathfrak{F} = (A, R)$ and an instantiated argument x we define the expansion $f_e(\mathfrak{F}, x)$ of $\mathfrak{F}$ with x by letting $f_e(\mathfrak{F}, x) = (A \cup \{x\}, R_x)$ be the cvAF where*

$$\begin{aligned} R_x = R &\cup \{(x, y) \mid y \in A, cl(x) \in vul(y)\} \\ &\cup \{(y, x) \mid y \in A, cl(y) \in vul(x)\}. \end{aligned}$$

We stipulate that $f_e(\mathfrak{F}, X)$ is a shorthand for successively expanding $\mathfrak{F}$ with each $x \in X$ in an arbitrary order.

Example 8.2.11. *Consider the cvAF $\mathfrak{F}$ with mutually attacking arguments $x_1 = (\{a, b\}, c)$ and $x_2 = (\{c, d, e\}, a)$. The expansion of $\mathfrak{F}$ with argument $x_3 = (\{a\}, c)$ induces the attacks (x_3, x_1) and (x_2, x_3). We depict both cvAFs below.*





cvAFs and Atomic ABA Frameworks. Our cvAFs are closely related to atomic ABA frameworks (cf. Definition 3.3.31) in dynamic scenarios as we will discuss next. There are several decisive observations we make about atomic ABA frameworks.

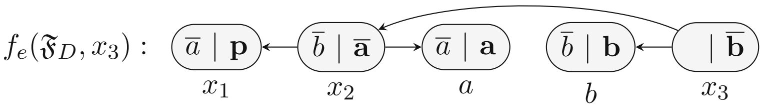
Lemma 8.2.12. *Given an atomic ABA $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$.*

- *If $a \in \mathcal{A}$, then $\mathfrak{F}_{D \cup \{r\}} = f_e(\mathfrak{F}_D, x)$ with $x = (\overline{a}, a)$;*
- *for each atomic (in D) rule $r = p \leftarrow S$, we have $\mathfrak{F}_{D \cup \{r\}} = f_e(\mathfrak{F}_D, x)$ with $x = (\overline{S}, p)$;*
- *for each $x = (\overline{S}, p)$, we have $\mathfrak{F}_{D \cup H} = f_e(\mathfrak{F}_D, x)$ with $H = (\mathcal{L}, \{p \leftarrow S\}, S, ^{-})$.*

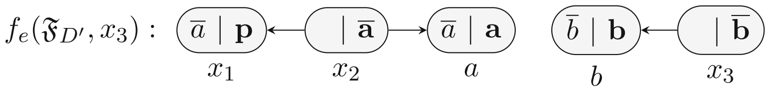
By moving from general to atomic ABA frameworks we do not lose expressive power; each framework can be transformed into an atomic one (cf. Section 3.3.2). The translation might result in an exponential blow-up in the number of rules. However, given an atomic ABA framework D we can be sure that the instantiated cvAF $\mathfrak{F}_D$ is of linear size in D.

Proposition 8.2.13. *If $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ is atomic, then $\mathfrak{F}_D$ has $|\mathcal{R}| + |\mathcal{A}|$ arguments.*

Example 8.2.14. *Let D be our running example and $\mathfrak{F}_D$ its instantiated cvAF. Adding a fact "$\overline{b}$." yields an additional instantiated argument $x_3 = (\emptyset, \overline{b})$: Since $\mathfrak{F}_D$ rightfully encodes that conclusion $\overline{b}$ is a threat to x_2, the instantiation of the resulting ABA framework can be directly computed from $\mathfrak{F}_D$ by adding the argument $x_3 = (\emptyset, \overline{b})$.*



If we consider the expansion of $\mathfrak{F}_{D'}$ instantiated from the ABA framework D' from Example 8.0.1 with the same argument x_3 instead, we obtain the following picture:



We obtain a similar cvAF, but x_2 does not have any vulnerability. Hence we are indeed able to distinguish the two instantiations as desired.

8.3 The cvAF Enforcement Problem

In this section we develop a notion of the enforcement problem for cvAFs and establish criteria for deciding enforceability. At first glance, this yields results applicable to atomic ABAs due to Lemma 8.2.12; we will, however, discuss some subtle details of the notions which one needs to be aware of.

In line with our enforcement notion from Definition 8.1.2, we define conclusion enforcement for cvAFs by requiring that no new argument with the target conclusion is introduced. In addition, we introduce a natural notion of argument enforcement.

Definition 8.3.1. *Let $\mathfrak{F} = (A, R)$ be a cvAF and σ a semantics. A conclusion p is* σ-enforceable *if there is a set X of instantiated arguments s.t. $p \notin cl(X)$ and p is credulously accepted in $f_e(\mathfrak{F}, X)$. An argument $x \in A$ is* σ-enforceable *if there is a set X of instantiated arguments s.t. $cl(x) \notin cl(X)$ and x is credulously accepted in $f_e(\mathfrak{F}, X)$.*

Example 8.3.2. *Let $\mathfrak{F}_D$ be our running example cvAF and consider the expansion $f_e(\mathfrak{F}_D, x_3)$ with $x_3 = (\emptyset, \bar{b})$ (cf. Example 8.2.14). Since $co(f_e(\mathfrak{F}_D, x)) = \{\{a, x_1, x_3\}\}$ with $cl(x_1) = p$ we obtain that conclusion p is co-enforceable.*

In the following we establish criteria to decide whether arguments and conclusions are enforceable in cvAFs. By definition, it suffices to focus on argument enforcement:

Proposition 8.3.3. *Let $\mathfrak{F} = (A, R)$ be a cvAF and σ a semantics. A conclusion $c \in cl(A)$ is enforceable iff there is some $x \in A$ with $cl(x) = c$ s.t. x is enforceable.*

The possible modifications of a cvAF are determined by the conclusions and vulnerabilities of its arguments. It is thus not possible to consider arbitrary expansions. We already saw this for our running example $\mathfrak{F}_{D'}$ where a is not enforceable since it is attacked by some argument without any vulnerability (cf. Example 8.2.14).

In general, arguments without any vulnerability will always be defended in complete-based semantics. This is not only the case within the given cvAF, but also for any conceivable expansion. Motivated by this observation, we call such arguments *strongly defeated.*

Definition 8.3.4. *For a cvAF* $\mathfrak{F} = (A, R)$*,* $x \in A$ *is* strongly defeated *if there is* $y \in A$ *with* $(y, x) \in R$ *and* $vul(y) = \emptyset$*.*

Example 8.3.5. *In our running example cvAF stemming from instantiating* D'*, the argument* x_1 *is strongly defeated. In fact, it is verifiable with reasonable effort that* x_2 *is part of the grounded extension in any possible expansion* $f_e(\mathfrak{F}, X)$*.*

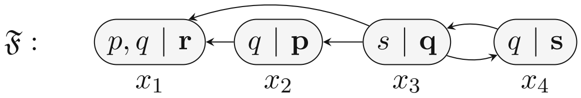
The following proposition formalizes that the behavior we observed in the previous example generalizes to any cvAF and verifies our intuition about strong defeat.

Proposition 8.3.6. *Let* $\mathfrak{F} = (A, R)$ *be a cvAF. If* $x \in A$ *is strongly defeated, then for each set* X *of instantiated arguments, the grounded extension of* $f_e(\mathfrak{F}, X)$ *attacks* x*.*

Proof. Let $y \in A$ with $(y, x) \in R$ and $vul(y) = \emptyset$. In each expansion $f_e(\mathfrak{F}, X)$, y remains unattacked. Hence $y \in G \in gr(f_e(\mathfrak{F}, X))$ for each set X of instantiated arguments. ☐

Hence strongly defeated arguments can never be enforced. It is therefore a reasonable conjecture that an argument is enforceable iff it is not strongly defeated. However, as the following example illustrates, the notion of strong defeat is not yet general enough.

Example 8.3.7. *Consider the cvAF* $\mathfrak{F}$ *depicted below.*



Suppose we want to enforce x_1*. In order to achieve this goal we have to add an argument defeating* x_2*. However, the only vulnerability of* x_2 *is* q *and due to* $q \in vul(x_1)$*, such an argument would defeat* x_1 *as well.*

In general, if there is some argument y with $(y, x) \in R$ and $vul(y) \subseteq vul(x)$, then x can never be defended by a conflict-free set. We call arguments of this kind *strongly unacceptable* since this holds also true for any expansion.

Definition 8.3.8. *For a cvAF* $\mathfrak{F} = (A, R)$*,* $x \in A$ *is* strongly unacceptable *if there is* $y \in A$ *with* $(y, x) \in R$ *and* $vul(y) \subseteq vul(x)$*.*

By definition, each strongly defeated argument is strongly unacceptable. For $\sigma \in \{co, pr, stb\}$ we are now ready to state our enforcement results.

Theorem 8.3.9. *Let* $\mathfrak{F} = (A, R)$ *be a cvAF and suppose* $\sigma \in \{co, pr, stb\}$*. An argument* $x \in A$ *is* σ*-enforceable if and only if it is not strongly unacceptable.*

Proof. ($\Rightarrow$) Suppose x is strongly unacceptable and let $x \in A$ with $(y, x) \in R$ and $vul(y) \subseteq vul(x)$. Assume that in some expansion $f_e(\mathfrak{F}, X)$ we have $x \in E$ for $E \in ad(f_e(\mathfrak{F}, X))$. Since E defends x, E attacks y. Due to $vul(y) \subseteq vul(x)$, E attacks x as well; contradiction.

($\Leftarrow$) Suppose x is not strongly unacceptable. First consider $\sigma \neq stb$. Then x is no self-attacker since otherwise it would be strongly unacceptable: Let $y := x$ and we get $(y, x) \in R$ and $vul(y) \subseteq vul(x)$. So let $y_1, \ldots, y_n$ be the set of arguments in $\mathfrak{F}$ attacking x. For each i we have $vul(y_i) \setminus vul(x) \neq \emptyset$ since x is not strongly unacceptable. Consider some $p_i \in vul(y_i) \setminus vul(x) \neq \emptyset$ for each i. Let $x_i = (p_i, \emptyset)$ be unattacked instantiated arguments with the p_i as respective claim. Set $X = \{x_1, \ldots, x_n\}$. It is straightforward to see that $x \cup X$ is admissible in $f_e(\mathfrak{F}, X)$.

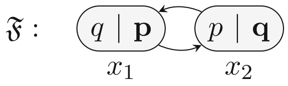
Now let $\sigma = stb$. Here in addition we need to ensure that there is at least one extension in our expansion. We will proceed by taking care of each self-attacker as well as each odd cycle in the given cvAF.

- Let $z_1, \ldots, z_m$ be the set of self-attacking arguments in $\mathfrak{F}$. Suppose for some i we have $vul(z_i) \subseteq vul(x)$. Then since $(z, z) \in R$ we also have $(z, x) \in R$ and we infer strong unacceptability of x; contradiction. So we infer $vul(z_i) \nsubseteq vul(x)$ for each i. Therefore, we can find $q_i \in vul(z_i) \setminus vul(x) \neq \emptyset$ for each i. Let $x'_i = (q_i, \emptyset)$ be unattacked instantiated arguments with the q_i as respective claim. Set $X' = \{x'_1, \ldots, x'_n\}$; we have considered each self-attacking argument in $\mathfrak{F}$.
- Let $O = \{o_1, \ldots, o_n\}$ be an arbitrary but fixed set of arguments forming an odd cycle in $\mathfrak{F}$. We argue that there is some $o_i \in O$ satisfying the usual condition, i.e., $vul(o_i) \nsubseteq vul(x)$. Otherwise each $o_i \in O$ attacks x since $(o_i, o_j) \in R$ implies $(o_i, x) \in R$ due to $vul(o_j) \subseteq vul(x)$ (we get $(o_i, o_j) \in R$ from the fact that O is a cycle). From this we get that x is strongly unacceptable; a contradiction. So take $o_i \in O$ with $vul(o_i) \nsubseteq vul(x)$ and construct an argument x'' attacking it as usual. Since O was arbitrary, we proceed like this for each odd cycle, obtaining a third set X'' of arguments.

In $f_e(\mathfrak{F}, X \cup X' \cup X'')$ the set $C = x \cup X \cup X' \cup X''$ is admissible and each odd cycle is resolved. Due to [22, Theorem 5.7], C can be extended to a stable extension of $\mathfrak{F}$. $\square$

For grounded semantics, however, we need to consider further unacceptability notions. The reason why Theorem 8.3.9 does not hold for grounded semantics is that an argument might be capable of defending itself, but is still not part of the iterative procedure which yields the grounded extension. To illustrate this we consider the following example.

Example 8.3.10. *Suppose we aim to gr-enforce x_1 in $\mathfrak{F}$:*



Since $gr(\mathfrak{F}) = \emptyset$ *we would have to add an argument defeating* x_2 *to defend* x_1*. However, an argument achieving this would possess* p *as conclusion which we want to avoid for this version of the enforcement notion. Indeed,* x_1 *is not* gr*-enforceable.*

In general, for grounded semantics we require a notion which is similar to strong unacceptability, while taking the special case we just illustrated into account.

Definition 8.3.11. *For a cvAF* $\mathfrak{F} = (A, R)$*,* $x \in A$ *is* strongly gr-unacceptable *if there is* $y \in A$ *with* $(y, x) \in R$ *and* $vul(y) \setminus \{cl(x)\} \subseteq vul(x)$*.*

The following condition characterizes gr-enforceability for cvAFs. Although it may appear technical at first glance, it simply ensures that an argument z can be defeated without attacking x, y, or introducing the target claim $cl(x) = cl(y)$.

Proposition 8.3.12. *Let* $\mathfrak{F} = (A, R)$ *be a cvAF. An argument* $x \in A$ *is* gr*-enforceable if and only if one of the following two conditions hold:*

- x *is not strongly* gr*-unacceptable,*
- *there is some* $y \in A$ *with* $cl(y) = cl(x) = q$ *s.t.*
  - *if* z *attacks* y*, then* $vul(z) \setminus (vul(x) \cup vul(y) \cup \{q\}) \neq \emptyset$*,*
  - *if* z *attacks* x*, then* $q \in vul(z)$ *or* $vul(z) \setminus (vul(x) \cup vul(y)) \neq \emptyset$*.*

Proof. ($\Leftarrow$) First suppose x is not gr-unacceptable. As usual let $w_1, \ldots, w_n$ be the set of attackers of x. We have $(vul(w_i) \setminus \{cl(x)\}) \setminus vul(x) \neq \emptyset$, so we take one conclusion $p_i \in (vul(w_i) \setminus \{cl(x)\}) \setminus vul(x)$, introduce corresponding instantiated arguments $(p_i, \emptyset)$ and obtaining a set X s.t. x is defended by X in $f_e(\mathfrak{F}, W)$.

Now suppose the second condition is true and consider $y \in A$ as described.

- Let $z_1, \ldots, z_n$ be the set of arguments attacking y. As usual, we take conclusions $p_i \in vul(z_i) \setminus (vul(x) \cup vul(y) \cup \{q\})$.
- Let $z'_1, \ldots, z'_m$ be the set of arguments attacking x. For each z'_i with $q \notin vul(z'_i)$ consider a conclusion $q_i \in vul(z) \setminus (vul(x) \cup vul(y))$.

Let Z be the set of instantiated arguments with the considered conclusions as claims and no vulnerabilities. By construction, Z defends y in $f_e(\mathfrak{F}, Z)$ and defeats each attacker of x not having q as vulnerability; arguments of this kind are defeated due to y being defended. That is, $Z \cup \{x, y\}$ is part of the grounded extension of $f_e(\mathfrak{F}, Z)$.

($\Rightarrow$) Suppose both conditions are false, i.e., x is strongly gr-unacceptable and there is no y satisfying the two mentioned conditions. If x is even strongly unacceptable, we are done since this would even prevent us from enforcing x w.r.t. co semantics.

So suppose x is not strongly unacceptable, but strongly gr-unacceptable. Then there is some argument z attacking x and $vul(z) \setminus \{q\} \subseteq vul(x)$. Hence, in order for x to be in the grounded extension, we need to ensure defense of some argument different from x with claim q. Take some y with $cl(y) = q$ (if none exists, we are done). By assumption, at least one of the mentioned conditions is wrong.

- Suppose $vul(z) \setminus (vul(x) \cup vul(y) \cup \{q\}) = \emptyset$ for some attacker z of y. However, this means by introducing an argument not having q as conclusion we can never ensure defeat of z without also defeating either x or y. We cannot introduce arguments which defeat x and defeating y means we need to move to another y' with claim q.
- Now suppose some z attacking x with $q \notin vul(z)$ satisfies $vul(z) \setminus (vul(x) \cup vul(y)) = \emptyset$. As before, this means we cannot defend x from z without introducing arguments which also defeat either x or y; again this means we need to move to another y'. ☐

Let us now discuss corresponding results for conclusion enforcement. To enforce a conclusion $p \in cl(A)$ we need to enforce an argument $x \in A$ with $cl(x) = p$. Thus, as a corollary of Theorem 8.3.9 and Proposition 8.3.12 we obtain:

Corollary 8.3.13. *Let $\mathfrak{F} = (A, R)$ be a cvAF and $\sigma \in \{ad, co, pr, stb\}$. A conclusion $p \in cl(A)$ is σ-enforceable iff there is an argument $a \in A$ with $cl(a) = p$ which is not strongly unacceptable; it is gr-enforceable iff there is an argument $a \in A$ with $cl(a) = p$ which is not strongly gr-unacceptable.*

8.3.1 Consequences for Assumption-based Argumentation

The introduced unacceptability notions yield syntactical conditions to decide the cvAF enforcement problem in polynomial time. In view of this, it might seem that Lemma 8.2.12 now implies tractability of the enforcement problem for atomic ABA frameworks. However, when inspecting the construction for the proof of Theorem 8.1.4 we see that the constructed ABA framework is atomic itself.

Corollary 8.3.14. *Deciding whether assumption a (conclusion p) is enforceable w.r.t. σ is* NP*-hard even for atomic ABA frameworks.*

The reason why this is no contradiction to tractability in cvAFs is rather subtle: When considering an arbitrary expansion $f_e(\mathfrak{F}, X)$, it might happen that the resulting cvAF does not correspond to a flat ABA framework anymore due to $cl(X) \cap \mathcal{A} \neq \emptyset$. So even if we start with a cvAF corresponding to some flat ABA framework, we do not have a one to one correspondence between expansions of the cvAF and flat ABA framework extending the initial one.

To ensure that it is not necessary to introduce arguments with assumptions as conclusion when enforcing an argument $x \in A$, we recall the fragment of *ABA frameworks with separated contraries* (cf. Definition 3.3.26) where assumptions do not have out-going attacks: an ABA framework $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ has *separated contraries* if $\mathcal{A} \cap \overline{\mathcal{A}} = \emptyset$. We are now ready to introduce a tractable fragment for the ABA enforcement problem.

Theorem 8.3.15. *Deciding whether an argument or conclusion is enforceable for atomic flat ABAs with separated contraries is tractable.*

Proof. Let $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ be an atomic flat ABA framework with separated contraries. We apply Lemma 8.2.12; from Proposition 3.3.25 obtain that for each $p \in \mathcal{L}$ we have that p is enforceable in D iff the p is enforceable in $\mathfrak{F}_D = (A_D, R_D)$ disregarding any expansion $f_e(\mathfrak{F}, X)$ where $cl(X) \cap \mathcal{A} \neq \emptyset$. The proofs given for the enforcement results for cvAFs only require addition of arguments with out-going attacks. Since D has separated contraries, $vul(A_D) \cap \mathcal{A} = \emptyset$ and we can assume $cl(X) \cap \mathcal{A} \neq \emptyset$ in each expansion $f_e(\mathfrak{F}, X)$ without loss of generality, i.e., the conclusion p is enforceable in $\mathfrak{F}_D = (A_D, R_D)$ disregarding any expansion $f_e(\mathfrak{F}, X)$ where $cl(X) \cap \mathcal{A} \neq \emptyset$ iff the conclusion p is enforceable in $\mathfrak{F}_D = (A_D, R_D)$. Hence p is enforceable in D iff the conclusion p is enforceable in $\mathfrak{F}_D = (A_D, R_D)$. By Corollary 8.3.14 we obtain tractability of the enforcement problem in the considered ABA fragment. ☐

We want to emphasize that moving from flat ABA to flat atomic ABA does not change the complexity class of the enforcement problem; but additionally requiring separated contraries does, i.e., we found a rather minor condition pushing the enforcement problem over the edge to tractability.

8.4 The cvAF Strong Equivalence Problem

In this section, we establish methods to decide strong equivalence for cvAFs. We define further unacceptability notions, tailored for this setting. In accordance with the standard literature on strong equivalence we then can decide this problem for two cvAFs by comparing their so-called *kernels*, that is, we transform both cvAFs into a semantics-dependent normal form.

Let us point out the following crucial difference: In contrast to strong equivalence characterizations in Dung AFs [142], SETAFs [93], and CAFs (cf. Chapter 7) where kernels are constructed by removing redundant *attacks*, we identify redundant *arguments*. The kernels in cvAFs are constructed by *removing* as well as *manipulating* arguments that fall in certain redundancy categories.

We start by defining an appropriate strong equivalence notion for cvAFs.

Definition 8.4.1. *Two cvAFs $\mathfrak{F}$, $\mathfrak{G}$ are* strongly equivalent *w.r.t. a semantics σ, for short $\mathfrak{F} \equiv_s^\sigma \mathfrak{G}$, if for each set X of instantiated arguments $\sigma_i(f_e(\mathfrak{F}, X)) = \sigma_i(f_e(\mathfrak{G}, X))$ holds.*

Example 8.4.2. *Consider again the cvAFs $\mathfrak{F}_D$ and $\mathfrak{F}_{D'}$ from Example 8.0.1. Judging from earlier results we anticipate that they are not strongly equivalent to each other.*

Indeed, if we recall the expansions of $\mathfrak{F}_D$ *and* $\mathfrak{F}_{D'}$ *from Example 8.2.14 where we add the argument* $x_3 = (\emptyset, \bar{b})$ *to both frameworks, we obtain that* $\{a, p, \bar{b}\}$ *is stable in* $f_e(\mathfrak{F}_D, x_3)$ *but not in* $f_e(\mathfrak{F}_{D'}, x_3)$*. Hence* $\mathfrak{F}_D$ *and* $\mathfrak{F}_{D'}$ *are not strongly equivalent w.r.t. stable semantics.*

In the above example, it was quite easy to come up with an appropriate counter example. Not only that finding a counter example might be more involved in other situations, it is usually not possible to verify strong equivalence by testing all possible expansions because there might be infinitely many of them. Instead, we identify semantics-dependent kernels – checking strong equivalence reduces to computing and comparing the respective kernels.

Let us start with some general observations regarding *redundancies* of cvAFs. We first recall a redundancy notion which we have already encountered in Section 3.1 in the context of CAFs: an argument x in a CAF $\mathcal{F}$ is called redundant w.r.t. argument y iff they have the same claim and attack the same arguments, but x is attacked by strictly more arguments than y, i.e., $y^- \subset x^-$. This concept is naturally adapted to cvAFs as follows:

Definition 8.4.3. *For a cvAF* $\mathfrak{F} = (A, R)$*,* $x \in A$ *is* redundant *if there is* $y \in A$ *with* $cl(y) = cl(x)$ *and* $vul(y) \subset vul(x)$*.*

Example 8.4.4. *The argument* x_2 *from the cvAF* $\mathfrak{F}_D$ *from our running example is redundant w.r.t.* $x = (\emptyset, \bar{a})$ *because* $cl(x) = cl(x_2) = \bar{a}$ *and* $vul(x) = \emptyset \subset \{\bar{b}\} = vul(x_2)$*.*

As shown in Section 3.1, redundant arguments can be removed without changing the conclusion-σ-extensions of a given cvAF for $\sigma \in \{gr, co, pr, stb\}$.

Proposition 8.4.5. *For a cvAF* $\mathfrak{F} = (A, R)$*, a semantics* $\sigma \in \{gr, co, pr, stb\}$ *and a redundant argument* $x \in A$*, it holds that* $\sigma_i(\mathfrak{F}) = \sigma_i(\mathfrak{F} \setminus \{x\})$*.*

Next, we reconsider the unacceptability notions from Section 8.3. We have shown that strongly defeated arguments cannot be enforced; in fact, they can be removed without changing the σ-extensions.

Proposition 8.4.6. *For a cvAF* $\mathfrak{F} = (A, R)$*, semantics* $\sigma \in \{gr, co, pr, stb\}$*, and a strongly defeated argument* $x \in A$*, it holds that* $\sigma_i(\mathfrak{F}) = \sigma_i(\mathfrak{F} \setminus \{x\})$*.*

Proof. Let $\mathfrak{F}' = \mathfrak{F} \setminus \{x\}$, and let y with $\textsc{Vul}(y) = \emptyset$ denote some argument which strongly defeats x. Note that y is contained in the grounded extension of both $\mathfrak{F}$ and $\mathfrak{F}'$; moreover, the grounded extension of $\mathfrak{F}$ and $\mathfrak{F}'$ coincides since $y \in \Gamma_{\mathfrak{F}}(\emptyset)$ defeats x. Therefore,

$$\Gamma^i_{\mathfrak{F}}(\emptyset) \subseteq \Gamma^i_{\mathfrak{F}'}(\emptyset) \qquad \text{and} \qquad \Gamma^i_{\mathfrak{F}'}(\emptyset) \subseteq \Gamma^{i+1}_{\mathfrak{F}}(\emptyset).$$

We obtain

$$gr(\mathfrak{F}) = \bigcup_{i \in \mathbb{N}} \Gamma^i_{\mathfrak{F}}(\emptyset) = \bigcup_{i \in \mathbb{N}} \Gamma^i_{\mathfrak{F}'}(\emptyset) = gr(\mathfrak{F}').$$

Moreover, $E^+_{\mathfrak{F}} \setminus \{x\} = E^+_{\mathfrak{F}'}$ for every set of arguments E which is complete in $\mathfrak{F}$ or $\mathfrak{F}'$ since the grounded extension is contained in each complete extension. Hence $\Gamma_{\mathfrak{F}}(E) = \Gamma_{\mathfrak{F}'}(E)$ for each set $G \subseteq E$. Since each complete extension is a superset of G, we obtain $co(\mathfrak{F}) = co(\mathfrak{F}')$. It follows also that preferred semantics coincide. Regarding stable semantics, we argue analogously since each stable extension is a superset of G. ☐

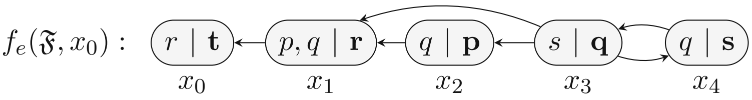
For stable semantics we can make an even stronger assertion: Not only strongly defeated, but also strongly unacceptable arguments can be deleted without affecting the outcome.

Proposition 8.4.7. *For a cvAF $\mathfrak{F} = (A, R)$ and a strongly unacceptable argument $x \in A$, it holds that $stb_i(\mathfrak{F}) = stb_i(\mathfrak{F} \setminus \{x\})$.*

Proof. Let $\mathfrak{F}' = \mathfrak{F} \setminus \{x\}$ and let x be strongly unacceptable w.r.t. $y \in A$, i.e., $cl(x) = cl(y)$ and $vul(y) \subseteq vul(x)$. Observe that $E \in cf(\mathfrak{F})$ iff $E \in cf(\mathfrak{F}')$ for every E with $x \notin E$; moreover, x does not belong to any admissible extension of $\mathfrak{F}$ and $\mathfrak{F}'$ since x cannot be defended against y without being attacked by its defender (using $vul(y) \subset vul(x)$). We obtain that x is either attacked by an admissible set or undecided. If y is contained in a stable extension, x is defeated; in case y is not contained in a stable extension, y is attacked and thus also x is attacked using $vul(y) \subset vul(x)$. Consequently, the argument x can be safely removed without changing the stable extensions of $\mathfrak{F}$. ☐

Considering grounded, complete, and preferred semantics, we observe that strongly unacceptable arguments are not necessarily defeated – removing them thus potentially results in a change of the σ_i-extensions.

Example 8.4.8. *Consider cvAF $\mathfrak{F}$ from Example 8.3.7 and a new argument $x_0 = (\{r\}, t)$:*



$f_e(\mathfrak{F}, x_0)$ has three complete conclusion-extensions: $\emptyset$ (the grounded extension), $\{s, p, t\}$, and $\{q, t\}$. Recall that x_1 is strongly unacceptable w.r.t. x_2. Removing x_1 would make x_0 unattacked, changing the grounded extension to $\{t\}$.

Strongly unacceptable arguments can neither be enforced nor deleted in such situations. This means that on semantics level, it is not possible to distinguish if such arguments are self-attacking or not. We show this by proving that the semantics of the cvAF remain unchanged after turning x into a self-attacker. Formally, this is achieved by removing it and expanding the resulting cvAF with some argument x' which is analogous to x, except having also its claim as vulnerability; formally, $x' = (vul(x) \cup \{cl(x)\}, cl(x))$.

Proposition 8.4.9. *For a cvAF* $\mathfrak{F} = (A, R)$*, a semantics* $\sigma \in \{gr, co, pr, stb\}$*, and a strongly unacceptable argument* $x \in A$*, it holds that* $\sigma_i(\mathfrak{F}) = \sigma_i(f_e(\mathfrak{F} \setminus \{x\}, x'))$ *for* $x' = (vul(x) \cup \{cl(x)\}, cl(x))$*.*

Proof. Let $\mathfrak{F}' = f_e(\mathfrak{F} \setminus \{x\}, x')$ and assume x is strongly unacceptable w.r.t. $y \in A$. As outlined in the proof of Proposition 8.4.7, x can never appear in an admissible extension. We moreover observe that $E^+_F = E^+_{F'}$ for every conflict-free set E, $y \notin E$, since (x, x) is the only attack which has been introduced. We thus obtain $ad_i(\mathfrak{F}) = ad_i(\mathfrak{F}')$. Moreover, the grounded extension is preserved by adding this self-attack since it does not remove nor introduce new unattacked arguments (or any arguments defended by them). We thus obtain $\sigma_i(\mathfrak{F}) = \sigma_i(\mathfrak{F}')$ for $\sigma \in \{co, gr, pr, stb\}$. $\square$

8.4.1 Complete Kernel for cvAFs

We are ready to consider our first cvAF kernel. Following Proposition 8.4.9, the first adjustment we carry out is a modification on vulnerability level: Each strongly unacceptable argument x is turned into a self-attacker by adding $cl(x)$ to $vul(x)$. In the next step, we remove all strongly defeated and redundant arguments.

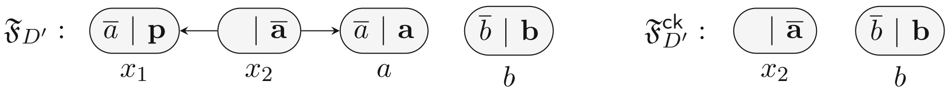
Definition 8.4.10. *For a cvAF* $\mathfrak{F} = (A, R)$*, let* X *denote the set of all strongly unacceptable arguments in* A *and let*

$$(A', R') = f_e(\mathfrak{F} \setminus X, \{(vul(x) \cup \{cl(x)\}, cl(x)) \mid x \in X\}).$$

We define the complete kernel $\mathfrak{F}^{\mathsf{ck}} = (A^{\mathsf{ck}}, R^{\mathsf{ck}})$ *with*

$$A^{\mathsf{ck}} = A' \setminus \{x \in A' \mid x \text{ is strongly defeated or redundant}\},$$
$$R^{\mathsf{ck}} = R' \cap (A^{\mathsf{ck}} \times A^{\mathsf{ck}}).$$

Example 8.4.11. *The cvAF* $\mathfrak{F}_D$ *from our running example coincides with its complete kernel since no arguments are strongly defeated, unacceptable or redundant. That is, we obtain* $\mathfrak{F}^{\mathsf{ck}}_D = \mathfrak{F}^{\mathsf{gk}}_D = \mathfrak{F}_D$*. For* $\mathfrak{F}_{D'}$*, we obtain the following picture:*



Proposition 8.4.12. $\mathfrak{F} \equiv^\sigma_s \mathfrak{F}^{\mathsf{ck}}$ *for every cvAF* $\mathfrak{F}$ *and for* $\sigma \in \{co, gr, pr, stb\}$*.*

Proof. Consider a set of instantiated arguments X. First, by Proposition 8.4.9, we can modify all strongly unacceptable arguments of $\mathfrak{F}$ without changing semantics. Let $A_{unac} \subseteq A$ denote the set of unacceptable arguments in $\mathfrak{F}$. We obtain $\sigma_i(f_e(\mathfrak{F}', X)) = \sigma_i(f_e(\mathfrak{F}, X))$ for $\mathfrak{F}' = (A', R') = (\mathfrak{F} \setminus A_{unac}) \cup \{(\textsc{Vul}(x) \cup \{cl(x)\}, cl(x)) \mid x \in A_{unac}\}$. By Proposition 8.4.5 and 8.4.6, we can delete redundant and strongly unacceptable arguments as well. Let $A_{red} \subseteq A'$ and $A_{sdef} \subseteq A'$ denote the set of redundant and strongly defeated arguments of $\mathfrak{F}'$, respectively. We obtain $\sigma_i(f_e(\mathfrak{F}'', X)) = \sigma_i(f_e(\mathfrak{F}, X))$ for $\mathfrak{F}'' = \mathfrak{F} \setminus (A_{red} \cup A_{sdef})$. By definition of the complete kernel, it holds that $\mathfrak{F}'' = \mathfrak{F}^{\mathsf{ck}}$. We obtain $\sigma_i(f_e(\mathfrak{F}^{\mathsf{ck}}, X)) = \sigma_i(f_e(\mathfrak{F}, X))$, hence $\mathfrak{F} \equiv_s^{co} \mathfrak{F}^{\mathsf{ck}}$. □

Corollary 8.4.13. *$\sigma_i(\mathfrak{F}) = \sigma_i(\mathfrak{F}^{\mathsf{ck}})$ for every cvAF $\mathfrak{F}$ and for $\sigma \in \{co, gr, pr, stb\}$.*

Next we show that kernelization behaves as expected: the complete kernel does neither contain redundant nor strongly defeated arguments; and each strongly unacceptable argument is self-attacking. For this, we consider the syntactical effects of our modifications.

Observation 8.4.14. *Removing arguments from a given cvAF $\mathfrak{F}$ does not add novel redundant, strongly unacceptable, or strongly defeated arguments.*

We show that the modification of unacceptable arguments can be done iteratively.

Lemma 8.4.15. *Given a cvAF $\mathfrak{F} = (A, R)$ and a strongly unacceptable argument $x \in A$. Let $x' = (vul(x) \cup \{cl(x)\}, cl(x))$ and let $\mathfrak{F}' = f_e(\mathfrak{F} \setminus \{x\}, x') = (A', R')$. Then, for all $y \neq x \in A$, y is strongly unacceptable in $\mathfrak{F}$ iff y is strongly unacceptable in $\mathfrak{F}'$.*

Proof. Consider a strongly unacceptable argument $y \in A$ in $\mathfrak{F}$. Then there is $z \in A$ with $vul(z) \subseteq vul(y)$ and $(z, y) \in R$ in $\mathfrak{F}$. First assume $z \neq x$. Then it holds that $z \in A'$, witnessing unacceptability of y in $\mathfrak{F}'$. Otherwise, in case $z = x$, there is $z' \in A$ with $vul(z') \subseteq vul(x) = vul(z)$ such that $(z', x) \in R$. Consequently, $(z', y) \in R$ using $vul(z) \subseteq vul(y)$, showing that y is strongly unacceptable in $\mathfrak{F}'$. For the other direction, consider a strongly unacceptable argument $y \in A'$ in $\mathfrak{F}'$. There is a witness $z \in A'$ of strong unacceptability of y in $\mathfrak{F}'$. By construction, z is also a witness in $\mathfrak{F}$. □

We observe that we might obtain novel redundant arguments when turning unacceptable arguments into self-attacker. Let x be an argument with $vul(x) = \{c, d\}$, and let y be a strongly unacceptable argument with claim $cl(y) = c$ which is attacked by claims $vul(y) = \{d, e\}$ in $\mathfrak{F}$. Turning y into a self-attacker thus makes x redundant in $\mathfrak{F}'$.

Lemma 8.4.16. *Given a cvAF $\mathfrak{F}$ and arguments $x, y \in A$, $x \neq y$. Let y be redundant/strongly defeated in $\mathfrak{F}$. Then x is redundant or strongly defeated in $\mathfrak{F}$ iff x is redundant or strongly defeated in $\mathfrak{F} \setminus \{y\}$.*

Proof. In case x is redundant or strongly defeated in $\mathfrak{F} \setminus \{y\}$ then there is a witness z in $\mathfrak{F} \setminus \{y\}$. As mentioned in Remark 8.4.14, the claim-attacks are not affected by removing certain arguments. We thus obtain that z witnesses that x is redundant or strongly defeated in $\mathfrak{F}$. Also, in case x is strongly defeated in $\mathfrak{F}$, it is clear that x is contained in $\mathfrak{F} \setminus \{y\}$ since y cannot serve as witness of x being strongly defeated since $vul(y) \neq \emptyset$.

Now, let y be strongly defeated in $\mathfrak{F}$. In case x is redundant w.r.t. y in $\mathfrak{F}$, there is some $z \in A$ with $(z, y) \in R$. We obtain x is strongly defeated (using $vul(y) \subseteq vul(x)$).

Let y be redundant in $\mathfrak{F}$ and let x be redundant w.r.t. y in $\mathfrak{F}$. Then there is $z \in A$ with $vul(z) \subseteq vul(y)$ and $cl(z) = cl(y)$, thus witnessing redundancy of x. □

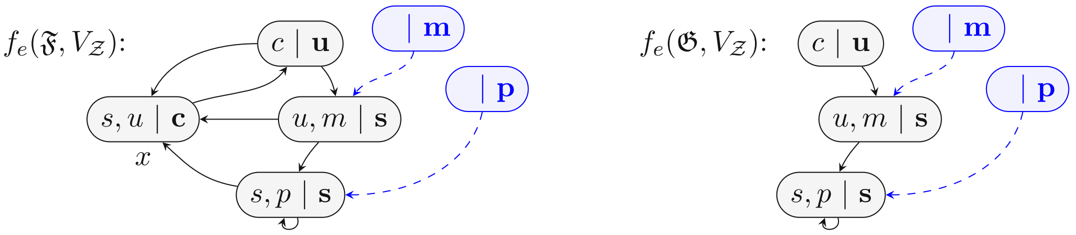


Figure 8.2: Illustration of Case 1 in the proof of Lemma 8.4.18 with $V_{\mathcal{Z}} = \{(\emptyset, m), (\emptyset, p)\}$.

Proposition 8.4.17. *The complete kernel* $\mathfrak{F}^{\mathsf{ck}}$ *of a cvAF* $\mathfrak{F}$ *does neither contain redundant nor strongly defeated arguments, and each strongly unacceptable argument is self-attacking.*

Proof. We first modify strongly unacceptable arguments. By Lemma 8.4.15, the modification does not add novel strongly unacceptable arguments, thus this procedure can be done iteratively and it is guaranteed that each strongly unacceptable argument is self-attacking after this modification. Next, we iteratively delete redundant and strongly defeated arguments. By Observation 8.4.14, the deletion of arguments does not introduce novel strongly unacceptable, redundant, or strongly defeated arguments. Moreover, by Lemma 8.4.16, redundant and strongly defeated arguments can be removed without producing novel redundant or strongly defeated arguments. □

We show that complete kernels of strongly equivalent cvAFs contain the same claims.

Lemma 8.4.18. *For two cvAFs* $\mathfrak{F}$ *and* $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{co} \mathfrak{G}$ *implies* $cl(A_{\mathfrak{F}^{\mathsf{ck}}}) = cl(A_{\mathfrak{G}^{\mathsf{ck}}})$.

Proof. Consider an argument $x \in A_{\mathfrak{F}^{\mathsf{ck}}}$ with claim $cl(x) = c$. Towards a contradiction, assume that there is no argument $y \in A_{\mathfrak{G}^{\mathsf{ck}}}$ with $cl(y) = c$. We may assume $co_i(\mathfrak{F}^{\mathsf{ck}}) = co_i(\mathfrak{G}^{\mathsf{ck}})$, hence we deduce that x does not occur in any complete extension of $\mathfrak{F}^{\mathsf{ck}}$. Hence it does not occur in any admissible extension. Consequently, x receives incoming attacks.

Case 1 Suppose x is no self-attacker. The overall idea is as follows: We construct a set of instantiated arguments X in order to deal with all arguments that attack x. We introduce isolated arguments attacking (most of) them; this is possible due to our definition of the kernel. Then $f_e(\mathfrak{F}^{\mathsf{ck}}, X)$ has an admissible extension containing the argument x with claim c, where in $\mathfrak{G}^{\mathsf{ck}}$ claim c does not occur at all. Consider the set

$$\mathcal{Z} = \{z \in A_{\mathfrak{F}^{\mathsf{ck}}} \mid (z, x) \in R_{\mathfrak{F}}\}$$

of arguments attacking x. Since x is no self-attacker, we have $vul(z) \not\subseteq vul(x)$, i.e., $vul(z) \setminus vul(x) \neq \emptyset$ for each $z \in \mathcal{Z}$ (otherwise, $vul(z) \subseteq vul(x)$ and $(z, x) \in R$ implies that x is strongly unacceptable, hence x would be self-attacking in the kernel). We let

$$V_{\mathcal{Z}} = \{v_e = (\emptyset, e) \mid e \in vul(z) \setminus vul(x), z \in \mathcal{Z}, e \neq c\},$$

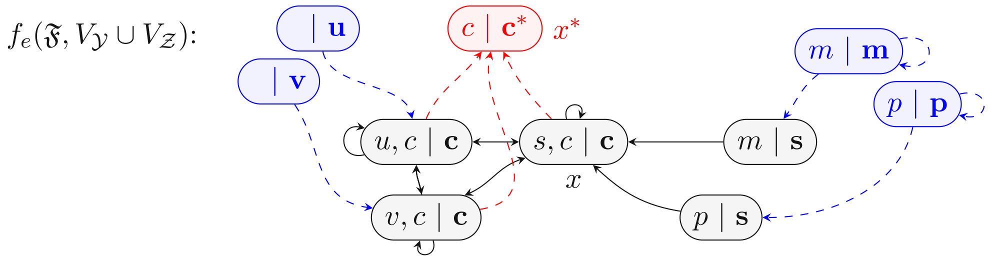


Figure 8.3: Illustration of Case 2.1 in the proof of Lemma 8.4.18. Novel arguments are in color with dashed attacks; left we depict arguments with claim c, i.e., the set $\mathcal{Y}$, and the novel arguments $V_{\mathcal{Y}}$ defeating them; right, we depict arguments attacking x and the novel self-attacking arguments which attack them. The novel argument x^* (in red) is undecided in the cvAF $\mathfrak{F}$ and unattacked (hence accepted) in the cvAF $\mathfrak{G}$.

i.e., we defeat these attackers as long as this would not require introducing claim c. Having c as claim, x can now defend itself, i.e., $\{x\} \cup V_{\mathcal{Z}}$ is admissible in the obtained cvAF. See Figure 8.2 for an example of the construction.

Since c does not occur in $\mathfrak{G}^{\mathsf{ck}}$ this is a witness for the absence of strong equivalence.

Case 2 Now suppose each argument with claim c is a self-attacker and fix such x. Since x occurs in the kernel $\mathfrak{F}^{\mathsf{ck}}$, each attacker of x must itself possess attacking arguments.

The first step is to get rid of arguments with the same claim c. Consider the set

$$\mathcal{Y} = \{y \in A_{\mathfrak{F}^{\mathsf{ck}}} \mid cl(y) = c, y \neq x\}$$

of arguments with claim c. We consider arguments which defeat them; i.e., we let

$$\begin{aligned} V_{\mathcal{Y}} &= \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}, e \neq c\} \\ &= \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}\}. \end{aligned}$$

Now consider the set $\mathcal{Z} = \{z \in A_{\mathfrak{F}^{\mathsf{ck}}} \mid (z, x) \in R_{\mathfrak{F}^{\mathsf{ck}}}\} \setminus \mathcal{Y}$ of arguments attacking x. We introduce self-attacking arguments that attack (most of) the arguments $z \in \mathcal{Z}$:

$$V_{\mathcal{Z}} = \{v_e = (\{e\}, e) \mid e \in vul(z), z \in \mathcal{Z}, e \neq c\}.$$

This ensures that all $z \in \mathcal{Z}$ with $vul(z) \neq \{c\}$ are undecided in the resulting cvAF.

Case 2.1: Suppose there is no argument z attacking x with $x \neq z$ and $vul(z) = \{c\}$, i.e., if $(z, x) \in R_{\mathfrak{F}^{\mathsf{ck}}}$, then $vul(z) \setminus \{c\} \neq \emptyset$. Hence introducing a self-attacker for each claim except c as done before ensures that x is undecided in each admissible extension; moreover, bear in mind that there is no other realization of c left after introducing $V_{\mathcal{Y}}$.

Now, consider some fresh argument $x^*_{c^*} = (\{c\}, c^*)$ with novel claim c^* which is attacked by c. This way, we ensure that $x^*_{c^*}$ is attacked by the (always undecided) self-attacker x in $f_e(\mathfrak{F}^{\mathsf{ck}}, X)$, but unattacked in $f_e(\mathfrak{G}^{\mathsf{ck}}, X)$. See Figure 8.3 for an illustrative example.

$f_e(\mathfrak{F}, V_{\mathcal{Y}} \cup V_{\mathcal{Z}})$:

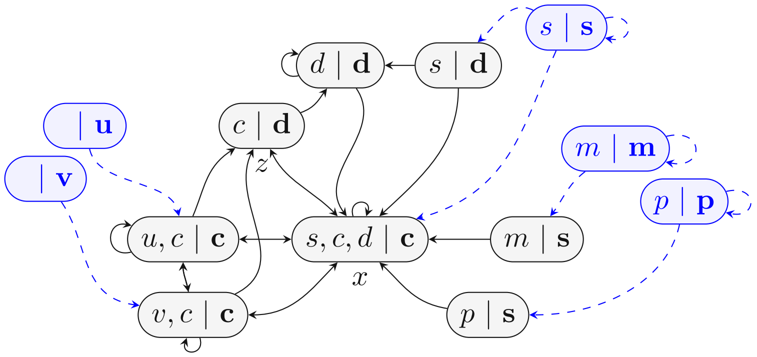


Figure 8.4: Illustration of Case 2.2 in the proof of Lemma 8.4.18. Novel arguments are in blue with dashed attacks. The argument $(\{c\}, d)$ is not grounded in the expansion of $\mathfrak{F}$ but unattacked (thus grounded) in the expansion of $\mathfrak{G}$.

Case 2.2: Suppose there is some $z \in \mathcal{Z}$ attacking x with $x \neq z$ and $vul(z) = \{c\}$.

Suppose $cl(z) = d$ and consider

$$\mathcal{Y}_z = \{y \in A_{\mathfrak{F}^{\mathsf{ck}}} \mid cl(y) = d = cl(z)\}$$

Observe that $\mathcal{Y}_z \subseteq \mathcal{Z}$ (since $d \in vul(x)$ by assumption z attacks x). Hence for each $y \in \mathcal{Y}_z$, for each vulnerability $e \in \textsc{Vul}(y)$ with $e \neq c$, we have introduced self-attacking arguments $(\{e\}, e)$ which attack y on e. Hence $z = (\{c\}, d)$ is the only argument with claim d which is not undecided (i.e., attacked by self-attacking arguments) in $f_e(\mathfrak{F}^{\mathsf{ck}}, V_{\mathcal{Y}} \cup V_{\mathcal{Z}})$. Hence there is no argument with claim d which is contained in the grounded extension of $f_e(\mathfrak{F}^{\mathsf{ck}}, V_{\mathcal{Y}} \cup V_{\mathcal{Z}})$. For an example of a cvAF $\mathfrak{F}$ expanded by $V_{\mathcal{Y}} \cup V_{\mathcal{Z}}$ see Figure 8.4.

In $f_e(\mathfrak{G}^{\mathsf{ck}}, V_{\mathcal{Y}} \cup V_{\mathcal{Z}})$, on the other hand, the argument z is unattacked and thus contained in the grounded extension. ☐

Theorem 8.4.19. *For two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{co} \mathfrak{G}$ iff $\mathfrak{F}^{\mathsf{ck}} = \mathfrak{G}^{\mathsf{ck}}$.*

Proof. First assume $\mathfrak{F}^{\mathsf{ck}} = \mathfrak{G}^{\mathsf{ck}}$ holds. By Proposition 8.4.12, it holds that $\mathfrak{F}^{\mathsf{ck}} \equiv_s^{co} \mathfrak{F}$ and $\mathfrak{G}^{\mathsf{ck}} \equiv_s^{co} \mathfrak{G}$. Thus we obtain $\mathfrak{F} \equiv_s^{co} \mathfrak{G}$ by transitivity.

For the other direction, assume $\mathfrak{F} \equiv_s^{co} \mathfrak{G}$. We show that in this case, the kernels of $\mathfrak{F}$ and $\mathfrak{G}$ coincide. It suffices to show that they contain the same arguments, that is, we show that for all $x \in A_{\mathfrak{F}^{\mathsf{ck}}}$ there is $y \in A_{\mathfrak{G}^{\mathsf{ck}}}$ with $cl(y) = cl(x)$ and $vul(y) = vul(x)$.

By Lemma 8.4.18, $\mathfrak{F}^{\mathsf{ck}}$ and $\mathfrak{G}^{\mathsf{ck}}$ contain the same claims. We show that *for all arguments x in $\mathfrak{F}^{\mathsf{ck}}$ there is some argument y in $\mathfrak{G}^{\mathsf{ck}}$ such that $cl(x) = cl(y) = c$ and $vul(y) \subseteq vul(x)$.*

Let $x \in A_{\mathfrak{F}^{\mathsf{ck}}}$ with $cl(x) = c$. Then there is some argument y with claim c in $\mathfrak{G}^{\mathsf{ck}}$. Towards a contradiction, assume that for all $y \in A_{\mathfrak{G}^{\mathsf{ck}}}$ with $cl(y) = c$ we have $vul(y) \nsubseteq vul(x)$. Let $\mathcal{Y} = \{y \in A_{\mathfrak{G}^{\mathsf{ck}}} \mid cl(y) = c\}$ denote all arguments with claim c in $A_{\mathfrak{G}^{\mathsf{ck}}}$. Then for all

$y \in \mathcal{Y}$ there is a claim $e \in vul(y)$ with $e \notin vul(x)$. We introduce arguments

$$V_{\mathcal{Y}} = \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}, e \neq c\}$$

in order to defeat all arguments in $\mathfrak{G}^{\mathsf{ck}}$ with claim c without introducing a novel argument with claim c. Now, let $\mathfrak{F}' = f_e(\mathfrak{F}^{\mathsf{ck}}, V_{\mathcal{Y}})$ and $\mathfrak{G}' = f_e(\mathfrak{G}^{\mathsf{ck}}, V_{\mathcal{Y}})$.

Case 1 Suppose $c \in vul(x)$, i.e., x is self-attacking. Then each argument with claim c in $\mathfrak{G}^{\mathsf{ck}}$ is attacked by arguments in $V_{\mathcal{Y}}$. The cvAF $\mathfrak{G}'$ has no argument with claim c since all such arguments are strongly defeated by $V_{\mathcal{Y}}$. On the other hand, x is contained in the kernel of $\mathfrak{F}'$. By Lemma 8.4.18, $\mathfrak{F}'$ and $\mathfrak{G}'$ are not strongly equivalent to each other, contradiction to our assumption.

Case 2 Now assume x is not self-attacking. In this case, $\mathfrak{G}'$ might still contain a single argument y with claim c and $vul(y) = vul(x) \cup \{c\}$. Thus the conclusion c does not appear in any conflict-free extension of $(\mathfrak{G}')^{\mathsf{ck}}$. We proceed analogous as in the proof of Lemma 8.4.18, Case 1, and introduce arguments to defend x in $(\mathfrak{F}')^{\mathsf{ck}}$ in order to guarantee that x appears in an admissible extension in the resulting cvAF. Then $\mathfrak{F}^{\mathsf{ck}}$ and $\mathfrak{G}^{\mathsf{ck}}$ do not yield the same admissible extensions after expansion.

We obtain that for every argument $x \in A_{\mathfrak{F}^{\mathsf{ck}}}$ there is exactly one argument $y \in A_{\mathfrak{G}^{\mathsf{ck}}}$ such that $cl(x) = cl(y)$ and $vul(x) = vul(y)$: Consider an argument $y \in A_{\mathfrak{G}^{\mathsf{ck}}}$ such that $cl(x) = cl(y) = c$ and $vul(x) \supseteq vul(y)$. By symmetry, there is $z \in A_{\mathfrak{F}^{\mathsf{ck}}}$ with $cl(z) = c$ such that $vul(y) \supseteq vul(z)$. Thus $vul(x) \supseteq vul(y) \supseteq vul(z)$. Since $\mathfrak{F}^{\mathsf{ck}}$ is redundancy-free, we obtain $vul(x) = vul(y) = vul(z)$; by assumption $\mathfrak{F}^{\mathsf{ck}}$, $\mathfrak{G}^{\mathsf{ck}}$ do not contain equivalent arguments, we conclude $x = z$ (by well-formedness, x and z attack the same arguments and are thus equivalent).

We thus obtain that $\mathfrak{F}^{\mathsf{ck}}$ and $\mathfrak{G}^{\mathsf{ck}}$ contain the same arguments in case $\mathfrak{F}$ and $\mathfrak{G}$ are strongly equivalent w.r.t. complete semantics. Since all attacks in cvAFs are determined by the claims and vulnerabilities of the arguments they contain, we thus conclude $\mathfrak{F}^{\mathsf{ck}} = \mathfrak{G}^{\mathsf{ck}}$. ☐

8.4.2 Preferred Kernel for cvAFs

Towards a kernel for preferred semantics, we consider a special case of strong unacceptability that affects only preferred semantics.

Definition 8.4.20. *For a cvAF $\mathfrak{F} = (A, R)$, $x \in A$ is* strongly pr-unacceptable *if x is strongly unacceptable w.r.t. $y \in A$ and $vul(y) = \{cl(x)\}$.*

Note that each strongly pr-unacceptable argument is strongly unacceptable; also, each strongly pr-unacceptable argument is self-attacking because $vul(y) = \{cl(x)\} \subseteq vul(x)$. It turns out that such arguments can be removed without affecting preferred semantics.

Proposition 8.4.21. *For a cvAF $\mathfrak{F} = (A, R)$ and a strongly pr-unacceptable argument $x \in A$, $pr_i(\mathfrak{F}) = pr_i(\mathfrak{F} \setminus \{x\})$.*

Proof. Let $\mathfrak{F}' = f_e(\mathfrak{F} \setminus \{x\}, x')$ and recall that x can never appear in an admissible extension as it it strongly unacceptable. Let x be strongly pr-unacceptable w.r.t. $y \in A$. Then $\Gamma_{\mathfrak{F}}(\{y\}) \subseteq \Gamma_{\mathfrak{F}'}(\{y\})$ for every $z \in A \setminus \{x\}$, i.e., every argument $z \neq x$ defends the same arguments in $\mathfrak{F}'$ which are defended by z in $\mathfrak{F}$. We obtain $ad_i(\mathfrak{F}) \subseteq ad_i(\mathfrak{F}')$.

To prove $pr_i(\mathfrak{F}) = pr_i(\mathfrak{F}')$ we show that for every $E \in ad(\mathfrak{F}')$, there is $D \in ad(\mathfrak{F})$ such that $E \subseteq D$. In case $E \in ad(\mathfrak{F})$, we are done (taking $D = E$). In case $E \notin ad(\mathfrak{F})$, there is $z \in E$ such that $(x, z) \in R$ and z is not defended by E in F. In case $E \cup \{y\} \in cf(\mathfrak{F})$ we are done (not that in this case, $D = E \cup \{y\}$ is admissible). Now assume $E \cup \{y\}$ is not conflict-free. Observe that $(y, y) \notin R$ by assumption $vul(y) = \{cl(x)\}$. In case there is $v \in E$ such that $(v, y) \in R$ we have $cl(v) = c$ and thus $(v, x) \in R$ by well-formedness, contradiction to $E \notin ad(\mathfrak{F})$. In case $(y, v) \in R$ for some $v \in E$ we have some $w \in E$ which defends v against w (since E is admissible in $\mathfrak{F}'$) thus we arrive again at a contradiction since $(w, y) \in R$ implies $(w, x) \in R$. It follows that $D = E \cup \{y\}$ is an admissible superset of E in $\mathfrak{F}$. We have shown that the preferred extensions of $\mathfrak{F}$ and $\mathfrak{F}'$ coincide. ☐

The preferred kernel refines the complete kernel:

Definition 8.4.22. *For a cvAF* $\mathfrak{F} = (A, R)$, *let* $\mathfrak{F}^{\mathsf{ck}} = (A^{\mathsf{ck}}, R^{\mathsf{ck}})$ *be as in Definition 8.4.10. We define the* preferred kernel $\mathfrak{F}^{\mathsf{pk}} = (A^{\mathsf{pk}}, R^{\mathsf{pk}})$ *with*

$$A^{\mathsf{pk}} = A^{\mathsf{ck}} \setminus \{x \in A^{\mathsf{ck}} \mid x \textit{ is strongly } pr\textit{-unacceptable}\},$$
$$R^{\mathsf{pk}} = R^{\mathsf{ck}} \cap (A^{\mathsf{pk}} \times A^{\mathsf{pk}}).$$

For our running example cvAFs $\mathfrak{F}_D$ and $\mathfrak{F}_{D'}$, it holds that $\mathfrak{F}_D^{\mathsf{pk}} = \mathfrak{F}_D^{\mathsf{ck}}$ and $\mathfrak{F}_{D'}^{\mathsf{pk}} = \mathfrak{F}_{D'}^{\mathsf{ck}}$.

Proposition 8.4.23. $\mathfrak{F} \equiv_s^{pr} \mathfrak{F}^{\mathsf{ck}}$ *for every cvAF* $\mathfrak{F}$.

Proof. Consider a set of instantiated arguments X. By Proposition 8.4.12, we obtain $pr_i(f_e(\mathfrak{F}^{\mathsf{ck}}, X)) = pr_i(f_e(\mathfrak{F}, X))$. Let Let $A_{punac} \subseteq A^{\mathsf{ck}}$ denote the set of strongly pr-unacceptable arguments of $\mathfrak{F}^{\mathsf{ck}}$. By Proposition 8.4.21, we can delete strongly pr-unacceptable arguments iteratively without changing preferred extensions. We obtain $pr_i(f_e(\mathfrak{F}', X)) = pr_i(f_e(\mathfrak{F}, X))$ for $\mathfrak{F}' = \mathfrak{F} \setminus A_{punac}$. By definition of the preferred kernel, it holds that $\mathfrak{F}' = \mathfrak{F}^{\mathsf{pk}}$. Hence we obtain $\mathfrak{F} \equiv_s^{pr} \mathfrak{F}^{\mathsf{pk}}$. ☐

Corollary 8.4.24. $pr_i(\mathfrak{F}) = pr_i(\mathfrak{F}^{\mathsf{ck}})$ *for every cvAF* $\mathfrak{F}$.

We show that the preferred kernel does not contain redundant, strongly defeated, and strongly pr-unacceptable arguments; moreover, each strongly unacceptable argument is self-attacking. The latter follows by Lemma 8.4.15. It remains to show that redundant, strongly defeated, and strongly pr-unacceptable arguments can be removed iteratively.

Lemma 8.4.25. *Given a cvAF* $\mathfrak{F}$ *and arguments* $x, y \in A$, $x \neq y$. *Let* y *be redundant/strongly defeated/strongly* pr*-unacceptable in* $\mathfrak{F}$. *Then* x *is redundant, strongly defeated, or strongly* pr*-unacceptable in* $\mathfrak{F}$ *iff* x *is redundant, strongly defeated, or strongly* pr*-unacceptable in* $\mathfrak{F} \setminus \{y\}$.

Proof. First observe that if x is redundant, strongly defeated, or strongly pr-unacceptable in $\mathfrak{F} \setminus \{y\}$ then there is a witness z in $\mathfrak{F} \setminus \{y\}$. As mentioned in Remark 8.4.14, the claim-attacks are not affected by removing certain arguments. We thus obtain that z witnesses that x is redundant, strongly defeated, or strongly pr-unacceptable in $\mathfrak{F}$. Also, in case x is strongly defeated in $\mathfrak{F}$, it is clear that x is contained in $\mathfrak{F} \setminus \{y\}$ since y is not unattacked and thus cannot witness that x is is strongly defeated.

Let y be strongly defeated in $\mathfrak{F}$. In case x is redundant w.r.t. y in $\mathfrak{F}$, there is some unattacked $z \in A$ with $(z, y) \in R$. Thus we obtain that also x is strongly defeated (using $vul(y) \subseteq vul(x)$, i.e., $(z, x) \in R$). In case x is strongly pr-unacceptable w.r.t. y in $\mathfrak{F}$, there is some unattacked $z \in A$, $(z, y) \in R$, moreover, $cl(z) = cl(x)$ (using $vul(y) = \{cl(x)\}$. Consequently we obtain that x is redundant in $\mathfrak{F}$ and in $\mathfrak{F} \setminus \{y\}$.

Let y be redundant in $\mathfrak{F}$ and let x be redundant w.r.t. y in $\mathfrak{F}$. Then there is $z \in A$ with $vul(z) \subseteq vul(y)$ and $cl(z) = cl(y)$, thus witnessing redundancy of x. In case x is strongly pr-unacceptable w.r.t. y in $\mathfrak{F}$, there is $z \in A$ with $cl(z) = cl(y)$ and $vul(z) \subset vul(y) = \{cl(x)\}$, thus $vul(z) = \emptyset$; moreover, $(z, x) \in R$ using $cl(z) = cl(y) \in vul(x)$. We obtain that x is strongly defeated. □

By Lemma 8.4.15, Observation 8.4.14, and Lemma 8.4.25, we obtain the following result.

Proposition 8.4.26. *For any cvAF $\mathfrak{F}$, the kernel $\mathfrak{F}^{\mathsf{pk}}$ does neither contain redundant, non-self-attacking strongly unacceptable, strongly defeated nor pr-unacceptable arguments.*

We show that preferred kernels of two strongly equivalent cvAFs contain the same claims.

Lemma 8.4.27. *For two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{pr} \mathfrak{G}$ implies $cl(A_{\mathfrak{F}^{\mathsf{pk}}}) = cl(A_{\mathfrak{G}^{\mathsf{pk}}})$.*

Proof. Let $x \in A_{\mathfrak{F}^{\mathsf{pk}}}$ with $cl(x) = c$. Towards a contradiction, assume that there is no argument $y \in A_{\mathfrak{G}^{\mathsf{pk}}}$ with $cl(y) = c$. Since we may assume $pr_{cl}(\mathfrak{F}^{\mathsf{pk}}) = pr_{cl}(\mathfrak{G}^{\mathsf{pk}})$ we have x does not occur in any preferred extension of $\mathfrak{F}^{\mathsf{pk}}$. Hence it does not occur in any admissible extension. Consequently, x receives incoming attacks. We proceed similar as in the proof of Lemma 8.4.18.

Case 1 Suppose x is no self-attacker. This case is analogous to the proof of Lemma 8.4.18.

Case 2 Now suppose each argument with claim c is a self-attacker and fix such x. Since x occurs in the kernel $\mathfrak{F}^{\mathsf{pk}}$, each attacker of x must itself possess attacking arguments. This case is analogous to Case 2.1 in the proof of Lemma 8.4.18. Since the preferred kernel does not contain pr-unacceptable arguments, it holds that each attacker z of x contains some vulnerability $e \in vul(z)$ with $e \neq c$. Hence a case analogous to Case 2.2 in the proof of Lemma 8.4.18 can never occur. □

Theorem 8.4.28. *For two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{pr} \mathfrak{G}$ iff $\mathfrak{F}^{\mathsf{pk}} = \mathfrak{G}^{\mathsf{pk}}$.*

Proof. Analogous to the proof of Theorem 8.4.19, we first assume $\mathfrak{F}^{\mathsf{pk}} = \mathfrak{G}^{\mathsf{pk}}$. By Proposition 8.4.23, it holds that $\mathfrak{F}^{\mathsf{pk}} \equiv_s^{pr} \mathfrak{F}$ and $\mathfrak{G}^{\mathsf{pk}} \equiv_s^{pr} \mathfrak{G}$. Thus we obtain $\mathfrak{F} \equiv_s^{pr} \mathfrak{G}$.

For the other direction, assume $\mathfrak{F} \equiv_s^{pr} \mathfrak{G}$. By Lemma 8.4.27, $\mathfrak{F}^{\mathsf{pk}}$ and $\mathfrak{G}^{\mathsf{pk}}$ contain the same claims. Analogous to the proof of Theorem 8.4.19, it can be shown that for all arguments x in $\mathfrak{F}^{\mathsf{pk}}$ there is some argument y in $\mathfrak{G}^{\mathsf{pk}}$ such that $cl(x) = cl(y) = c$ and $vul(y) \subseteq vul(x)$. Hence $\mathfrak{F}^{\mathsf{pk}}$ and $\mathfrak{G}^{\mathsf{pk}}$ contain the same arguments. We obtain $\mathfrak{F}^{\mathsf{pk}} = \mathfrak{G}^{\mathsf{pk}}$. ☐

8.4.3 Grounded Kernel for cvAFs

Next we consider the case for grounded semantics. As we have demonstrated in the scope of our enforcement results, grounded semantics give rise to a more general notion of strong unacceptability. We show that all strongly gr-unacceptable arguments can be turned into self-attacker without affecting grounded semantics.

Proposition 8.4.29. *Given a cvAF* $\mathfrak{F} = (A, R)$ *and a strongly* gr*-unacceptable argument* $x \in A$ *and let* $x' = (vul(x) \cup \{cl(x)\}, cl(x))$*. Then* $gr(\mathfrak{F}) = gr((f_e(\mathfrak{F} \setminus \{x\}, x'))$*.*

Proof. Let $\mathfrak{F}' = f_e(\mathfrak{F} \setminus \{x\}, x')$ and assume x is strongly gr-unacceptable w.r.t. $y \in A$. In case $x \notin gr(\mathfrak{F})$ we are done (turning x into a self-attacking argument does not change the grounded extension). In case $x \in gr(\mathfrak{F})$ there is $z \in gr(\mathfrak{F})$ such that $(z, y) \in R$. If $cl(z) \neq cl(x)$ we have $cl(z) \in vul(x)$ by assumption $vul(y) \setminus \{cl(x)\} \subseteq vul(x)$, that is, z attacks x, contradiction to $\{x, z\} \subseteq gr(\mathfrak{F})$. In case $cl(z) = cl(x)$, we have $cl(x) \in gr_i(\mathfrak{F}')$, and z attacks the same arguments as x by well-formedness, hence $gr_{cl}(\mathfrak{F}) = gr_i(\mathfrak{F}')$. ☐

The *grounded kernel* is defined analogously to the complete kernel by replacing X with the set of all strongly gr-unacceptable arguments in A.

Definition 8.4.30. *For a cvAF* $\mathfrak{F} = (A, R)$*, let* X *denote the set of all strongly* gr*-unacceptable arguments in* A *and let*

$$(A', R') = f_e(\mathfrak{F} \setminus X, \{(vul(x) \cup \{cl(x)\}, cl(x)) \mid x \in X\}).$$

We define the grounded kernel $\mathfrak{F}^{\mathsf{gk}} = (A^{\mathsf{gk}}, R^{\mathsf{gk}})$ *with*

$$A^{\mathsf{gk}} = A' \setminus \{x \in A' \mid x \textit{ is strongly defeated or redundant}\},$$

and $R^{\mathsf{gk}} = R' \cap (A^{\mathsf{gk}} \times A^{\mathsf{gk}})$*.*

Proposition 8.4.31. $\mathfrak{F} \equiv_s^{gr} \mathfrak{F}^{\mathsf{ck}}$ *for every cvAF* $\mathfrak{F}$*.*

Proof. By Proposition 8.4.29, we can modify all strongly gr-unacceptable arguments of $\mathfrak{F}$ without changing semantics. Next, we iteratively remove all redundant and strong unacceptable arguments (cf. Proposition 8.4.5 and 8.4.6). ☐

Corollary 8.4.32. $gr_i(\mathfrak{F}) = gr_i(\mathfrak{F}^{\mathsf{ck}})$ *for every cvAF* $\mathfrak{F}$*.*

Lemma 8.4.33. *For a cvAF $\mathfrak{F} = (A, R)$ and a strongly gr-unacceptable argument $x \in A$. Let $x' = (vul(x) \cup \{cl(x)\}, cl(x))$ and let $\mathfrak{F}' = f_e(\mathfrak{F} \setminus \{x\}, x') = (A', R')$. Then, for all $y \neq x \in A$, y is strongly gr-unacceptable in $\mathfrak{F}$ iff y is strongly gr-unacceptable in $\mathfrak{F}'$.*

Proof. Let $y \in A$ be strongly gr-unacceptable in $\mathfrak{F}$. Then there is $z \in A$ with $vul(z) \setminus \{cl(y)\} \subseteq vul(y)$ and $(z, y) \in R$ in $\mathfrak{F}$. In case $z \neq x$ we are done (then $z \in A'$). In case $z = x$, we have $cl(x) \in vul(y)$. Replacing x in $\mathfrak{F}'$ with x', we obtain $vul(x') = vul(x) \cup \{cl(x)\}$, thus $vul(x') \setminus \{cl(y)\} \subseteq vul(y)$ and $(x', y) \in R$ showing that y is strongly (gr-)unacceptable in $\mathfrak{F}'$. In case $y \in A'$ is strongly gr-unacceptable in $\mathfrak{F}'$, there is a witness $z \in A'$ in $\mathfrak{F}$. Using $A' \subset A$ we obtain that y is strongly gr-unacceptable in $\mathfrak{F}$. ☐

Analogously to Proposition 8.4.17, we obtain the following result.

Proposition 8.4.34. *The grounded kernel $\mathfrak{F}^{\mathsf{gk}}$ of a cvAF $\mathfrak{F}$ does neither contain redundant, non-self-attacking strongly gr-unacceptable nor strongly defeated arguments.*

Lemma 8.4.35. *For two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{gr} \mathfrak{G}$ implies $cl(A_{\mathfrak{F}^{\mathsf{gk}}}) = cl(A_{\mathfrak{G}^{\mathsf{gk}}})$.*

Proof. Let $x \in A_{\mathfrak{F}^{\mathsf{gk}}}$ with $cl(x) = c$. Towards a contradiction, assume that there is no argument $y \in A_{\mathfrak{G}^{\mathsf{gk}}}$ with $cl(y) = c$. Since we may assume $gr_{cl}(\mathfrak{F}^{\mathsf{gk}}) = gr_{cl}(\mathfrak{G}^{\mathsf{gk}})$, x does not occur in the grounded extension of $\mathfrak{F}^{\mathsf{gk}}$. Consequently, x receives incoming attacks.

<u>Case 1</u> Suppose x is no self-attacker. Consider the set $\mathcal{Z} = \{z \in A_{\mathfrak{F}^{\mathsf{gk}}} \mid (z, x) \in R_{\mathfrak{F}}\}$ of arguments attacking x. Since x is no self-attacker, by definition of the kernel we have $vul(z) \setminus (vul(x) \cup \{cl(x)\}) \neq \emptyset$ for each $z \in \mathcal{Z}$. Thus by letting

$$V_{\mathcal{Z}} = \{v_e = (\emptyset, e) \mid e \in vul(z) \setminus (vul(x) \cup \{cl(x)\}), z \in \mathcal{Z}\},$$

we defeat these attackers without introducing claim c. Thus c appears in the grounded extension of $f_e(\mathfrak{F}^{\mathsf{gk}}, V_{\mathcal{Z}})$ but not in $f_e(\mathfrak{G}^{\mathsf{gk}}, V_{\mathcal{Z}})$.

<u>Case 2</u> Now suppose each argument with claim c is a self-attacker and fix such x. Since x occurs in the kernel $\mathfrak{F}^{\mathsf{gk}}$, each attacker of x must itself possess attacking arguments.

First, we get rid of arguments with the same claim c. Let $\mathcal{Y} = \{y \in A_{\mathfrak{F}^{\mathsf{gk}}} \mid cl(y) = c, y \neq x\}$ denote the set of arguments with claim c. We consider arguments which defeat them; this time we can get rid of all of them via

$$\begin{aligned} V_{\mathcal{Y}} &= \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}, e \neq c\} \\ &= \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}\}. \end{aligned}$$

By introducing a self-attacker to each claim except c we ensure that all arguments except the unattacked ones are attacked and hence undecided in the unique grounded extension; in particular, x is. Thus consider $V = \{v_e = (\emptyset, e) \mid e \in cl(\mathfrak{F}^{\mathsf{gk}}), e \neq c\}$. With the usual technique—introducing a fresh argument attacked by c—we separate the cvAFs. ☐

Theorem 8.4.36. *For two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{gr} \mathfrak{G}$ iff $\mathfrak{F}^{\mathsf{gk}} = \mathfrak{G}^{\mathsf{gk}}$.*

Proof. First assume $\mathfrak{F}^{\mathsf{gk}} = \mathfrak{G}^{\mathsf{gk}}$ holds. By Proposition 8.4.12, it holds that $\mathfrak{F}^{\mathsf{gk}} \equiv_s^{gr} \mathfrak{F}$ and $\mathfrak{G}^{\mathsf{gk}} \equiv_s^{gr} \mathfrak{G}$. Thus we obtain $\mathfrak{F} \equiv_s^{gr} \mathfrak{G}$ by transitivity.

For the other direction, assume $\mathfrak{F} \equiv_s^{gr} \mathfrak{G}$. By Lemma 8.4.35, it holds that $\mathfrak{F}^{\mathsf{gk}}$ and $\mathfrak{G}^{\mathsf{gk}}$ contain the same claims. To show that for all arguments x in $\mathfrak{F}^{\mathsf{gk}}$ there is some argument y in $\mathfrak{G}^{\mathsf{gk}}$ such that $cl(x) = cl(y) = c$ and $vul(y) = vul(x)$, we proceed analogous to the proof of Theorem 8.4.19. Hence we obtain $\mathfrak{F}^{\mathsf{gk}} = \mathfrak{G}^{\mathsf{gk}}$. □

8.4.4 Stable Kernel for cvAFs

Finally, we consider stable semantics. We start with the crucial observation that the particular conclusion of self-attacking arguments is not of importance.

Example 8.4.37. *Consider the following two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$:*

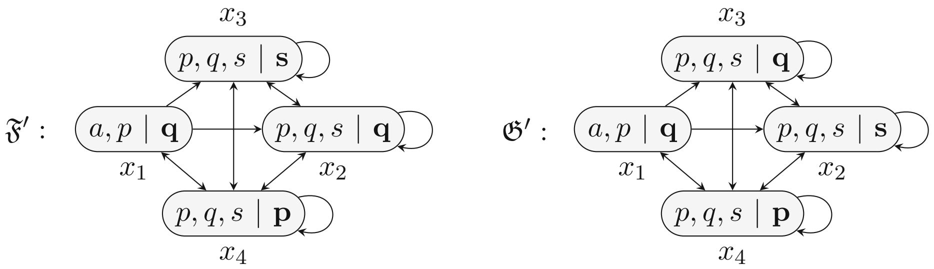
The only difference between $\mathfrak{F}$ and $\mathfrak{G}$ is the claim of the self-attacker x_2. Both $\mathfrak{F}$ and $\mathfrak{G}$ have the same unique stable extension $\{q\}$. As we will see, this is not a coincidence: for stable semantics, self-attacking arguments are indistinguishable w.r.t. their claims.

Proposition 8.4.38. *Given a cvAF $\mathfrak{F} = (A, R)$ and a self-attacking argument $x \in A$. For any $s \in vul(x)$, it holds that $stb_i(\mathfrak{F}) = stb_i(f_e(\mathfrak{F}, \{(vul(x), s)\}))$.*

Proof. Let $\mathfrak{F}' = f_e(\mathfrak{F}, \{(vul(x), s)\})$ and let $y = (vul(x), s)$. Then $x, y \notin E$ for all stable extensions E in $\mathfrak{F}$ and $\mathfrak{F}'$. The statement thus follows by observing that y is attacked by a stable extension $E \in stb(\mathfrak{F}')$ iff E attacks x in $\mathfrak{F}'$ iff E attacks x in $\mathfrak{F}$. □

Hence we can *add* all such self-attackers without changing stable semantics.

Example 8.4.39. *By adding all self-attackers $(vul(x_2), s)$ with $s \in vul(x_2)$ to our cvAFs $\mathfrak{F}$ and $\mathfrak{G}$ from Example 8.4.37 we obtain the following identical frameworks:*

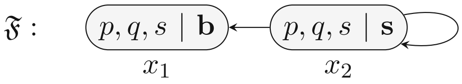


As shown in Proposition 8.4.7, strongly unacceptable arguments can be removed under stable semantics. However, as we observe in the above example, all of the arguments x_2, x_3, x_4 are strongly unacceptable w.r.t. to each other. Hence we just remove *strictly* strongly unacceptable arguments to guarantee that our kernel is well-defined.

Definition 8.4.40. *For a cvAF* $\mathfrak{F} = (A, R)$*,* $x \in A$ *is* strictly strongly unacceptable *if there is* $y \in A$ *with* $(y, x) \in R$ *and* $vul(y) \subsetneq vul(x)$*.*

To ensure that we catch all redundancies we need to take care of another issue.

Example 8.4.41. *Consider the following cvAF* $\mathfrak{F}$*:*



The argument x_1 *is not strictly strongly unacceptable w.r.t.* x_2 *hence it might be unsafe to remove it as observed above. However, if we apply the usual modification for strongly unacceptable arguments—adding the claim to the set of vulnerabilities—we obtain* $vul(x_2) \subsetneq vul(x_1)$*, i.e.,* x_1 *is now strictly strongly unacceptable w.r.t.* x_2*.*

To catch all redundancies we first have to add all ‘missing’ self-attackers including the modifications of strongly unacceptable arguments before we can delete all redundant, strongly defeated, and strictly strongly unacceptable arguments. Putting these observations together, we construct the stable kernel by performing the following steps:

1. we modify all strongly unacceptable arguments x by adding $cl(x)$ as vulnerability;
2. for each self-attacking argument x we add arguments $(vul(x), c)$ for all $c \in vul(x)$;
3. we delete all redundant, strongly defeated, and strictly strongly unacceptable arguments.

Definition 8.4.42. *For a cvAF* $\mathfrak{F} = (A, R)$*, let* X *denote the set of all strongly unacceptable arguments in* A *and let*

$$\mathfrak{F}' = (A', R') = f_e(\mathfrak{F} \setminus X, \{(vul(x) \cup \{cl(x)\}, cl(x)) \mid x \in X\}).$$

Now, let Y *denote the set of all self-attacking arguments in* A *and let*

$$\mathfrak{F}'' = (A'', R'') = f_e(\mathfrak{F}', \{(vul(x), s) \mid x \in Y, s \in vul(x)\}).$$

We define the stable kernel $\mathfrak{F}^{\mathsf{sk}} = (A^{\mathsf{sk}}, R^{\mathsf{sk}})$ *with*

$$\begin{aligned} A^{\mathsf{sk}} &= A' \setminus \{x \in A'' \mid x \textit{ is strongly defeated, redundant,} \\ &\qquad\qquad\qquad \textit{or strictly strongly unacceptable}\}, \\ R^{\mathsf{sk}} &= R'' \cap (A^{\mathsf{sk}} \times A^{\mathsf{sk}}). \end{aligned}$$

By iterative application of Proposition 8.4.38, 8.4.5, 8.4.7, and 8.4.6 we obtain

Proposition 8.4.43. $\mathfrak{F} \equiv_s^{gr} \mathfrak{F}^{\mathsf{sk}}$ *for every cvAF* $\mathfrak{F}$*.*

We show that the deletion of strongly unacceptable, redundant, or strongly defeated arguments does not change strong unacceptability, redundancy, or strong defeat of other arguments. Hence such arguments can be iteratively removed.

Lemma 8.4.44. *For a cvAF* $\mathfrak{F} = (A, R)$ *and a strongly unacceptable/redundant/strongly defeated argument* $y \in \mathfrak{F}$*,* $y \neq x \in A$ *is strongly unacceptable, redundant, or strongly defeated in* $\mathfrak{F}$ *iff* x *is strongly unacceptable, redundant, or strongly defeated in* $\mathfrak{F} \setminus \{y\}$*.*

Proof. We first observe that if x is strongly unacceptable/redundant/strongly defeated in $\mathfrak{F} \setminus \{y\}$ then there is a witness z in $\mathfrak{F} \setminus \{y\}$. As mentioned in Remark 8.4.14, the claim-attacks are not affected by removing certain arguments. We thus obtain that z witnesses that x is strongly unacceptable/redundant/strongly defeated in $\mathfrak{F}$. Also, in case x is strongly defeated, it is clear that x is contained in $\mathfrak{F} \setminus \{y\}$ since y is not unattacked and thus cannot serve as witness for x being strongly defeated.

- Let y be strongly unacceptable. First, let x be strongly unacceptable in $\mathfrak{F}$. In case y witnesses strong unacceptability of x in $\mathfrak{F}$, there is z with $vul(z) \subseteq vul(y)$ and $cl(z) \in vul(y)$. Then z witnesses unacceptability of x in $\mathfrak{F}$ since $vul(z) \subseteq vul(x)$ and $cl(z) \in vul(x)$. W.l.o.g. let z be minimal in the sense that there is no $u \in A$ with $vul(u) \subset vul(y)$ and $cl(u) \in vul(y)$. Then z is not strongly unacceptable in $\mathfrak{F}$ (otherwise, we find such an u, contradiction to the minimality assumption), and thus z witnesses unacceptability of x in $\mathfrak{F} \setminus \{y\}$.

  In case x is redundant in $\mathfrak{F}$ w.r.t. y, there is z with $vul(z) \subseteq vul(y)$ and $cl(z) \in vul(y)$. Thus $vul(z) \subseteq vul(x)$ and $cl(x) \in vul(y)$ shows that x is strongly unacceptable in $\mathfrak{F}$. We obtain x is strongly unacceptable in $\mathfrak{F} \setminus \{y\}$.
- Let y be strongly defeated. In case x is redundant w.r.t. y in $\mathfrak{F}$, there is some $z \in A$ with $(z, y) \in R$. Thus we obtain that also x is strongly defeated (using $vul(y) \subseteq vul(x)$, i.e., $(z, x) \in R$). In case x is strongly unacceptable w.r.t. y in $\mathfrak{F}$, also x is strongly defeated (using $vul(y) \subseteq vul(x)$).
- Let y be redundant. First, let x be redundant w.r.t. y. Then there is $vul(z) \subseteq vul(y)$ and $cl(z) = cl(y)$, thus witnessing redundancy of x. In case x is unacceptable w.r.t. y in $\mathfrak{F}$. There is $z \in A$ satisfying $vul(z) \subseteq vul(y)$ and $cl(z) = cl(y)$, and thus z witnesses unacceptability of x in $\mathfrak{F} \setminus \{y\}$. ☐

Proposition 8.4.45. *For a cvAF* $\mathfrak{F}$*, the stable kernel* $\mathfrak{F}^{\mathsf{sk}}$ *does neither contain redundant, strictly strongly unacceptable, nor strongly defeated arguments.*

We show that the stable kernels of two strongly equivalent cvAFs contain the same claims.

Lemma 8.4.46. *For two cvAFs* $\mathfrak{F}$ *and* $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{stb} \mathfrak{G}$ *implies* $cl(A_{\mathfrak{F}^{\mathsf{sk}}}) = cl(A_{\mathfrak{G}^{\mathsf{sk}}})$.

Proof. Let $x \in A_{\mathfrak{F}^{\mathsf{sk}}}$ with $cl(x) = c$. Towards a contradiction, assume that there is no argument $y \in A_{\mathfrak{G}^{\mathsf{sk}}}$ with $cl(y) = c$. Since we may assume $stb_i(\mathfrak{F}^{\mathsf{sk}}) = stb_i(\mathfrak{G}^{\mathsf{sk}})$ in this case we deduce that x does not occur in any stable extension of $\mathfrak{F}^{\mathsf{sk}}$. Hence it does not occur in any admissible extension. Consequently, x receives incoming attacks.

<u>Case 1</u> Suppose x is no self-attacker. We have to deal with three kinds of arguments:

- same claim as c (we block these arguments),
- attacking x (we block these arguments, whenever x cannot do this on its own),
- odd cycles (we disrupt all of them).

Then c appears in a stable extension of $\mathfrak{F}$ but not in $\mathfrak{G}$, because we will never add claim c.

Consider the set $\mathcal{Y} = \{y \in A_{\mathfrak{F}^{\mathsf{sk}}} \mid cl(y) = c, y \neq x\}$ of arguments with claim c. By assumption, none of these arguments attacks x. We consider arguments which defeat them unless this would require either defeating x as well or adding claim c. We let

$$V_{\mathcal{Y}} = \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}, e \neq c\}.$$

The remaining arguments $y \in \mathcal{Y}$ which are not attacked by $V_{\mathcal{Y}}$ must satisfy $vul(y) \setminus vul(x) = \{c\}$ and are therefore attacked by x (self-attackers).

Now consider the set $\mathcal{Z} = \{z \in A_{\mathfrak{F}^{\mathsf{sk}}} \mid (z, x) \in R_{\mathfrak{F}}, (x, z) \notin R_{\mathfrak{F}}\}$ of arguments attacking x without receiving a counter-attack. For $z \in \mathcal{Z}$ it holds that $cl(z) \in vul(x)$ and therefore, by our definition of the stable kernel, it cannot be the case that $vul(z) \subseteq vul(x)$. Moreover, $c \notin vul(z)$ since that would imply a counterattack from x. Therefore with

$$V_{\mathcal{Z}} = \{v_e = (\emptyset, e) \mid e \in vul(z) \setminus vul(x), z \in \mathcal{Z}\}$$

we get rid of them and we have now already ensured that $\{x\}$ becomes admissible.

Now consider the set $\mathcal{S} = \{s \in A_{\mathfrak{F}} \mid (s, s) \in R_{\mathfrak{F}}\}$ of self-attacking arguments. By definition of the stable kernel, we have $vul(s) \not\subseteq vul(x)$ for all $s \in \mathcal{S}$. Therefore with $V_{\mathcal{S}} = \{v_e \mid e \in vul(s) \setminus vul(x), s \in \mathcal{S}\}$ we get rid of them without attacking x.

Now consider any odd cycle $\mathcal{O} = \{o_1, \ldots, o_n\}$ occurring in $\mathfrak{F}^{\mathsf{sk}}$. Our goal is to argue that $\bigcup vul(o_i) \subseteq vul(x)$ is impossible; i.e., we can disrupt the odd cycle without attacking x. Assume the contrary, i.e., suppose $\bigcup vul(o_i) \subseteq vul(x)$. Then $cl(o_i) \in vul(x)$ for each i. Since $vul(o_i) \subseteq vul(x)$ this implies that x is unacceptable contradicting the construction of the stable kernel $\mathfrak{F}^{\mathsf{sk}}$. Thus, by adding appropriate arguments we can disrupt the odd cycles and therefore, the admissible set $\{x\}$ can be extended to a stable extension.

<u>Case 2</u> Now suppose each argument with claim c is a self-attacker and fix such x. Again, we have to deal with three kinds of arguments:

- same claim as c (block these arguments),
- attacking x (we block all of these arguments),
- odd cycles (we disrupt all of them).

Then $\mathfrak{F}$ has no stable extension, but one after we add claim c, where in $\mathfrak{G}$ adding c does not change anything.

Consider the set $\mathcal{Y} = \{y \in A_{\mathfrak{F}^{\mathsf{pk}}} \mid cl(y) = c, y \neq x\}$ of arguments with claim c. We consider arguments which defeat them; this time we can get rid of all of them via

$$\begin{aligned} V_{\mathcal{Y}} &= \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}, e \neq c\} \\ &= \{v_e = (\emptyset, e) \mid e \in vul(y) \setminus vul(x), y \in \mathcal{Y}\}. \end{aligned}$$

Now consider the set $\mathcal{Z} = \{z \in A_{\mathfrak{F}^{\mathsf{sk}}} \mid (z, x) \in R_{\mathfrak{F}^{\mathsf{sk}}}\} \setminus \mathcal{Y}$ of arguments attacking x and not having claim c. For $z \in \mathcal{Z}$ it holds that $cl(z) \in vul(x)$ and therefore, by our definition of the stable kernel, it cannot be the case that $vul(z) \subseteq vul(x)$. Moreover, $c \in x^{cl-}$ implies $c \notin vul(z) \setminus vul(x)$. Therefore with $V_{\mathcal{Z}} = \{v_e = (\emptyset, e) \mid e \in vul(z) \setminus vul(x), z \in \mathcal{Z}\}$ we get rid of them without introducing claim c. Moreover, we deal with the set of self-attackers $\mathcal{S} = \{s \in A_{\mathfrak{F}} \mid (s, s) \in R_{\mathfrak{F}}\}$ as before via $V_{\mathcal{S}} = \{v_e \mid e \in vul(s) \setminus vul(x), s \in \mathcal{S}\}$.

Now consider any odd cycle $\mathcal{O} = \{o_1, \dots, o_n\}$ occurring in $\mathfrak{F}^{\mathsf{sk}}$. Towards a contradiction, suppose $\bigcup vul(o_i) \subseteq vul(x)$. Then $cl(o_i) \in vul(x)$ for each i. Since $vul(o_i) \subseteq vul(x)$ this would, however, imply that x is unacceptable contradicting the construction of the stable kernel $\mathfrak{F}^{\mathsf{sk}}$. Thus, by adding appropriate arguments we ensure that $\mathfrak{F}$ has no stable extension, but with the self-attacker x being the only odd cycle.

Therefore, $f_e(\mathfrak{F}^{\mathsf{sk}}, X)$ has no stable extension, but adding an isolated argument with claim c resolves this; meanwhile, adding this argument to $f_e(\mathfrak{G}^{\mathsf{sk}}, X)$ does not influence whether or not there is a stable extension. □

We are ready to present our characterization result for cvAF strong equivalence with respect to stable semantics (the proof proceeds analogous to the proof of Theorem 8.4.19).

Theorem 8.4.47. *For two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$, $\mathfrak{F} \equiv_s^{stb} \mathfrak{G}$ iff $\mathfrak{F}^{\mathsf{sk}} = \mathfrak{G}^{\mathsf{sk}}$.*

Our results yield criteria to check strong equivalence for any two cvAFs without testing each possible expansion or searching for counter-examples. Consider, for instance, our cvAFs $\mathfrak{F}$ and $\mathfrak{G}$ from Example 8.4.37. Since their stable kernels are syntactically equivalent (cf. Example 8.4.41), we conclude $\mathfrak{F} \equiv_s^{stb} \mathfrak{G}$ by Theorem 8.4.47.

8.4.5 Consequences for Assumption-based Argumentation

By transferring the above results in the context of ABA we obtain that deciding strong equivalence for flat, atomic ABA frameworks with separated contraries is tractable.

Theorem 8.4.48. *For two atomic, flat ABA frameworks* $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^{-})$ *and* $D' = (\mathcal{L}, \mathcal{R}', \mathcal{A}', ^{-})$ *with separated contraries, deciding* $D \equiv_s^\sigma D'$ *is tractable.*

Proof. We construct the cvAFs $\mathfrak{F}_D$ and $\mathfrak{F}_{D'}$ corresponding to D and D'.

Let us first consider the instantiated arguments corresponding to assumptions, i.e., $X^D = \{(\{\overline{a}\}, a) \mid a \in \mathcal{A}\}$ in $\mathfrak{F}_D$ resp. $X^{D'} = \{(\{\overline{a}\}, a) \mid a \in \mathcal{A}'\}$ in $\mathfrak{F}_{D'}$. We make the following observation: For each assumption-argument $x \in X^D \cup X^{D'}$, it holds that x is either i) strongly defeated or ii) strongly unacceptable, or iii) remains unchanged in the kernel of the corresponding cvAF. Let us discuss all other cases.

- In case x is strictly strongly unacceptable, it is strongly defeated: $vul(x) = \{\overline{a}\}$ is a singleton, it holds that x is attacked by some argument with no vulnerabilities.
- Strong gr-unacceptability is equivalent to strong unacceptability for ABA frameworks which separate contraries; hence it suffices to discuss the latter.
- It cannot be redundant because $cl(x)$ does appear as conclusion of some other argument (we consider flat ABA frameworks).
- It cannot be strongly pr-unacceptable because $cl(x)$ cannot appear as vulnerability of any argument since we assume that D and D' separate contraries.

So let us consider the cases i), ii), and iii) mentioned above.

i) Suppose $x \in X^D \cup X^{D'}$ is strongly defeated. By our previous results, we can remove the assumption from the corresponding ABA framework without changing the semantics (even considering arbitrary expansions). Hence, we can w.l.o.g. assume that no assumption is strongly defeated in D or D'.

ii) Let X_{su}^D and $X_{su}^{D'}$ denote the set of assumption-arguments that are strongly unacceptable in $\mathfrak{F}_D$ and $\mathfrak{F}_{D'}$, respectively. We show that $X_{su}^D \neq X_{su}^{D'}$ implies that $D \not\equiv_s^\sigma D'$. By symmetry, it suffices to consider some $x \in X_{su}^D \setminus X_{su}^{D'}$

First note that if $x = (\{\overline{a}\}, a) \in X_{su}^D$ is strongly unacceptable, there must be some argument y occurring in $\mathfrak{F}_D$ which attacks x (i.e., $cl(y) = \overline{a}$) and satisfies $vul(y) \subseteq vul(x)$ (i.e., $vul(y) = \emptyset$ or $vul(y) = \{\overline{a}\}$), where the case $vul(y) = \emptyset$ is excluded since x is not strongly defeated. Thus, y is of the form $y = (\{\overline{a}\}, \overline{a})$.

Since D is atomic, the only way to induce such an argument y is by $\overline{a} \leftarrow a$. If this rule occurs in D' as well, then $a \in \mathcal{A}'$ since D' is atomic; thus $x \in X_{su}^{D'}$ contradicting our assumption. So $\overline{a} \leftarrow a$. does not occur in D', and we proceed as follows:

(a) Suppose $\overline{a} \in Th_{D'}(\emptyset)$ (i.e., $\overline{a}$ must be a fact since D' is atomic). Then $\overline{a}$ is a fact in D', but by assumption not in D. Then consider

$$\mathcal{R}_H = \{\overline{a} \leftarrow . \mid a \in (A \cup \mathcal{A}') \setminus \{a\}\}$$

and let $H = \{\mathcal{L}, \mathcal{R}_H, \mathcal{A} \cup \mathcal{A}', {}^{-}\}$. Hence D and D' are not strongly equivalent ($\overline{a}$ is accepted in $D' \cup H$, but not in $D \cup H$).

(b) Suppose $\overline{a} \notin Th_{D'}(\emptyset)$; and assume for the moment $a \in \mathcal{A}'$. Since the rule $\overline{a} \leftarrow a.$ does not occur in D', our reasoning from above shows that a is neither strongly unacceptable nor strongly defeated in D' (and the other cases cannot occur). Therefore, our enforcement results show that a can be enforced in D', but not in D (yielding a suitable counter-example for strong equivalence). Finally, if $a \notin \mathcal{A}'$, then first add $H = (\{a\}, \emptyset, \{a\}, \{a \mapsto \overline{a}\})$ and apply the same argument afterwards.

Hence $X_{su}^{D} \neq X_{su}^{D'}$ implies that $D \not\equiv_s^\sigma D'$, i.e., we can assume $X_{su}^{D} = X_{su}^{D'}$.

iii) Let X_n^D and $X_n^{D'}$ denote the set of assumption-arguments that remain unchanged in the kernel of $\mathfrak{F}_D$ and $\mathfrak{F}_{D'}$, respectively. By our enforceability results we can enforce each of them, hence we immediately obtain that $X_n^D \neq X_n^{D'}$ implies that $D \not\equiv_s^\sigma D'$.

To summarize, we may w.l.o.g. assume $\mathcal{A} = \mathcal{A}'$, otherwise we can handle the ABA frameworks with the above arguments. Now we are ready to apply our cvAF results. Given $\mathcal{A} = \mathcal{A}'$, the following holds.

$$\begin{align}
(\mathfrak{F}_D^p)^{k(\sigma)} = (\mathfrak{F}_{D'}^p)^{k(\sigma)} &\Leftrightarrow \mathfrak{F}_D^p \equiv_s^\sigma \mathfrak{F}_{D'}^p \tag{8.1}\\
&\Leftrightarrow \text{for each set of inst. args } X : \sigma_i(f_e(\mathfrak{F}_D^p, X)) = \sigma_i(f_e(\mathfrak{F}_{D'}^p, X)) \tag{8.2}\\
&\Leftrightarrow \text{for each ABA } H : \sigma_i(\mathfrak{F}_{D \cup H}^p) = \sigma_i(\mathfrak{F}_{D' \cup H}^p) \tag{8.3}\\
&\Leftrightarrow \text{for each ABA } H : \sigma_i(\mathfrak{F}_{D \cup H}) = \sigma_i(\mathfrak{F}_{D' \cup H}) \tag{8.4}\\
&\Leftrightarrow \text{for each ABA } H : \sigma_{Th}(D \cup H) = \sigma_{Th}(D' \cup H) \tag{8.5}\\
&\Leftrightarrow D \equiv_s^\sigma D' \tag{8.6}
\end{align}$$

Equivalence (8.1) follows from Theorems 8.4.19, 8.4.36, 8.4.28, and 8.4.47 for the respective semantics. Equivalence (8.2) is by definition of strong equivalence for cvAFs.

(8.3) The crucial observation is that each rule r with assumptions that appear in the frameworks at hand corresponds to an instantiated argument and vice versa.

($\Rightarrow$) Given ABA $H = (\mathcal{L}, \mathcal{R}'', \mathcal{A}'', {}^{-})$, we let $X = \{(\overline{A}, p) \mid p \leftarrow A \in \mathcal{R}'', A \subseteq \mathcal{A} \cup \mathcal{A}''\}$. By Lemma 8.2.12, it holds that $\mathfrak{F}_{D \cup \{r\}}^p = f_e(\mathfrak{F}_D^p, \{(\overline{A}, p)\})$ for each rule $r = p \leftarrow A$ with $A \subseteq \mathcal{A} \cup \mathcal{A}''$. We obtain $\mathfrak{F}_{D \cup H}^p = f_e(\mathfrak{F}_D^p, X)$ and $\mathfrak{F}_{D' \cup H}^p = f_e(\mathfrak{F}_{D'}^p, X)$. Since $\sigma_i(f_e(\mathfrak{F}_D^p, X)) = \sigma_i(f_e(\mathfrak{F}_{D'}^p, X))$ we thus obtain $\sigma_i(\mathfrak{F}_{D \cup H}^p) = \sigma_i(\mathfrak{F}_{D' \cup H}^p)$.

($\Leftarrow$) Given a set of arguments X, we consider an expansion $H = (\mathcal{L}, \mathcal{R}'', \mathcal{A}'', {}^{-})$ such that all arguments in X are instantiated. For this, we need to ensure that $D \cup H$ contains all necessary assumptions, that is, we let $\mathcal{A}'' = \bigcup_{(\overline{A}, p) \in X} A$. Now, we add a rule for each argument in X, i.e., we let $\mathcal{R}'' = \{p \leftarrow A \mid (\overline{A}, p) \in X\}$. By Lemma 8.2.12, we obtain $\mathfrak{F}_{D \cup H}^p = f_e(\mathfrak{F}_D^p, X)$. Thus the statement follows.

(8.4) (⇒) Consider a ABA H. Let $H' = H \cup H_{\mathcal{A}}$ where $H_{\mathcal{A}}$ is the ABA framework consisting of the assumptions (and their contraries), i.e., $H_{\mathcal{A}} = (\mathcal{A} \cup \overline{\mathcal{A}}, \emptyset, \mathcal{A}, ^{-})$. By our assumption it holds that $\sigma_i(\mathfrak{F}^p_{D\cup H'}) = \sigma_i(\mathfrak{F}^p_{D'\cup H'})$. Hence we can add the assumptions to the instantiation: it holds that

$$\sigma_i(\mathfrak{F}^p_{D\cup H\cup H_{\mathcal{A}}}) = \sigma_i(\mathfrak{F}_{D\cup H}) = \sigma_i(\mathfrak{F}_{D'\cup H}) = \sigma_i(\mathfrak{F}^p_{D'\cup H\cup H_{\mathcal{A}}}).$$

(⇐) By Proposition 3.3.25, we can remove the assumptions from the extensions.

Equivalence (8.5) is by Proposition 3.3.20; finally, equivalence (8.6) is by definition of strong equivalence for ABA. Thus, to decide strong equivalence between D and D', it suffices to check

(i) $\mathcal{A} \setminus \{a \in \mathcal{A} \mid \overline{a} \leftarrow \in \mathcal{R}\} = \mathcal{A}' \setminus \{a \in \mathcal{A}' \mid \overline{a} \leftarrow \in \mathcal{R}'\}$; if this is not the case, we have $D \not\equiv^{\sigma}_{s} D'$; otherwise, we check

(ii) syntactical equivalence of the σ-kernels of $\mathfrak{F}^p_D$ and $\mathfrak{F}^p_{D'}$. ☐

As in the case of enforcement, we want to emphasize that moving from flat ABA to flat atomic ABA does not change the complexity class of this problem. However, if we additionally require that the frameworks have separated assumptions we obtain the desired tractable fragment.

8.5 Tractability Results for Logic Programs

We consider normal logic programs (LPs); for an overview we refer to Section 3.2. Given an LP P, the corresponding instantiated cvAF is denoted by $\mathfrak{F}_P$. We proceed as for ABA frameworks by applying the cvAF results. We give an LP version of Lemma 8.2.12.

Lemma 8.5.1. *Given an atomic LP P.*

- *For each atomic rule $r = c \leftarrow \mathit{not}\ B$, we have $\mathfrak{F}_{P\cup\{r\}} = f_e(\mathfrak{F}_P, x)$ with $x = (B, c)$.*
- *For each argument $x = (B, c)$, it holds that $\mathfrak{F}_{P\cup\{r\}} = f_e(\mathfrak{F}_P, x)$ with $r = c \leftarrow \mathit{not}\ B$.*

This suffices in order to efficiently investigate our two problems we considered before. The relation is even much closer since we do not need to handle additional assumptions. In accordance with our general definitions of enforcement and strong equivalence in non-monotonic reasoning formalisms, we define the LP enforcement problem as follows.

Definition 8.5.2. *Let P be an LP and σ a semantics. An atom p is σ-*enforceable *if there is a set R of rules s.t. $head(r) \neq p$ for all $r \in R$ and p is credulously accepted in $P \cup R$ w.r.t. semantics σ.*

Proposition 8.5.3. *Consider a semantics σ. Deciding atom-enforceability w.r.t. σ for the class of normal LPs is* NP*-hard.*

Proof. Let φ be a boolean formula given by clauses C over variables in X. The corresponding logic program P contains the following rules:

- the atomic rule '$p_\varphi \leftarrow \texttt{not}\ C$';
- rules '$p_c \leftarrow \{l \mid \neg l \in c\}, \texttt{not}\ \{l \in X \mid l \in c\}$' for each clause $c \in C$.

Intuitively, a clause-atom c is contained in a stable model M iff c is false in M. Hence we can accept φ iff $c \notin M$ for all $c \in C$. We show φ is satisfiable iff p_φ is enforceable in P. Since each stable model is well-founded in P it suffices to focus on stable semantics.

First assume φ is satisfiable. Assume M is a model of φ. We add each $x \in M$ as fact. We show that $Q = M \cup \{p_\varphi\}$ is a stable model of $P \cup M$. Consider $c \in C$. If $c \cap M \neq \emptyset$ then the rule r with $head(r) = p_c$ contains $\texttt{not}\ x$ for some $x \in c \cap M$. Hence the rule r is satisfied by Q. Likewise, if $c \cap M = \emptyset$ we have some $x \in X$ with $x \notin M$ and $\neg x \in c$. Hence the rule r with $head(r) = p_c$ satisfies $x \in body(r)$. Hence Q satisfies r.

Now assume p_φ is enforceable. Let R denote the set of rules which enforce p_φ, and let M denote the model of $R \cup P$ which contains p_φ. Then M does not contain any $c \in C$ (otherwise, p_φ would not be acceptable). Now, we show that $N = M \cap X$ is a model of φ. Again, for each rule $r \in P$ corresponding to a clause in $c \in C$, there is either some $x \in N$ with $\texttt{not}\ x \in body(r)$—in this case, $x \in c$ hence c is satisfied; or there is some $x \in X \setminus N$ with $x \in body(r)$—then $\neg x \in c$ and thus c is satisfied. □

Thus the enforcement problem is intractable for LPs in general. From our cvAF results we obtain tractability for atomic LPs.

Theorem 8.5.4. *For atomic LPs, deciding whether some atom is enforceable is tractable.*

Proof. By Corollary 8.3.13, we have for any atom a: a is enforceable in P iff a is credulously accepted in $P \cup H$ for some H iff a is credulously accepted in $f_e(\mathfrak{F}_P, X)$ for some X iff a is enforceable in $\mathfrak{F}_P$; the latter is tractable. □

Let us next discuss strong equivalence for atomic LPs. In general, we define strong equivalence for LP relative to a LP-fragment $\mathfrak{C}$ as follows.

Definition 8.5.5. *Two LPs $P, P' \in \mathfrak{C}$ are* strongly equivalent *w.r.t. a semantics σ in the fragment $\mathfrak{C}$, for short $P \equiv_s^\sigma P'$, if for each LP $R \in \mathfrak{C}$, it holds that $\sigma(P \cup R) = \sigma(P' \cup R)$.*

Without the requirement of P, P', and R being atomic, intractability of strong equivalence is well-known [146, 133]. With our results, we obtain a tractable fragment here as well.

Theorem 8.5.6. *Deciding strong equivalence in the class of atomic LPs is tractable.*

Proof. Immediate from Theorems 8.4.19, 8.4.36, 8.4.28, and 8.4.47: for two LPs P and P' it holds that P is atomic strongly equivalent to P' iff $\sigma(P \cup R) = \sigma(P' \cup R)$ for each atomic set of rules R iff $\sigma(f_e(\mathfrak{F}_P, H)) = \sigma(f_e(\mathfrak{F}_{P'}, H))$ for $H = \bigcup\{(B, c) \mid c \leftarrow \texttt{not}\ B \in R\}$ for each R iff $\sigma(f_e(\mathfrak{F}_P, X)) = \sigma(f_e(\mathfrak{F}_{P'}, X))$ for each set X of instantiated arguments iff $\mathfrak{F}_P^{k(\sigma)} = cvF_{P'}^{k(\sigma)}$; the latter is tractable. □

For stable model semantics, we obtain an even more general result: strong equivalence between two atomic LPs is tractable even if we consider expansions with rules that are non-atomic. For this, we will first show that each atomic LP P is strongly equivalent to the program obtained by re-translating the stable kernel $\mathfrak{F}_P^{\mathsf{sk}}$ w.r.t. stable semantics.

Proposition 8.5.7. *Let P be an atomic LP and let P^{sk} denote the LP $P_{\mathfrak{F}_P^{\mathsf{sk}}}$. It holds that P and P^{sk} are strongly equivalent w.r.t. stable semantics in the class of normal LPs.*

Proof. In the following, we use the terms instantiated arguments and atomic rules interchangeably. For simplicity, we will talk about redundant, strongly unacceptable, and strongly defeated rules instead of formally switching between the formalisms. By Lemma 8.5.1, these concepts are indeed transferable to the realm of atomic LPs.

Consider a set of rules H. We show that M is a stable model of $P' = P \cup H$ iff M is a stable model of $P'' = P^{\mathsf{sk}} \cup H$. The underlying observation is that the reduct of P'/M coincides with P''/M in case M is a model of P' or P''.

($\Rightarrow$) First assume M is a stable model of P'. We show that $P'/M = P''/M$. For each rule in the reduct obtained from $r \in H$, the statement holds true. Moreover notice that all other rules not originating from rules in H are facts. Assume $a. \in P'/M$ but $a. \notin P''/M$. Let $r \in P$ denote some rule with $head(r) = a$ which has survived the reduct modifications. That is, each negated literal in the body of r is false, i.e., $neg(r) \subseteq \mathcal{L}(P \cup H) \setminus M$.

Now, since $a. \notin P''/M$ we have either (1) r is deleted when building the kernel of P or (2) r is strongly unacceptable in P and thus the modified rule $r' = a \leftarrow neg(r) \cup \{\texttt{not}\ a\}$ is deleted when building the reduct of $P^{\mathsf{sk}} \cup H$.

Let us first deal with <u>case 2:</u> let $t \in P$ be a rule witnessing unacceptability of r in P. That is, $neg(t) \subseteq neg(r)$ and $head(t) \in neg(r)$. Hence each atom $b \in neg(t)$ is false (not contained in M) and thus $\texttt{not}\ b$ is removed from the body of t when forming the reduct. We obtain that the rule $head(t)$. is contained in P'. Since M is a model of P', it holds that $head(t) \in M$ (by (a) in Definition 3.2.3). Consequently, $M \cap neg(r) \neq \emptyset$, i.e., r is removed when constructing the reduct, contradiction to our above assumption.

In <u>case 1</u>, rule r is deleted when constructing the kernel P^{sk}. That is r is either strongly defeated, strictly strongly unacceptable, or redundant in P. In the former case, there is some fact $b. \in P$ such that $b \in neg(r)$. Hence $b \in M$ and we obtain that r is deleted when constructing the reduct P'. In case r is strictly strongly unacceptable, we proceed as in case 2. In case r is redundant, consider a rule $s \in P$ with $neg(s) \subset neg(r)$ and $head(r) = head(s)$. W.l.o.g., let s be minimal in that aspect (i.e., there is no rule s' with $neg(s') \subset neg(r)$ and $head(r) = head(s')$ and $neg(s') \subset neg(s)$). If s is contained in P^{sk}, then we have $neg(s) \subseteq \mathcal{L}(P \cup H) \setminus M$ and $head(s) = a$, hence we have found a witness showing that the fact $a.$ is contained in P'' as well. In case s is not contained in P^{sk}, it holds that s is either strictly strongly unacceptable (we proceed as in case 2) or strongly defeated (we proceed as above). Hence we obtain that $P'/M \subseteq P''/M$.

For the other direction, assume $a. \in P''/M$ but $a. \notin P'/M$. Let $r \in P$ denote some rule with $head(r) = a$ which has survived the reduct modifications in P''/M. That is, each negated literal in the body of r is false, i.e., $neg(r) \subseteq \mathcal{L}(P \cup H) \setminus M$. Now, since $a. \notin P'/M$ we have either (1) r is not contained in P but $r' = a \leftarrow neg(r) \setminus \{\texttt{not } a\}$ is unacceptable in P or (2) r is a self-attacker which has been added when building the kernel of P. In any other cases, r would be contained in P as well. In both cases, $\texttt{not } a \in body(r)$ implies $M \cap neg(r) \neq \emptyset$; contradiction to r witnessing $a. \in P''/M$.

We obtain $P'/M = P''/M$ for each stable model M of P' which implies that M is a stable model of P'' as well.

($\Leftarrow$) For the other direction, consider a stable model M of P''. We show that We show that $P'/M = P''/M$. For each rule in the reduct obtained from some rule $r \in H$, the statement holds true. Moreover notice that all other rules not originating from rules in H are facts. In case $a. \in P''/M$ but $a. \notin P'/M$ we proceed as above (notice that we did not make use of the fact that M was a model of P' and not of P'').

For the other direction, let us assume $a. \in P'/M$ but $a. \notin P''/M$. Let $r \in P$ denote some rule with $head(r) = a$ which has survived the reduct modifications. That is, each negated literal in the body of r is false, i.e., $neg(r) \subseteq \mathcal{L}(P \cup H) \setminus M$.

Again, we distinguish the cases (1) r is deleted when building the kernel of P or (2) r is strongly unacceptable in P and thus the modified rule $r' = a \leftarrow neg(r) \cup \{\texttt{not } a\}$ is deleted when building the reduct of $P^{\mathsf{sk}} \cup H$.

Case 2: let $t \in P$ be a rule witnessing unacceptability of r in P which is minimal in this aspect, i.e., $\{head(t)\} \cup neg(t)$ is $\subseteq$-minimal among all such rules. Then it holds that $t' = head(t) \leftarrow body(t) \cup \{\texttt{not } head(t)\}$ is contained in P^{sk}. Moreover, $neg(t') \subseteq neg(r)$. Since M is a model of P'' we obtain that $neg(t') \cap M \neq \emptyset$ (otherwise, it holds that $head(t) \in M$ and $head(t) \notin M$ by definition of stable model semantics). Hence $neg(r) \cap M \neq \emptyset$, contradiction to our assumption $neg(r) \subseteq \mathcal{L}(P \cup H) \setminus M$.

Case 1: we perform only syntactical modifications, that is, we can proceed analogous to case 1 for the other direction. This concludes the proof of the statement. □

By our above results, we obtain that strong equivalence w.r.t. stable semantics coincides in the class of atomic and normal LPs when we compare atomic LPs.

Theorem 8.5.8. *$P \equiv_s^{stb} Q$ in the class of atomic LPs iff $P \equiv_s^{stb} Q$ in the class of normal LPs for any two atomic LPs P and Q.*

Proof. In case P and Q are not strongly equivalent in the class of atomic LPs we obtain that they are not strongly equivalent in the class of normal LPs as the former is a special case. Now assume P and Q are strongly equivalent in the class of atomic LPs. Then their stable kernels coincide (by Theorem 8.4.47). By Proposition 8.5.7, we obtain $P \equiv_s^{stb} P^{\mathsf{sk}} = Q^{\mathsf{sk}} \equiv_s^{stb} Q$ in the class of normal LPs. □

Corollary 8.5.9. *Deciding whether two atomic LPs P and Q are strongly equivalent w.r.t. stable semantics in the class of normal LPs is tractable.*

8.6 Summary

In this chapter, we investigated strong equivalence and claim enforcement in the context of instantiations. Hereby, we focused on assumption-based argumentation. We showed that in general, both tasks are intractable for ABA. Inspired by tractability of the corresponding problems for CAFs, we proposed an adjusted instantiation procedure via cvAFs to obtain a closer relation between the knowledge base and the corresponding abstract representation. Our cvAFs consider not only claims but also vulnerabilities of arguments. With this, we were able to capture the hidden weaknesses of arguments. We showed that strong equivalence and enforcement is tractable for cvAFs. We provided several novel redundancy notions for instantiated arguments that gave rise to kernel characterizations for strong equivalence and criteria for claim enforcement. Exploiting the close correspondence between atomic ABA frameworks and cvAFs, we demonstrated how our cvAF tractability results yield tractable fragments of ABA for these problems.

Similar to other studies in the area of abstract argumentation, we characterize strong equivalence via semantics-dependent kernels (cf. [142, 85], or Chapter 7). However, there is an important difference: in contrast to strong equivalence characterizations in CAFs, AFs, or other abstract representations of argumentation, our kernels are obtained by manipulating and removing arguments. While for AFs and CAFs, two strongly equivalent frameworks necessarily agree on their arguments and on their self-attacker, two cvAFs can be strongly equivalent to each other although they do not even agree on the number of arguments. This is a distinguishing feature of cvAFs.

Finally, we applied our techniques to LPs as well. We showed that the problem of atom enforcement for LPs is intractable. Using our cvAF results, we obtained tractable fragments for strong equivalence and enforcement. Here, we want to point out that our redundancy notions for instantiated arguments may be of independent interest when transferring them to rules in LPs (as done in Proposition 8.5.7, where we considered the LP obtained from constructing the stable kernel of the corresponding cvAF). Indeed, as demonstrated in the proof of Proposition 8.5.7, our redundancy notions for instantiated arguments give rise to certain rules that can safely be removed without changing the outcome under each possible expansion. Hence our results relate to the line of research that deals with syntactic rule manipulations in LPs [105, 143].

CHAPTER 9

Discussion

9.1 Results at a Glance

This thesis provides a thorough analysis of argumentation semantics in terms of claims. We introduce a novel class of semantics (hybrid semantics). We study fundamental properties of claim semantics in static and dynamical settings. Below, we give a brief overview over the results presented in this thesis.

Claims are everywhere. In Chapter 3, we survey the role of claims in non-monotonic reasoning formalisms and show that the conclusion-focused evaluation of argumentation frameworks can make the connection to other non-monotonic reasoning formalisms even stronger. Semantics-preserving relations are summarized in Figure 9.1.

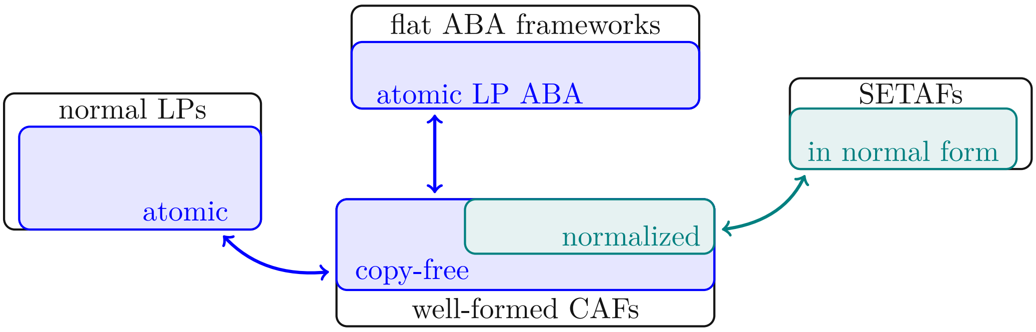
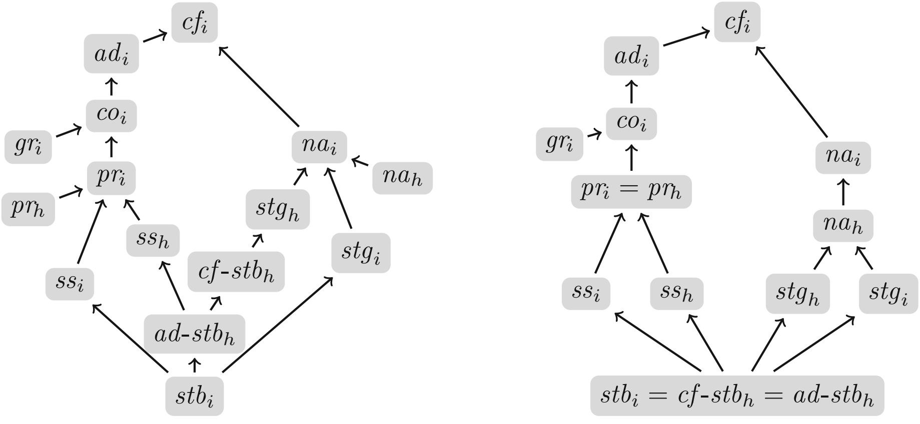


Figure 9.1: Translations between CAFs, ABA frameworks, SETAFs, and LPs, preserving inherited admissible, complete, preferred, grounded, and stable semantics as well as hybrid semi-stable, stage, and naive semantics. Fragments which are (up to isomorphism) in one-to-one correspondence are indicated in blue and teal, respectively.



(a) Relations between CAF semantics. (b) Relations for well-formed CAFs.

Figure 9.2: Relations between semantics for general CAFs (a) and well-formed CAFs (b). An arrow from σ to τ indicates that $\sigma(\mathcal{F}) \subseteq \tau(\mathcal{F})$ for each (well-formed) CAF $\mathcal{F}$.

A hybrid approach. In particular in the abstract argumentation community, the evaluation of justified claims is often considered as byproduct of argument acceptance: claims are extracted in the final step of the procedure after computing the argument-extensions. The claims simply inherit the acceptance status of the corresponding arguments. As we show, these variants yield often unsatisfactory results: they do not satisfy intuitive properties and fail to cover corresponding evaluation methods of formalisms closely related to argumentation. In Chapter 4, we propose semantics that incorporate claims in the evaluation (hybrid semantics). Figure 9.2 presents the relations between inherited (σ_i) and hybrid (σ_h) semantics for CAFs and for well-formed CAFs (cf. Theorem 4.2.2).

Principles. We consider in this work three different categories of principles: meta-principles, basic principles, and set-theoretical principles, the latter being critical for our expressiveness results. Tables 9.1 and 9.2 present selected principles from each category.

We considered the principles *realizability*, *argument-name independence*, *unique realizability*, *language independence*, and *maximal realizability* in Section 5.1.1. Apart from the language independence principle, all considered principles are new. See Tables 5.1 and 5.2 for a complete overview.

In Section 5.1.2, we adapt the principles *conflict-freeness*, *defense*, *admissibility*, *naivety*, *reinstatement*, and *CF-reinstatement* to CAF semantics. We obtain two different versions of naivety. We furthermore consider the novel principle *justified rejection*. We refer to Tables 5.3 and 5.4 for a full overview.

In Section 5.1.3, we consider *I-maximality*, *downward closure*, *tightness*, *conflict-sensitivity*, *cautious closure*, *weak cautious closure*, and *unique completion*, the latter three being novel set-theoretical principles. See Tables 5.5 and 5.6 for a complete overview.

	Unique Real.	Maximal Real.	CF-Reinst.	Just. Reject.	Cautious Closure	w-Cautious Closure	Unique Compl.
cl_c	✗	✗	✗	✓	✗	✗	✗
ad_i	✗	✗	✗	✗	✗	✗	✗
gr_i	✓	✓	✗	✗	✓	✓	✓
co_i	✗	✗	✗	✗	✗	✗	✗
pr_i	✗	✗	✗	✗	✗	✗	✗
pr_h	✗	✗	✗	✗	✓	✓	✓
stb_i	✗	✗	✗	✓	✗	✗	✗
$cf\text{-}stb_h$	✗	✗	✗	✓	✗	✗	✗
$ad\text{-}stb_h$	✗	✗	✗	✓	✗	✗	✗
ss_i	✗	✗	✗	✗	✗	✗	✗
ss_h	✗	✗	✗	✗	✗	✗	✗
na_i	✗	✗	✗	✓	✗	✗	✗
na_h	✗	✗	✓	✓	✓	✓	✓
stg_i	✗	✗	✗	✗	✗	✗	✗
stg_h	✗	✗	✗	✓	✗	✗	✗

Table 9.1: Principles w.r.t. general CAFs (excerpt).

	Unique Real.	Maximal Real.	CF-Reinst.	Just. Reject.	Cautious Closure	w-Cautious Closure	Unique Compl.
cl_c	✗	✓	✗	✓	✗	✗	✗
ad_i	✗	✓	✗	✗	✓	✓	✓
gr_i	✓	✓	✓	✗	✓	✓	✓
co_i	✓	✓	✓	✗	✗	✓	✓
pr_i	✓	✓	✓	✗	✓	✓	✓
pr_h	✗	✓	✓	✗	✓	✓	✓
stb_i	✓	✓	✓	✓	✓	✓	✓
$cf\text{-}stb_h$	✗	✓	✓	✓	✓	✓	✓
$ad\text{-}stb_h$	✗	✓	✓	✓	✓	✓	✓
ss_i	✓	✓	✓	✗	✓	✓	✓
ss_h	✗	✓	✓	✗	✓	✓	✓
na_i	✓	✓	✗	✓	✗	✗	✗
na_h	✗	✓	✓	✓	✓	✓	✓
stg_i	✓	✓	✓	✓	✓	✓	✓
stg_h	✗	✓	✓	✓	✓	✓	✓

Table 9.2: Principles w.r.t. well-formed CAFs (excerpt).

Expressiveness. We study the expressiveness of inherited and hybrid semantics in terms of signatures. For a semantics ρ, the signature of ρ w.r.t. CAFs resp. well-formed CAFs is defined as follows:

$$\begin{aligned}\Sigma_{\rho}^{CAF} &= \{\rho(\mathcal{F}) \mid \mathcal{F} \text{ is a CAF}\}\\ \Sigma_{\rho}^{wf} &= \{\rho(\mathcal{F}) \mid \mathcal{F} \text{ is a well-formed CAF}\}\end{aligned}$$

For general CAFs, our results are as follows (cf. Theorem 5.2.2):

$$\begin{aligned}\Sigma_{gr_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid |\mathbb{S}| = 1\}\\ \Sigma_{cf_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \mathbb{S} \text{ is downwards closed}\}\\ \Sigma_{ad_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \emptyset \in \mathbb{S}\}\\ \Sigma_{co_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \textstyle\bigcap_{S\in\mathbb{S}} S \in \mathbb{S}\}\\ \Sigma_{\rho}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \mathbb{S} \text{ is I-maximal}\},\ \rho \in \{pr_h, na_h\}\\ \Sigma_{\rho}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} = \{\emptyset\} \text{ or } \emptyset \notin \mathbb{S}\},\ \rho \in \{stb_i, cf\text{-}stb_h, ad\text{-}stb_h\}\\ \Sigma_{\rho}^{CAF} &= \Sigma_{stb_c}^{CAF} \setminus \{\emptyset\},\ \rho \in \{pr_i, na_i, ss_i, ss_h, stg_i, stg_h\}\end{aligned}$$

For well-formed CAFs, we obtain the following results (cf. Theorem 5.2.10):

$$\begin{aligned}\Sigma_{gr_i}^{CAF} &= \{\mathbb{S} \subseteq 2^C \mid |\mathbb{S}| = 1\}\\ \Sigma_{cf_i}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \mathbb{S} \text{ is downwards-closed}\}\\ \Sigma_{ad_i}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \emptyset \in \mathbb{S}, \mathbb{S} \text{ is cautiously closed}\}\\ \Sigma_{co_i}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \neq \emptyset, \textstyle\bigcap_{S\in\mathbb{S}} S \in \mathbb{S}, \mathbb{S} \text{ is weak-cautiously closed}\\ &\qquad \text{and satisfies unique completion}\}\\ \Sigma_{\rho}^{wf} &= \{\mathbb{S} \subseteq 2^C \mid \mathbb{S} \text{ is I-maximal}\},\ \rho \in \{stb_i, cf\text{-}stb_h, ad\text{-}stb_h\}\\ \Sigma_{\rho}^{wf} &= \Sigma_{stb_c}^{wf} \setminus \{\emptyset\},\ \rho \in \{pr_i, pr_h, na_h, ss_i, ss_h, stg_i, stg_h\}\end{aligned}$$

In general, we observe that claim semantics are more expressive than AF semantics, even when we restrict ourselves to well-formed CAFs. In particular, for general CAFs, we observe that the signatures admit very soft constraints.

Computational complexity of reasoning. We study skeptical and credulous acceptance, verification of a claim-set, the non-emptiness problem, coherence, and concurrence for CAF semantics. Table 9.3 and Table 9.4 give an overview over the complexity results presented in this work (cf. Chapter 6). We observe a rise in complexity compared to AF semantics in particular for verification in the general case. For hybrid naive semantics, skeptical acceptance is even harder to decide than for inherited naive semantics. Interestingly, for verification, the complexity of h-naive semantics drops compared to i-naive semantics. As our concurrence results show, it is in general quite hard to decide whether two variants of a semantics coincide.

ρ	$Cred_\rho^\Delta$	$Skept_\rho^{CAF}$	$Skept_\rho^{wf}$	Ver_ρ^{CAF}	Ver_ρ^{wf}	NE_ρ^Δ
$ad\text{-}stb_h$	NP-c	coNP-c	coNP-c	**NP**-c	in P	NP-c
$cf\text{-}stb_h$	NP-c	coNP-c	coNP-c	**NP**-c	in P	NP-c
na_h	in P	$\underline{\boldsymbol{\Pi_2^{\mathsf{P}}}\text{-c}}$	**coNP**-c	$\underline{\textbf{DP}\text{-c}}$	in P	in P
pr_h	NP-c	Π_2^{P}-c	Π_2^{P}-c	$\underline{\textbf{DP}\text{-c}}$	coNP-c	NP-c
ss_h	Σ_2^{P}-c	Π_2^{P}-c	Π_2^{P}-c	$\boldsymbol{\Sigma_2^{\mathsf{P}}}$-c	coNP-c	NP-c
stg_h	Σ_2^{P}-c	Π_2^{P}-c	Π_2^{P}-c	$\boldsymbol{\Sigma_2^{\mathsf{P}}}$-c	coNP-c	in P

Table 9.3: Complexity of hybrid semantics for CAFs, $\Delta \in \{CAF, wf\}$. Results that deviate from AF semantics are bold-face; results that deviate from those w.r.t. inherited semantics are underlined.

Con_σ^Δ	pr	na	$\tau\text{-}stb$	ss	stg
$\Delta = CAF$	Π_2^{P}-c	coNP-c	Π_2^{P}-c	Π_3^{P}-c	Π_3^{P}-c
$\Delta = wf$	trivial	coNP-c	trivial	Π_2^{P}-c	Π_2^{P}-c

	$Con_{stb_h}^\Delta$	Coh_τ^Δ
$\Delta = CAF$	Π_2^{P}-c	Π_2^{P}-c
$\Delta = wf$	trivial	Π_2^{P}-c

Table 9.4: Complexity of deciding Con_σ^{CAF}, Con_σ^{wf} (left) and $Con_{stb_h}^\Delta$, Coh_τ^Δ (right).

Ordinary, strong, and renaming equivalence. We study ordinary, strong, and renaming strong equivalence for CAFs in Chapter 7. For ordinary equivalence, we show that, apart from the few dependencies presented in Propositions 7.1.3, 7.1.5 and 7.1.6, equivalence between the different semantics is largely independent. For strong equivalence, we obtain characterizations in terms of kernels, i.e., semantics-dependent sub-frameworks. For any two compatible CAFs $\mathcal{F}$ and $\mathcal{G}$,

$$\mathcal{F} \equiv_s^\rho \mathcal{G} \text{ iff } \mathcal{F}^{csk} = \mathcal{G}^{csk} \text{ for } \sigma \in \{cf\text{-}stb_h, stg_h\} \text{ (Theorem 7.2.15)}$$

$$\mathcal{F} \equiv_s^\rho \mathcal{G} \text{ iff } \mathcal{F}^{sk} = \mathcal{G}^{sk} \text{ for } \sigma \in \{stb_i, stg_i\} \text{ (Theorem 7.2.16)}$$

$$\mathcal{F} \equiv_s^\rho \mathcal{G} \text{ iff } \mathcal{F}^{ak} = \mathcal{G}^{ak} \text{ for } \sigma \in \{ad_i, pr_i, pr_h, ss_i, ss_h, ad\text{-}stb_h\}$$
$$\text{(Theorems 7.2.16, 7.2.23 and 7.2.24)}$$

$$\mathcal{F} \equiv_s^{co_i} \mathcal{G} \text{ iff } \mathcal{F}^{ck} = \mathcal{G}^{ck} \text{ (Theorem 7.2.16)}$$

$$\mathcal{F} \equiv_s^{gr_i} \mathcal{G} \text{ iff } \mathcal{F}^{gk} = \mathcal{G}^{gk} \text{ (Theorem 7.2.16)}$$

$$\mathcal{F} \equiv_s^\rho \mathcal{G} \text{ iff } \mathcal{F}^{nk} = \mathcal{G}^{nk} \text{ for } \sigma \in \{cf_i, na_i, na_h\} \text{ (Theorems 7.2.16 and 7.2.24)}$$

For well-formed CAFs, we have $\mathcal{F}^{csk} = \mathcal{G}^{csk}$ iff $\mathcal{F}^{sk} = \mathcal{G}^{sk}$ (cf. Proposition 7.2.26). Moreover, renaming strong equivalence is characterized by kernel isomorphism:

$$\mathcal{F} \equiv_s^\rho \mathcal{G} \text{ iff } \mathcal{F}^{k(\rho)} \text{ and } \mathcal{G}^{k(\rho)} \text{ for are isomporphic (cf. Theorem 7.3.9).}$$

Here, $\mathcal{F}^{k(\rho)}$ denotes the kernel which characterizes strong equivalence w.r.t. semantics ρ.

Our computational complexity results regarding all considered equivalence notions for CAFs are summarized in Table 9.5.

	gr_i	pr_i	pr_h	α	β
$\textsc{Ver-OE}_\rho$	in P	Π_3^{P}-c	Π_2^{P}-c	Π_2^{P}-c	Π_3^{P}-c
$\textsc{Ver-OE}_\rho^{wf}$	in P	Π_2^{P}-c	Π_2^{P}-c	coNP-c	Π_2^{P}-c
$\textsc{Ver-SE}_\rho$	tractable (in P) for all semantics				
$\textsc{Ver-SER}_\rho$	complexity of graph isomorphism				
$\textsc{Ver-SER}_\rho^{wf}$	tractable (in P) for all semantics				

Table 9.5: Complexity of deciding ordinary equivalence for CAFs ($\textsc{Ver-OE}_\rho$) and well-formed CAFs ($\textsc{Ver-OE}_\rho^{wf}$), strong equivalence for CAFs ($\textsc{Ver-SE}_\rho$), and renaming strong equivalence for CAFs ($\textsc{Ver-SER}_\rho$) and well-formed CAFs ($\textsc{Ver-SER}_\rho^{wf}$) where $\alpha \in \{cf_i, ad_i, co_i, na_i, na_h, stb_i, cf\text{-}stb_h, ad\text{-}stb_h\}$ and $\beta \in \{ss_i, ss_h, stg_i, stg_h\}$.

cvAFs - an instantiation for dynamics. In Chapter 8, we introduce a framework that captures the dynamical behavior of instantiations whose attack relation is always well-formed. We introduce *cvAFs* (claim and vulnerability augmented AFs) which generalize CAFs by additionally having information about the vulnerabilities of the arguments. We assume that cvAFs are always well-formed. We focus on complete, grounded, preferred, and stable semantics. We study claim enforcement and strong equivalence for this framework. Our strong equivalence results are as follows: For any two cvAFs $\mathfrak{F}$ and $\mathfrak{G}$,

$$\mathfrak{F} \equiv_s^{co} \mathfrak{G} \text{ iff } \mathfrak{F}^{\mathsf{ck}} = \mathfrak{G}^{\mathsf{ck}} \text{ (cf. Theorem 8.4.19)}$$
$$\mathfrak{F} \equiv_s^{gr} \mathfrak{G} \text{ iff } \mathcal{F}^{\mathsf{gk}} = \mathcal{G}^{\mathsf{gk}} \text{ (cf. Theorem 8.4.36)}$$
$$\mathfrak{F} \equiv_s^{pr} \mathfrak{G} \text{ iff } \mathcal{F}^{\mathsf{pk}} = \mathcal{G}^{\mathsf{pk}} \text{ (cf. Theorem 8.4.28)}$$
$$\mathfrak{F} \equiv_s^{stb} \mathfrak{G} \text{ iff } \mathcal{F}^{\mathsf{sk}} = \mathcal{G}^{\mathsf{sk}} \text{ (cf. Theorem 8.4.47)}$$

In contrast to the kernels for CAFs which are constructed by removing redundant attacks, the kernels for cvAFs identify arguments that are redundant w.r.t. a specific semantics. This is unique in the literature on strong equivalence in argumentation.

Our enforcement results are as follows (cf. Theorem 8.3.9, Proposition 8.3.12, and Corollary 8.3.13): an argument is enforceable w.r.t. semantics $\sigma \in \{co, pr, stb\}$ iff it is not strongly unacceptable and it is *gr*-enforceable it is not strongly *gr*-unacceptable. An argument is strongly unacceptable iff it is attacked by an argument having the same or less vulnerabilities. Hence it is impossible to defend this argument without introducing new attackers. Strong *gr*-unacceptability weakens strong unacceptability by ignoring the argument's claim in the set of vulnerabilities.

Results for ABA and LPs. We can use the translations presented in Chapter 3 to transfer results for CAFs to ABA and LP. We obtain (a) signature results for the class of normal LPs and the class of flat ABA frameworks with separated contraries; (b) complexity results for atomic LP ABA and atomic LPs (i.e., for the fragments in one-to-one correspondence, cf. Figure 9.1); (c) concise representations of the instantiations (using our strong equivalence results for CAFs). Moreover, the principles and properties from Chapter 5 can yield new insights as well.

In Chapter 8, we furthermore show that enforcing a conclusion (an assumption) and deciding strong equivalence for flat, atomic ABA frameworks is intractable (cf. Theorem 8.1.4 and 8.1.7). Likewise, we show that enforcement for LPs is intractable as well (cf. Proposition 8.5.3). Using our cvAF results we obtain the following tractability results:

- deciding whether an argument or conclusion is enforceable for atomic flat ABAs with separated contraries is tractable (cf. Theorem 8.3.15);
- deciding strong equivalence in the class of atomic, flat ABA frameworks with separated contraries is tractable (cf. Theorem 8.4.48);
- deciding atom enforcement in the class of atomic LPs is tractable (cf. Theorem 8.5.4);
- deciding strong equivalence in the class of atomic LPs is tractable (cf. Theorem 8.5.6);
- deciding whether two atomic LPs P and Q are strongly equivalent w.r.t. stable semantics in the class of normal LPs is tractable (cf. Corollary 8.5.9).

9.2 Related Work

9.2.1 Generalized Instantiations

Both CAFs and cvAFs constitute a generalization of Dung's abstract frameworks. In the last decades, several generalizations of AFs have been proposed [49]. We have already mentioned AFs with collective attacks (SETAFs) [141] and Abstract Dialectical Frameworks [50] due to their connection to CAFs (cf. Chapter 3). Other generalizations allow for uncertain attacks and arguments [30, 31], recursive attacks [14], the incorporation of preferences [125] or values [34], or consider a support relation at the abstract level [64, 63]. All of the aforementioned formalisms capture generalized scenarios and extend the modeling capacities of AFs. CAFs and cvAFs, on the other hand, establish a closer connection to the underlying knowledge base by keeping more information about the structure of the arguments on the abstract level. In contrast to structured formalisms, we do not equip our frameworks with argument construction capabilities. Although CAFs and cvAFs incorporate elements from structured argumentation by taking the claim and vulnerabilities into account, both models belong to the abstract argumentation family.

In particular in the context of structured argumentation, there are several abstract representations which generalize AFs by keeping track of parts of the argument structure. Several models in the literature on structured argumentation identify arguments with pairs (X, φ) where X is the support and φ is the claim of the argument. We mention argumentation-based models of defeasible logic programming [113], logic-based argumentation [40] and their cores [8]. A similar representation of arguments appears in the context of ABA and dynamic argumentation frameworks in which arguments are identified with pairs consisting of premises and claim [159, 165]. In contrast to these representations, cvAFs consider the vulnerabilities of the arguments. Corsi and Fermüller consider *semi-abstract argumentation frameworks* [67, 68] which assign each argument a logical formula that represents the claim of the argument. Their model is closely related to CAFs, however, in CAFs, no particular structure of the claim is assumed. Using the additional structure of the claim, Corsi and Fermüller identify attack rules and establish a logic of argumentation in their work. In [15], Baroni et al. consider an *argument-conclusion structure* which consists of a language, a set of arguments, and a set of claims (i.e., they disregard the attack relation). The argument-conclusion structure is used to model claim-labellings. We discuss this in more detail in the next subsection.

9.2.2 Floating Conclusions

The difference between argument and claim acceptance has evoked several discussions among researchers in the area of non-monotonic reasoning. Early discussions in that matter focus on inheritance networks[1] and can be traced back to the late eighties [121, 122, 166, 169]. In the center of these discussions are *floating conclusions*, as termed by Makinson and Schlechta [136], which are claims that appear as conclusions of different extensions. The skeptical evaluation of floating conclusion has evoked several debates throughout the non-monotonic reasoning community. Generally speaking, an element of a defeasible theory is *skeptically accepted* if it appears in each possible outcome. In abstract argumentation, an argument is skeptically accepted if it is contained in each extension. However, as Stein points out, “there are facts which are true in all credulous extensions, but which have no justification in the intersection of those extensions.” [169]

Should these claims be skeptically accepted? In the scope of a debate on floating conclusions between John F. Horty and Henry Prakken in the early two-thousands, the issue has also gained attention in the argumentation community, shedding light on the heterogeneous approaches to handle claim acceptance in argumentation. As discussed in [120], to identify the set of skeptically accepted claims, one could either (1) compute the skeptically accepted arguments and extract their claims; or (2) compute the extensions of the argumentation system, extract the claims, and then compute the intersection of the claims to obtain all skeptically acceptable statements of the system. Hence approach (2) accepts floating conclusions while approach (1) does not. Approach (1), in the context of inheritance networks also described as *directly skeptical* [122], is used in ASPIC+ and appears also in Pollock’s “Defeasible reasoning.” [149, Section 3]. The approach led to several criticisms in subsequent work, pointing out the intrinsic inability to handle floating conclusions. In [136], Makinson and Schlechta propose an alternative way to compute skeptically accepted conclusions by commuting the evaluation order which in turn corresponds to approach (2). This more generous approach is questioned by Horty [120] by presenting

---

[1] Roughly speaking, an inheritance network is a graph with positive and negative inference links between categories of descending specificity, formalizing positive and negative inheritance between categories such as Animal, Bird, Penguin, Flying subject, etc. We refer to [174] for an overview.

several examples in which the acceptance of floating conclusions is deemed unintuitive. Prakken suggested in [150] that a more appropriate modeling of the controversial instances could help to avoid the aforementioned issues.

In the present work, we adapt the more generous approach to claim acceptance and accept floating conclusions. This treatment of floating conclusions can be witnessed in several non-monotonic reasoning formalisms. Apart from the adaption in inheritance networks, approach (2) furthermore appears in ABA and in Reiter's default logic [160]. We show that this approach admits the same complexity for all apart of naive semantics like the AF counterparts. In this regard, it would be interesting to study the complexity of approach (1) in more detail.

We furthermore mention an interesting user study related to this context. In 2010, Rahwan et al. [156] conduced a user study on the issue of *floating reinstatement*, a problem in similar spirit as floating conclusions. The objective was to test the plausibility of floating reinstatement compared to standard reinstatement. An argument is said to be *reinstated* if it is defended against its attacker; in its simplest form, the reinstating argument is unattacked. Floating reinstatement, on the other hand, considers a slightly more involved situation in which two reinstating arguments are mutually attacking. Interestingly, the study suggests that both methods of reinstatement are considered equally plausible. Although similar studies directly addressing floating conclusions still need to be carried out, the results of this experimental evaluation indicate that humans have less issues with switching justifications as suspected.

9.2.3 Labelling Statements

The justification of statements and in particular the different possibilities to consider a statement acceptable has been studied in the context of labelling-based semantics. Labelling-based semantics constitute an alternative approach to evaluate the acceptance status in argumentation frameworks [124, 54, 56]. Although the main focus in the literature lies on argument-labellings, there is some work on claim-labellings. Caminada et al. [61] and, in recent work, also Rocha and Cozman [163] consider claim-labellings in the context of translations between logic programs and argumentation. They identify differences between argument- and claim-labellings when it comes to maximization. Baroni et al. [15] and, in subsequent work, Baroni and Riveret [12] propose a multi-labelling system that admits several stages of labellings. They identify two main approaches for statement justification: the argument-focused and the conclusion-focused approach.

Before diving into detail, let us give a short recap on labelling-based semantics.

Labelling-based semantics first appear in [124, 54] and have further been developed in numerous subsequent work, see, e.g., [56, 58]. A *labelling* is a function $\lambda : A \to \mathcal{L}$ which assigns each argument a label, indicating its acceptance status. Three-valued labellings $\mathcal{L} = \{\mathsf{in}, \mathsf{out}, \mathsf{undec}\}$ give rise to labelling-based counter-parts of all extension-based semantics considered in this work. Intuitively, an argument is accepted if it is assigned the label in, rejected if it is assigned the label out, and undecided if it is assigned the label

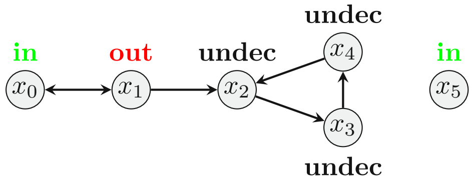


Figure 9.3: Complete labelling for AF $F = (A, R)$ (cf. Example 2.1), corresponding to the complete extension $\{x_0, x_5\}$.

undec. Each extension-based semantics σ under consideration can be characterized via labellings that satisfy certain characteristics. As an example, we recall the definition of a *complete labelling* (we refer to [16] for a comprehensive overview): For an AF $F = (A, R)$,

- $a \in A$ is labelled in iff each attacker is labelled out;
- $a \in A$ is labelled out iff there is some attacker that is labelled in.

Each complete extension corresponds to a complete labelling. Figure 9.3 gives an example of a complete labelling of an AF.

Lifting argument-labellings to claim-labellings. To investigate the relation between logic programming and argumentation semantics, Caminada et al. [61] lift labels of arguments to labels of claims by selecting the 'best' label among all arguments with the same claim according to the order in > undec > out.

Example 9.2.1. *Consider the following claim assignment our running example: $cl(x_0) = cl(x_3) = c$, $cl(x_1) = d$, $cl(x_2) = e$, $cl(x_4) = f$, and $cl(x_5) = g$. According to the order of the labellings, claim c is labelled* in*, d is labelled* out*, and e and f are labelled* undec.

In their work, Caminada et al. compare maximization of argument- and claim-labellings of their LP instantiation. They show that maximization of the in-labelled part under complete semantics yields the same result while maximization of the undec-labelled part does not. The former shows that preferred AF semantics and regular LP semantics coincide while the latter reveals the difference between semi-stable and L-stable semantics. As discussed in Section 3.2, they show that complete, grounded, preferred and stable semantics of both formalisms correspond to each other while semi-stable and L-stable semantics do not coincide.

Interestingly, when adapting the claim-labels to CAFs we obtain a labelling-based formalization of hybrid complete, grounded, preferred, *ad*-stable, and semi-stable semantics: given a complete argument extension E and a claim c, then

- claim c is accepted iff there is some argument a with claim c such that $a \in E$ iff there is some argument with claim c that is labelled in;

- claim c is defeated iff E attacks each argument with claim c iff each argument with claim c has label out.

The labels characterize complete semantics. When maximizing the in-labelled part, we obtain h-preferred semantics; minimizing the in-labelled part yields h-grounded semantics; minimizing the undec-labelled part yields h-semi-stable semantics; and requiring that the set of undec-labelled claims is empty yields h-ad-stable semantics. In fact, this has been recently addressed in the work of Rocha and Cozman: in [162], they adapt h-semi-stable semantics to probabilistic argumentation and study the connection to probabilistic logic programs; in [163], they generalize CAFs by incorporating a support relation and show that this model 1-1 corresponds to normal LPs.

As the attentive reader might have noticed, our results regarding maximization of in-labelled part differ from the results obtained by Caminada et al.: we show that h-preferred semantics and i-preferred semantics do not coincide. Likewise, we show that h-ad-stable and i-stable semantics differ. The reason for this mismatch is that Caminada et al. exclusively focus on LP instantiations which yield well-formed frameworks as we have discussed in Section 3.2. In well-formed frameworks, the aforementioned semantics coincide. However, they fail to yield the same results in the general case. Hence our results reveal that the observations made by Caminada et al. crucially depend on the graph-structure. The coincidence between preferred and stable semantics is only obtained when the attack relation satisfies well-formedness.

Moreover, due to their focus on LP semantics, Caminada et al. consider exclusively complete-based semantics in their work. It would be interesting to study the labelling-based approach also for admissible semantics and for semantics that are based on conflict-freeness. We expect that the lifting operation proposed by Caminada et al. can be adapted to the labelling-based versions of conflict-free and admissible semantics. We believe that research on this matter can yield valuable insights.

Multiple stages of labellings. Baroni et al. [15, 12] generalize labellings to multiple stages. They use their model to demonstrate that the argument-focused and the conclusion-focused approach differ from each other. In [15] they consider a multi-labelling model that consists of four stages: the argument acceptance stage, argument justification stage, claim acceptance stage, and claim justification stage. Intuitively, the argument acceptance stage corresponds to the labelling-based semantics as considered in the literature. The argument justification stage captures different reasoning modes such as credulous and skeptical reasoning: these acceptance notions can be interpreted as (partial) labelling-function $\mu : 2^{\mathcal{L}} \rightarrow \{\mathsf{skep}, \mathsf{cred}\}$ that assigns labels skep or cred to sets of labels in $\mathcal{L}$. The claim acceptance and claim justification stages can be seen as claim-focused counter-parts. Similar to our work, Baroni et al. consider the evaluation of claims as independent and as the final step of the procedure. They discuss the different ways that can lead to the claim justification stage. Figure 9.4 gives an overview over the different stages. Having settled the acceptance of arguments (stage AA) one can either

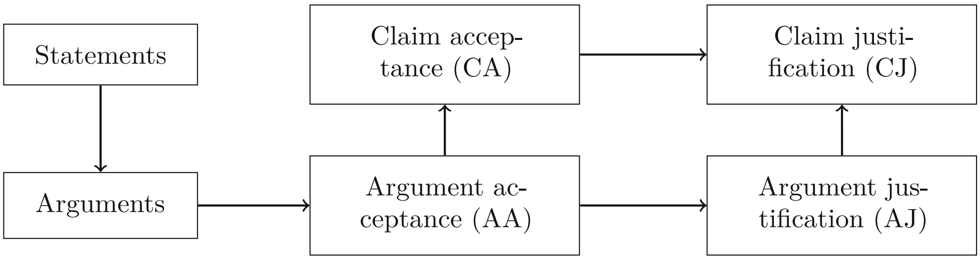


Figure 9.4: Different paths to claim justification. Adapted from [15].

1. identify justified arguments (stage AJ) and extract justified claims (stage CJ); or
2. identify accepted claims (stage CA) and extract justified claims (stage CJ).

Their model captures the different ways to treat floating conclusions. Approach 1 accepts only conclusions that correspond to skeptically accepted arguments while approach 2 accepts floating conclusions as well. They compare their model to different formalisms and show that ASPIC+ employs approach 1 while ABA follows approach 2. They show in particular that the two approaches are incomparable. In [12], Baroni and Riveret extend the model to arbitrary labelling stages.

Although the work by Baroni et al. has a somewhat different focus, the overall idea behind their approach is certainly related to the present thesis. In the same spirit of the present work, they put the main emphasis on statement justification. Moreover, they discover a fundamental difference between argument-focused and conclusion-focused approaches to statement justification. Although they did not consider different methods to obtain claim semantics in their work, we believe that multi-labelling systems are a powerful tool to investigate argument-focused and conclusion-focused evaluation methods on a very general level. Here, one could, for instance, consider the evaluation of a semantics itself as a multi-stage process (e.g., treating maximization of admissible sets as separate step which can be performed on argument- or on claim-level).

9.2.4 Analysis of Claim Semantics

Principles, postulates and properties of argumentation semantics have been considered in different facets for different (structured and abstract) argumentation formalisms, e.g., [11, 175, 116, 5, 57, 60, 101]. Likewise, expressiveness of argumentation semantics is an important topic that has been considered for different abstract formalisms [84, 85]. In contrast to most of the aforementioned works which investigate principles and expressiveness in terms of arguments, our studies focus on semantics in terms of claims. While there is naturally a close correspondence if not dependence between these two viewpoints the differences are considerable as shown in the present work. We also want to highlight in this regard in particular the work by Amgoud, Caminada, Gorogiannis, and Hunter [5, 57, 116] which study rationality postulates for logic-based argumentation systems also in terms of the conclusion-focused outcome. In contrast to our analysis they focus on consistency and closure properties. In our work, claims are considered abstract in order to investigate structural properties of the outcome.

We furthermore mention results on expressiveness and principle-based investigations for AFs with collective attacks (SETAFs): As noted in Remark 5.2.11 and discussed in Section 3.4 and 4.3.4, well-formed CAFs and SETAFs are closely related. On the one hand, we thus obtain an alternative characterization of the signatures for well-formed CAFs from signature results presented in [85]. In particular, we obtain that the respective properties coincide, i.e., set-conflict-sensitivity coincides with cautious closure and set-com-closure is equivalent to weak cautious closure and unique completion. While set-conflict-sensitivity and set-com-closure are formalized in terms of potential conflicts our formulations are conflict-independent and yield an alternative view on the SETAF characterizations. On the other hand, the close relation between well-formed CAFs and SETAFs reveals interesting parallels between our principle-based analysis for well-formed CAFs and the principle-based analysis of SETAF semantics recently conducted in [101]. Indeed, we obtain similar results regarding the common principles we investigated, i.e., for conflict-freeness, defense, admissibility, (CF-)reinstatement, h-naivety, and I-maximality. Apart from these principles, they put their focus on the investigation of modularization, non-interference principles, and SCC-recursiveness utilizing the so-called reduct [25], while we conducted set-theoretical investigations and considered genuine principles for claim-focused reasoning.

Complexity-theoretic considerations are well-established in the argumentation community and have a long-standing tradition, we mention [82, 178, 90, 74] and refer to [89] for an overview. The main focus here lies, however, on reasoning questions that deal with arguments; apart from recent work regarding the computational complexity of inherited CAF semantics [92] the vast majority considers complexity-theoretic questions in terms of argument acceptability.

9.2.5 Dynamics

Our work extends research on dynamics in argumentation. In the last decades, researchers explored several different directions, including strong equivalence [142, 112, 19, 8, 104, 21], enforcement [18, 180, 46], argument revision [109, 65, 168], and, in general, changes of the knowledge base or the abstract representation [179, 164, 165, 108]. We also refer to [28, 76] for an overview in that matter.

Strong Equivalence. Our characterization results for strong equivalence for CAFs and cvAFs are in line with existing studies for related abstract argumentation formalisms which provide characterization results of strong equivalence in terms of kernels (cf., e.g., [142, 112, 93]). In this matter, we furthermore mention Baumann and Strass [21] who provide logic-based characterization results of strong equivalence in non-monotonic knowledge representation formalisms (in similar spirit to the characterization of strong equivalence for logic programs in terms of the logic of here-and-there [132]). Here, we furthermore mention equivalence characterizations of answer set programming semantics [107]. Notable are also the strong equivalence characterizations of labelling-based semantics [20] which have revealed subtle differences to the extension-based approach. Moreover, strong equivalence is similar in spirit to *stability* [171].

Although many CAF semantics can be reduced to testing kernel equivalence for AF semantics, we have seen that this is not always the case. Interestingly, we show that for cvAFs, kernels are constructed by removing arguments instead of attacks. This in indeed unique for abstract formalisms. However, in logic-based approaches, a similar behavior has been observed by Amgoud et al. [8]: they show that under certain conditions on the underlying logic, unnecessary arguments can be removed while retaining (strong) equivalence. In contrast to their work, our studies are independent of the underlying formalism of the instantiated argumentation system as we do not impose any further constraints on the arguments, their vulnerabilities, or their claims.

Enforcement. Enforcement has received much attention in the abstract argumentation community in recent years, we refer to, e.g., [180, 18, 179]. Enforcement in AFs is typically easy to characterize; often, research in this matter takes certain minimality criteria into account. Also in the context of structured argumentation, the topic has received quite some attention. In a recent paper [46] the authors study under which conditions in a structured argumentation formalism a given formula can be enforced. A type of conclusion enforcement appears also in the context of defeasible logic programming. The authors in [138] consider argumentative revision operators in the context of defeasible logic programming in order to warrant a desired conclusion. In contrast to our enforcement approach, their objective lies in revising a program such that an argument with the desired conclusion ends up undefeated.

Similar to our setting, [179] considers situations where an AF undergoes certain changes, but the permitted modifications are constrained. Constraints on the possibly reachable expansions of a given cvAF are intrinsic to our approach. Wallner [179] considers several different types of constraints and dynamic operators, also in connection with assumption-based argumentation. He considers an AF instantiated from a knowledge base and establish suitable enforcement operators for AFs that respect the underlying knowledge base. In contrast to our approach which focuses on establishing existence criteria, he derives certain constraints the AF instantiated from the ABA framework must satisfy, moreover, they consider minimal changes of the knowledge base. It is shown that deciding whether a set of arguments is enforceable in a given AF which corresponds to a particular knowledge base is intractable while we show intractability of deciding enforcement for native ABA elements (i.e., for assumptions and conclusions).

Revising knowledge. Our considerations regarding dynamic changes of knowledge bases is related to certain operations in the area of belief revision which deals, broadly speaking, with changing beliefs in the light of new information [4, 147]. The AGM model of belief revision [3] is among the most influential approaches. Epistemic states are modeled by belief sets over a formal language; basic changes include expansions, contractions, and revisions of a belief set. There are several relations between argumentation and belief revision [42]; we refer to [109] for an overview.

As pointed out in [108], the AGM model distinguished between changes at the knowledge level and the symbol level. Different knowledge bases might yield the same representation at the symbol level. Although they represent the same belief in the static sense, they could be different when it comes to dynamic changes. We note that this observation is in line with our considerations in Chapters 7 and 8. We observe that the (C)AF representation is insufficient to capture changes of the underlying knowledge base. We outline several obstacles of the abstract representation of ABA instantiations in Chapter 8. There is, however, an interesting difference to our work: in contrast to the AGM model where changes are studied at knowledge level, we overcome these obstacles by tuning the abstract model until it correctly represents the changes of the underlying knowledge base. With our cvAFs, we were able to identify a fragment of ABA and LP for which the symbol level captures the considered modifications of the knowledge level.

There is some work which deals with revising knowledge specifically in argumentation. Snaith and Reed [168] consider revision operations in ASPIC+. Falappa et al. [108] study changes in logic-based argumentation systems and how the modification of strict to defeasible rules gives rise to the changing of arguments and their attack relation. Hadjisoteriou and Kakas [118] develop a logic-based framework capable to express logic-based reasoning about actions and change. Pandzic [144] defines dynamic operations for default theories with justification formulae. Rotstein et al. [164, 165] consider a framework specifically designed for handling dynamic changes in argumentation through the consideration of varying evidences. They develop dynamic argumentation frameworks which keep track of the structure of the arguments and their sub-argument relation even at the abstract level and hence we observe certain parallels to our cvAFs. They consider the addition and the removal of arguments and study associated interactions.

Our work overlaps in particular with a certain aspect of argument revision, namely the expansions of knowledge bases. The basic assumption here is that existing knowledge remains and new information is integrated in existing beliefs. While all of the aforementioned work consider the addition of new information to a certain extent, we want to highlight here in particular the work by Cayrol et al. [65] who study framework expansions in the context of AFs. They focus on the addition of a new argument to an AF which may interact with existing arguments. In their work, they consider several different types of revision operators that impose certain properties of the outcome (for instance, decisive revision assumes that there is only one set of acceptable arguments after the revision) and establish conditions under which a given property is satisfied.

Redundancies. The redundancy notions that we discussed for LPs are similar in spirit to the line of research on syntactic transformations for LPs, see., e.g., [47, 105, 183, 134], that gave rise to alternative characterizations of strong equivalence [143, 51] and set the ground for further complexity analysis of LP fragments [106].

9.3 Future Work

We identify several directions for future work.

We consider further studies on interlinking the argument- and claim-level as one of the most promising future work directions. Although we are convinced that hybrid semantics as defined in this work constitute a reasonable way to incorporate claims into the evaluation procedure, we believe that alternative ways are worth investigating. Here, an in-depth study of conclusion-focused evaluation methods in related formalisms would be a promising starting point. Likewise, we believe that the work by Baroni et al. [15] offers interesting possibilities in this matter. As briefly mentioned in Section 9.2.3, the multi-stage process could be integrated in the evaluation process of a framework. In this way, we could obtain a formulation of CAF semantics that captures the differences between inherited and hybrid semantics. For preferred semantics, for instance, we can split the evaluation into the following stages: (AA) argument acceptance under admissible semantics, (AM) argument-maximization, (CM) claim-maximization, and (CA) claim acceptance. Then, the path (AA)–(AM)–(CA) corresponds to i-preferred semantics while the path (AA)-(CA)-(CM) corresponds to h-preferred semantics. Regarding the range-based semantics, defeat must be treated separately, giving rise to two different stages (CD) claim-defeat and (AD) argument-defeat. The path (AA)–(AD)–(CA) corresponds to i-stable semantics while the path (AA)-(CD)-(CA) corresponds to h-stable semantics. We consider these and similar investigations as an interesting avenue for future work.

Another promising direction is the investigation of hybrid semantics in relation with ASPIC and other structured argumentation formalisms, in particular the semantics which make use of claim-defeat. We believe that this concept could be of value to structured argumentation and to instantiation-based approaches since it opens novel possibilities in the evaluation, that have been, to the best of our knowledge, not yet formalized so far.

Moreover, extending our investigations to further concepts used in argumentation semantics would be a promising endeavor. In this regard, we consider studies on other inherited semantics based on e.g., strong or weak admissibility [11, 26] worth investigating.

Another interesting but yet completely unexplored research direction is the development of argumentation semantics which include both arguments and claims. So far, all considered semantics return either sets of jointly acceptable arguments or jointly acceptable claims. Indeed, as we have formalized in our argument-names independence principle, the particular arguments for a claim are theoretically exchangeable (as long as their relation within the fragment stays the same). Here, we believe that CAFs can provide mediating approaches by taking both the claims as well as arguments into account.

The present work provides first insights into the advantages a principle-based analysis of claim-focused argumentation semantics can offer. As already expressed in Chapter 5, the principles and properties formulated in this work capture the behavior of the considered CAF semantics to a different extent; in particular inherited semantics in unrestricted CAFs lack principles that characterize their distinct behavior. An important future work direction in this regard is thus to deepen the principle-based analysis on inherited semantics. Another challenging avenue for future work is to adapt more classical AF principles to the realm of claim-focused reasoning. Although the principle-based investigation we conducted in this work already collects many of the classical principles that have been considered in the literature there are a lot of other principles left that are worth studying in the context of claims (e.g., directionality and non-interference principles [11, 60]).

We point out that the characterization of the expressiveness of i-naive semantics remains an open problem. So far, we have shown that i-naive semantics is more expressive than the AF semantics corresponding to it. However, it is not possible to express arbitrary claim-extension-sets, as Lemma 5.2.34 demonstrates.

In the present work, we filled in several gaps regarding the computational complexity of claim-centered reasoning. There are, however, several promising future work directions in this matter, including the computational complexity of dealing with incomplete information on the arguments and attacks [31, 110], the problem of counting the number of extensions [13, 111], or enforcing the acceptance of a statement or a set of statements respecting certain minimality constraints [28, 180].

We also consider further studies regarding structural restrictions imposed by the attack relation as an interesting future work direction. In this regard, we mention recent and ongoing work on the impact of preferences on claims and the structure of CAFs [38, 39] where we study the effect of preference incorporation and identify novel CAF classes that lie between well-formed and general CAFs.

Structured argumentation plays a significant role in the present work. In this thesis, we focus on the instantiation-based representation of structured argumentation; in particular, we assume a tree-like structure of arguments. There are, however, several alternative approaches to evaluating the acceptability of claims in structured argumentation that feature different representations of the knowledge base or argument structures, e.g., in form of dependency graphs or dialectical procedures [78, 172, 69]. Most notably in our context is the work by Craven and Toni who propose a graph-based representation of ABA arguments [69]. Their compact representation of arguments as directed, acyclic graphs which represent the dependencies among literals in a knowledge base generalizes the traditional tree-based representation. Each so-called *argument graph* corresponds to several tree-based arguments. Interestingly, each such graph contains only a single argument for each claim [69, Theorem 4.12]. We consider further studies on the connection between CAFs and argument graphs an interesting avenue for future research.

Regarding our investigations of the dynamic setting, we identify several future work directions.

First, we consider exploring further formalisms where cvAFs are applicable, i.e., investigating suitability for e.g. ASPIC [137] or logic-based argumentation [40], a promising future work direction. As demonstrated in our LP section, this technique may lead to quickly obtained results.

Second, we want to extend our results for cvAFs to further semantics. So far, we have only considered the complete-based Dung semantics (i.e., complete, grounded, preferred, and stable semantics). It would be in particular interesting to see whether our kernels suffice to characterize h-semi-stable and i-admissible semantics (bearing in mind that these semantics are the correct choice to capture the corresponding ABA semantics).

Third, we consider studies on revision operators in the spirit of the AGM model a promising future work direction. Finding more reasoning tasks where cvAFs are applicable might contribute to this line of research. As a further future research direction in this matter we identify the design of efficient algorithms since our tractability results might serve as a promising starting point for such an endeavor.

Moreover, our cvAF results may be leveraged to study learning in rule-based (argumentation) formalisms (cf. [139, 73, 152]). In particular, we observe a connection between our cvAF enforcement results and the task of covering positive examples. Here, our (im-)possibility criteria could serve as a promising starting point to identify conditions under which a desired result can be achieved.

In relation to our strong equivalence results, we identify the study of parameterized equivalence notions as a promising endeavor for future work. Baumann et al. [24] consider an equivalence notion parameterized w.r.t. a set of core arguments which are not affected by the expansions. We consider similar studies for cvAFs parameterized w.r.t. a set of arguments or claims worth investigating.

Finally, it would also be interesting to utilize cvAFs to identify syntactic simplifications in knowledge bases. As already discussed in the context of LPs, our redundancy notions for cvAFs can be used on knowledge base level to obtain rule redundancies. Here, we believe that our cvAFs can serve as link to transfer established results on the abstract level, for instance, replacement patterns in AFs [86], to the knowledge base level.

Bibliography

[1] J. F. L. Alcântara and S. Sá. Equivalence results between SETAF and attacking abstract dialectical frameworks. In *Proceedings of the 19th International Workshop on Non-Monotonic Reasoning (NMR 2021)*, pages 139–148, 2021. URL `https://drive.google.com/open?id=1WSIl3TOrXBhaWhckWN4NLXoD9AVFKp5R`.

[2] J. F. L. Alcântara, S. Sá, and J. C. A. Guadarrama. On the equivalence between abstract dialectical frameworks and logic programs. *Theory and Practics of Logic Programming*, 19(5-6):941–956, 2019. doi:10.1017/S1471068419000280.

[3] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50 (2):510–530, 1985. doi:10.2307/2274239.

[4] J. J. Alferes, J. A. Leite, L. M. Pereira, H. Przymusinska, and T. C. Przymusinski. Dynamic updates of non-monotonic knowledge bases. *The Journal of Logic Programming*, 45(1-3):43–70, 2000. doi:10.1016/S0743-1066(99)00065-5.

[5] L. Amgoud. Postulates for logic-based argumentation systems. *International Journal of Approximate Reasoning*, 55(9):2028–2048, 2014. doi:10.1016/j.ijar.2013.10.004.

[6] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM 2010)*, volume 6379 of *Lecture Notes in Computer Science*, pages 42–55. Springer, 2010. doi:10.1007/978-3-642-15951-0_10.

[7] L. Amgoud and P. Besnard. Logical limits of abstract argumentation frameworks. *Journal of Applied Non-Classical Logics*, 23(3):229–267, 2013. doi:10.1080/11663081.2013.830381.

[8] L. Amgoud, P. Besnard, and S. Vesic. Equivalence in logic-based argumentation. *Journal of Applied Non-Classical Logics*, 24(3):181–208, 2014. doi:10.1080/11663081.2014.959332.

[9] S. Arora and B. Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009. ISBN 978-0-521-42426-4. URL `http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521424264`.

[10] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. R. Simari, M. Thimm, and S. Villata. Towards artificial argumentation. *AI Magazine*, 38(3):25–36, 2017. doi:10.1609/aimag.v38i3.2704.

[11] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007. doi:10.1016/j.artint.2007.04.004.

[12] P. Baroni and R. Riveret. Enhancing statement evaluation in argumentation via multi-labelling systems. *Journal of Artificial Intelligence Research*, 66:793–860, 2019. doi:10.1613/jair.1.11428.

[13] P. Baroni, P. E. Dunne, and M. Giacomin. On extension counting problems in argumentation frameworks. In *Proceedings of the 3rd Conference on Computational Models of Argument (COMMA 2010)*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 63–74. IOS Press, 2010. doi:10.3233/978-1-60750-619-5-63.

[14] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. AFRA: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52(1):19–37, 2011. doi:10.1016/j.ijar.2010.05.004.

[15] P. Baroni, G. Governatori, and R. Riveret. On labelling statements in multi-labelling argumentation. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, volume 285 of *Frontiers in Artificial Intelligence and Applications*, pages 489–497. IOS Press, 2016. doi:10.3233/978-1-61499-672-9-489.

[16] P. Baroni, M. Caminada, and M. Giacomin. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation*, chapter 4, pages 159–236. College Publications, 2018.

[17] P. Baroni, D. M. Gabbay, and M. Giacomin. *Handbook of Formal Argumentation*. College Publications, 2018. ISBN 9781848902756. URL `https://books.google.at/books?id=_OnTswEACAAJ`.

[18] R. Baumann. What does it take to enforce an argument? Minimal change in abstract argumentation. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 127–132, 2012. doi:10.3233/978-1-61499-098-7-127.

[19] R. Baumann. Normal and strong expansion equivalence for argumentation frameworks. *Artificial Intelligence*, 193:18–44, 2012. doi:10.1016/j.artint.2012.08.004.

[20] R. Baumann. Characterizing equivalence notions for labelling-based semantics. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016.*, pages 22–32. AAAI Press, 2016.

[21] R. Baumann and H. Strass. An abstract, logical approach to characterizing strong equivalence in non-monotonic knowledge representation formalisms. *Artificial Intelligence*, 305:103680, 2022. doi:10.1016/j.artint.2022.103680.

[22] R. Baumann and M. Ulbricht. On cycles, attackers and supporters - A contribution to the investigation of dynamics in abstract argumentation. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI 2021)*, pages 1780–1786. ijcai.org, 2021. doi:10.24963/ijcai.2021/245.

[23] R. Baumann, T. Linsbichler, and S. Woltran. Verifiability of argumentation semantics. In *Proceedings of the 6th International Conference of Computational Models of Argument (COMMA 2016)*, pages 83–94, 2016. doi:10.3233/978-1-61499-686-6-83.

[24] R. Baumann, W. Dvorák, T. Linsbichler, and S. Woltran. A general notion of equivalence for abstract argumentation. *Artificial Intelligence*, 275:379–410, 2019. doi:10.1016/j.artint.2019.06.006. URL `https://doi.org/10.1016/j.artint.2019.06.006`.

[25] R. Baumann, G. Brewka, and M. Ulbricht. Comparing weak admissibility semantics to their dung-style counterparts - reduct, modularization, and strong equivalence in abstract argumentation. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR 2020)*, pages 79–88, 2020. doi:10.24963/kr.2020/9.

[26] R. Baumann, G. Brewka, and M. Ulbricht. Revisiting the foundations of abstract argumentation - semantics based on weak admissibility and weak defense. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI 2020)*, pages 2742–2749. AAAI Press, 2020. URL `https://ojs.aaai.org/index.php/AAAI/article/view/5661`.

[27] R. Baumann, D. M. Gabbay, and O. Rodrigues. Forgetting an argument. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI 2020)*, pages 2750–2757. AAAI Press, 2020. URL `https://ojs.aaai.org/index.php/AAAI/article/view/5662`.

[28] R. Baumann, S. Doutre, J.-G. Mailly, and J. P. Wallner. Enforcement in formal argumentation. In *Handbook of Formal Argumentation*, volume 2, chapter 8. College Publications, 2021. also appears in IfCoLog Journal of Logics and their Applications 8(6):1623–1678.

[29] R. Baumann, A. Rapberger, and M. Ulbricht. Equivalence in argumentation frameworks with a claim-centric view - classical results with novel ingredients. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI 2022)*, pages 5479–5486. AAAI Press, 2022. URL `https://ojs.aaai.org/index.php/AAAI/article/view/20486`.

[30] D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Verification in incomplete argumentation frameworks. *Artificial Intelligence*, 264:1–26, 2018. doi:10.1016/j.artint.2018.08.001.

[31] D. Baumeister, M. Järvisalo, D. Neugebauer, A. Niskanen, and J. Rothe. Acceptance in incomplete argumentation frameworks. *Artificial Intelligence*, 295:103470, 2021. doi:10.1016/j.artint.2021.103470.

[32] D. Baumeister, D. Neugebauer, and J. Rothe. Collective acceptability in abstract argumentation. *FLAP*, 8(6):1503–1542, 2021. URL `https://collegepublications.co.uk/ifcolog/?00048`.

[33] T. J. M. Bench-Capon. Representation of case law as an argumentation framework. In *Proceedings of the 15th Annual Conference on Legal Knowledge and Information Systems (JURIX 2002)*, pages 103–112. IOS Press: Amsterdam, 2002. URL `http://jurix.nl/pdf/j02-11.pdf`.

[34] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. Log. Comput.*, 13(3):429–448, 2003. doi:10.1093/logcom/13.3.429.

[35] T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007. doi:10.1016/j.artint.2007.05.001.

[36] T. J. M. Bench-Capon, K. Atkinson, and A. Chorley. Persuasion and value in legal argument. *Journal of Logic and Computation*, 15(6):1075–1097, 2005. doi:10.1093/logcom/exi058.

[37] C. Berge. *Hypergraphs - combinatorics of finite sets*, volume 45 of *North-Holland mathematical library*. North-Holland, 1989. ISBN 978-0-444-87489-4.

[38] M. Bernreiter, W. Dvořák, A. Rapberger, and S. Woltran. The effect of preferences in abstract argumentation under a claim-centric view. In *Proceedings of the 20th International Workshop on Non-Monotonic Reasoning (NMR 2022)*, volume 3197 of *CEUR Workshop Proceedings*, pages 27–38. CEUR-WS.org, 2022. URL `http://ceur-ws.org/Vol-3197/paper3.pdf`.

[39] M. Bernreiter, W. Dvořák, A. Rapberger, and S. Woltran. The effect of preferences in abstract argumentation under a claim-centric view. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI 2023)*, 2023. to appear.

[40] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2):203–235, 2001. doi:10.1016/S0004-3702(01)00071-6.

[41] P. Besnard and A. Hunter. A review of argumentation based on deductive arguments. In *Handbook of Formal Argumentation*, chapter 9, pages 437–484. College Publications, 2018.

[42] G. Boella, C. da Costa Pereira, A. Tettamanzi, and L. W. N. van der Torre. Making others believe what they want. In *Proceedings of Artificial Intelligence in Theory and Practice II, 20th World Computer Congress (IFIP AI 2008)*, volume 276 of *International Federation for Information Processing*, pages 215–224. Springer, 2008. doi:10.1007/978-0-387-09695-7_21.

[43] A. Bondarenko, F. Toni, and R. A. Kowalski. An assumption-based framework for non-monotonic reasoning. In *Proceedings of the 2nd International Workshop on Logic Programming and Non-monotonic Reasoning (LPNMR 1993)*, pages 171–189. MIT Press, 1993.

[44] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997. doi:10.1016/S0004-3702(97)00015-5.

[45] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI 2016)*, pages 914–920. AAAI Press, 2016. URL `http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12465`.

[46] A. Borg and F. Bex. Enforcing Sets of Formulas in Structured Argumentation. In *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning (KR 2021)*, pages 130–140, 11 2021. doi:10.24963/kr.2021/13.

[47] S. Brass and J. Dix. Characterizations of the disjunctive stable semantics by partial evaluation. *The Journal of Logic Programming*, 32(3):207–228, 1997. doi:10.1016/S0743-1066(96)00115-X.

[48] G. Brewka, T. Eiter, and M. Truszczynski. Answer set programming at a glance. *Communications of the ACM*, 54(12):92–103, 2011. doi:10.1145/2043174.2043195.

[49] G. Brewka, S. Polberg, and S. Woltran. Generalizations of dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems*, 29(1):30–38, 2014. doi:10.1109/MIS.2013.122.

[50] G. Brewka, S. Ellmauthaler, H. Strass, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks. In *Handbook of Formal Argumentation*, chapter 5, pages 237–285. College Publications, 2018. also appears in IfCoLog Journal of Logics and their Applications 4(8):2263–2318.

[51] P. Cabalar. A three-valued characterization for strong equivalence of logic programs. In *Proceedings of the 18th National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI 2002)*, pages 106–111. AAAI Press / The MIT Press, 2002. URL `http://www.aaai.org/Library/AAAI/2002/aaai02-017.php`.

[52] E. Cabrio and S. Villata. Five years of argument mining: a data-driven analysis. In *Proceedings of the 27nth International Joint Conference on Artificial Intelligence (IJCAI 2018)*, pages 5427–5433. ijcai.org, 2018. doi:10.24963/ijcai.2018/766.

[53] M. Caminada. Semi-stable semantics. In *Proceedings of the 1st Conference on Computational Models of Argument (COMMA 2006)*, volume 144 of *Frontiers in Artificial Intelligence and Applications*, pages 121–130. IOS Press, 2006. URL `http://www.booksonline.iospress.nl/Content/View.aspx?piid=1932`.

[54] M. Caminada. On the issue of reinstatement in argumentation. In *Proceedings of the 10th European Conference on Logics in Artificial Intelligence (JELIA 2006)*, volume 4160 of *Lecture Notes in Computer Science*, pages 111–123. Springer, 2006. doi:10.1007/11853886_11.

[55] M. Caminada. Comparing two unique extension semantics for formal argumentation: ideal and eager. In *Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence (BNAIC 2007)*, pages 81–87. Utrecht University Press, 2007.

[56] M. Caminada. An algorithm for computing semi-stable semantics. In *Proceedings on the 9th European Conference Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007)*, volume 4724 of *Lecture Notes in Computer Science*, pages 222–234. Springer, 2007. doi:10.1007/978-3-540-75256-1_22.

[57] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5–6):286–310, 2007. doi:10.1016/j.artint.2007.02.003.

[58] M. Caminada and D. M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2):109–145, 2009. ISSN 0039-3215. doi:10.1007/s11225-009-9218-x.

[59] M. Caminada and C. Schulz. On the equivalence between assumption-based argumentation and logic programming. *Journal of Artificial Intelligence Research*, 60:779–825, 2017. doi:10.1613/jair.5581.

[60] M. Caminada, W. A. Carnielli, and P. E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22:1207–1254, 2012. doi:10.1093/logcom/exr033.

[61] M. Caminada, S. Sá, J. Alcântara, and W. Dvořák. On the equivalence between logic programming semantics and argumentation semantics. *International Journal of Approximate Reasoning*, 58:87–111, 2015. doi:10.1016/j.ijar.2014.12.004.

[62] M. Caminada, S. Sá, J. Alcântara, and W. Dvořák. On the difference between assumption-based argumentation and abstract argumentation. *IfCoLog Journal of Logic and its Applications*, 2(1):15–34, 2015. URL `http://www.collegepublications.co.uk/downloads/ifcolog00003.pdf`.

[63] C. Cayrol and M. Lagasquie-Schiex. Bipolar abstract argumentation systems. In *Argumentation in Artificial Intelligence*, pages 65–84. Springer, 2009. doi:10.1007/978-0-387-98197-0_4.

[64] C. Cayrol and M.-C. Lagasquie-Schiex. 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)*, volume 3571 of *Lecture Notes in Computer Science*, pages 378–389. Springer, 2005. doi:10.1007/11518655_33.

[65] C. Cayrol, F. D. de Saint-Cyr, and M. Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010. doi:10.1613/jair.2965.

[66] F. Cerutti, S. A. Gaggl, M. Thimm, and J. P. Wallner. Foundations of implementations for formal argumentation. In *Handbook of Formal Argumentation*, chapter 15, pages 688–767. College Publications, 2018. also appears in IfCoLog Journal of Logics and their Applications 4(8):2623–2706.

[67] E. A. Corsi and C. G. Fermüller. Logical argumentation principles, sequents, and nondeterministic matrices. In *Proceedings of the 6th International Workshop on Logic, Rationality, and Interaction (LORI 2017)*, volume 10455 of *Lecture Notes in Computer Science*, pages 422–437. Springer, 2017. doi:10.1007/978-3-662-55665-8_29.

[68] E. A. Corsi and C. G. Fermüller. Connecting fuzzy logic and argumentation frames via logical attack principles. *Soft Computing*, 23(7):2255–2270, 2019. doi:10.1007/s00500-018-3513-2.

[69] R. Craven and F. Toni. Argument graphs and assumption-based argumentation. *Artif. Intell.*, 233:1–59, 2016. doi:10.1016/j.artint.2015.12.004. URL `https://doi.org/10.1016/j.artint.2015.12.004`.

[70] K. Cyras, X. Fan, C. Schulz, and F. Toni. Assumption-based argumentation: Disputes, explanations, preferences. In *Handbook of Formal Argumentation*, chapter 7, pages 365–408. College Publications, 2018. also appears in IfCoLog Journal of Logics and their Applications 4(8):2407–2456.

[71] K. Cyras, Q. Heinrich, and F. Toni. Computational complexity of flat and generic assumption-based argumentation, with and without probabilities. *Artificial Intelligence*, 293:103449, 2021. doi:10.1016/j.artint.2020.103449.

[72] M. Diller, A. Haret, T. Linsbichler, S. Rümmele, and S. Woltran. An extension-based approach to belief revision in abstract argumentation. *International Journal of Approximate Reasoning*, 93:395–423, 2018. doi:10.1016/j.ijar.2017.11.013.

[73] Y. Dimopoulos and A. C. Kakas. Learning non-monotonic logic programs: Learning exceptions. In N. Lavrac and S. Wrobel, editors, *Machine Learning: ECML-95, 8th European Conference on Machine Learning, Heraclion, Crete, Greece, April 25-27, 1995, Proceedings*, volume 912 of *Lecture Notes in Computer Science*,

pages 122–137. Springer, 1995. doi:10.1007/3-540-59286-5_53. URL `https://doi.org/10.1007/3-540-59286-5_53`.

[74] Y. Dimopoulos and A. Torres. Graph theoretical structures in logic programs and default theories. *Theoretical Computer Science*, 170(1-2):209–244, 1996. doi:10.1016/S0304-3975(96)80707-9.

[75] Y. Dimopoulos, B. Nebel, and F. Toni. On the computational complexity of assumption-based argumentation for default reasoning. *Artificial Intelligence*, 141 (1/2):57–78, 2002. doi:10.1016/S0004-3702(02)00245-X.

[76] S. Doutre and J. Mailly. Constraints and changes: A survey of abstract argumentation dynamics. *Argument and Computation*, 9(3):223–248, 2018. doi:10.3233/AAC-180425.

[77] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995. doi:10.1016/0004-3702(94)00041-X.

[78] P. M. Dung, R. A. Kowalski, and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artif. Intell.*, 170(2):114–159, 2006. doi:10.1016/j.artint.2005.07.002. URL `https://doi.org/10.1016/j.artint.2005.07.002`.

[79] P. M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674, 2007. doi:10.1016/j.artint.2007.05.003.

[80] P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009. doi:10.1007/978-0-387-98197-0_10.

[81] P. E. Dunne and T. J. M. Bench-Capon. Coherence in finite argument systems. *Artificial Intelligence*, 141(1/2):187–203, 2002. doi:10.1016/S0004-3702(02)00261-8.

[82] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, 2011. doi:10.1016/j.artint.2010.09.005.

[83] P. E. Dunne, P. Marquis, and M. Wooldridge. Argument aggregation: Basic axioms and complexity results. In *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA 2012)*, volume 245 of *Frontiers in Artificial Intelligence and Applications*, pages 129–140. IOS Press, 2012. doi:10.3233/978-1-61499-111-3-129.

[84] P. E. Dunne, W. Dvořák, T. Linsbichler, and S. Woltran. Characteristics of multiple viewpoints in abstract argumentation. *Artificial Intelligence*, 228:153–178, 2015. doi:10.1016/j.artint.2015.07.006.

[85] W. Dvorák, J. Fandinno, and S. Woltran. On the expressive power of collective attacks. *Argument and Computation*, 10(2):191–230, 2019. doi:10.3233/AAC-190457.

[86] W. Dvorák, M. Järvisalo, T. Linsbichler, A. Niskanen, and S. Woltran. Preprocessing argumentation frameworks via replacement patterns. In *Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019, Proceedings*, volume 11468 of *Lecture Notes in Computer Science*, pages 116–132. Springer, 2019. doi:10.1007/978-3-030-19570-0_8.

[87] W. Dvořák, A. Keshavarzi Zafarghandi, and S. Woltran. Expressiveness of SETAFs and support-free ADFs under 3-valued semantics. In *Proceedings of the 8th International Conference on Computational Model of Argument (COMMA 2020)*, pages 191–202. IOS Press, 2020. doi:10.3233/FAIA200504.

[88] W. Dvorák, M. König, M. Ulbricht, and S. Woltran. A reduct-driven study of argumentation frameworks with collective attacks. In *Proceedings of the 19th International Workshop on Non-Monotonic Reasoning (NMR 2021)*, pages 285–294, 2021. URL `https://drive.google.com/file/d/1WSIl3TOrXBhaWhckWN4NLXoD9AVFKp5R/view`.

[89] W. Dvořák and P. E. Dunne. Computational problems in formal argumentation and their complexity. In *Handbook of Formal Argumentation*, chapter 14, pages 631–687. College Publications, 2018. also appears in IfCoLog Journal of Logics and their Applications 4(8):2557–2622.

[90] W. Dvořák and S. Woltran. Complexity of semi-stable and stage semantics in argumentation frameworks. *Information Processing Letters*, 110(11):425–430, 2010. doi:10.1016/j.ipl.2010.04.005.

[91] W. Dvořák and S. Woltran. On the intertranslatability of argumentation semantics. *Journal of Artificial Intelligence Research*, 41:445–475, 2011. doi:10.1613/jair.3318.

[92] W. Dvořák and S. Woltran. Complexity of abstract argumentation under a claim-centric view. *Artificial Intelligence*, 285:103290, 2020. doi:10.1016/j.artint.2020.103290.

[93] W. Dvořák, A. Rapberger, and S. Woltran. Strong equivalence for argumentation frameworks with collective attacks. In *Proceedings of the 42nd German Conference on AI (KI 2019)*, volume 11793 of *Lecture Notes in Computer Science*, pages 131–145. Springer, 2019. doi:10.1007/978-3-030-30179-8_11.

[94] W. Dvořák, S. A. Gaggl, A. Rapberger, J. P. Wallner, and S. Woltran. The ASPARTIX system suite. In *Proceedings of the 8th International Conference on Computational Models of Argument (COMMA 2020)*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, pages 461–462. IOS Press, 2020. doi:10.3233/FAIA200534.

[95] W. Dvořák, A. Rapberger, and J. P. Wallner. Labelling-based algorithms for SETAFs. In *Proceedings of the 3rd International Workshop on Systems and Algorithms for Formal Argumentation (SAFA 2020)*, volume 2672 of *CEUR Workshop Proceedings*, pages 34–46. CEUR-WS.org, 2020.

[96] W. Dvořák, A. Rapberger, J. P. Wallner, and S. Woltran. ASPARTIX-V19 - an answer-set programming based system for abstract argumentation. In *Proceedings of the 11th International Symposium on Foundations of Information and Knowledge Systems (FoIKS 2020)*, volume 12012 of *Lecture Notes in Computer Science*, pages 79–89. Springer, 2020. doi:10.1007/978-3-030-39951-1_5.

[97] W. Dvořák, A. Rapberger, and S. Woltran. On the relation between claim-augmented argumentation frameworks and collective attacks. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 721–728. IOS Press, 2020. doi:10.3233/FAIA200159.

[98] W. Dvořák, A. Rapberger, and S. Woltran. Argumentation semantics under a claim-centric view: Properties, expressiveness and relation to setafs. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR 2020)*, pages 341–350, 2020. doi:10.24963/kr.2020/35.

[99] W. Dvořák, A. Greßler, A. Rapberger, and S. Woltran. The complexity landscape of claim-augmented argumentation frameworks. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)*, pages 6296–6303. AAAI Press, 2021. URL `https://ojs.aaai.org/index.php/AAAI/article/view/16782`.

[100] W. Dvořák, M. König, A. Rapberger, J. P. Wallner, and S. Woltran. ASPARTIX-V - a solver for argumentation tasks using ASP. In *Workshop on Answer Set Programming and Other Computing Paradigms (ASPOCP 2021)*, 2021.

[101] W. Dvořák, M. König, M. Ulbricht, and S. Woltran. Rediscovering argumentation principles utilizing collective attacks. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022)*, 2022. URL `https://proceedings.kr.org/2022/13/`.

[102] W. Dvořák, A. Rapberger, and S. Woltran. A claim-centric perspective on abstract argumentation semantics: Claim-defeat, principles, and expressiveness. Technical Report DBAI-TR-2022-124, TU Wien, 2022. URL `https://www.dbai.tuwien.ac.at/research/report/dbai-tr-2022-124.pdf`.

[103] W. Dvořák, A. Greßler, A. Rapberger, and S. Woltran. The complexity landscape of claim-augmented argumentation frameworks. *Artificial Intelligence*, 317:103873, 2023. ISSN 0004-3702. doi:https://doi.org/10.1016/j.artint.2023.103873. URL `https://www.sciencedirect.com/science/article/pii/S000437022300019X`.

[104] W. Dvořák, A. Rapberger, and S. Woltran. On the different types of collective attacks in abstract argumentation: equivalence results for SETAFs. *Journal of Logic and Computation*, 30(5):1063–1107, 06 2020. doi:10.1093/logcom/exaa033.

[105] T. Eiter, M. Fink, H. Tompits, and S. Woltran. Simplifying logic programs under uniform and strong equivalence. In *Proceedings of the 7th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR 2004)*, volume 2923 of *Lecture Notes in Computer Science*, pages 87–99. Springer, 2004. doi:10.1007/978-3-540-24609-1_10.

[106] T. Eiter, M. Fink, H. Tompits, and S. Woltran. Complexity results for checking equivalence of stratified logic programs. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*, pages 330–335, 2007. URL `http://ijcai.org/Proceedings/07/Papers/051.pdf`.

[107] T. Eiter, M. Fink, and S. Woltran. Semantical characterizations and complexity of equivalences in answer set programming. *ACM Transactions on Computational Logic*, 8(3):17, 2007. doi:10.1145/1243996.1244000.

[108] M. A. Falappa, G. Kern-Isberner, and G. R. Simari. Explanations, belief revision and defeasible reasoning. *Artificial Intelligence*, 141(1/2):1–28, 2002. doi:10.1016/S0004-3702(02)00258-8.

[109] M. A. Falappa, G. Kern-Isberner, and G. R. Simari. Belief revision and argumentation theory. In *Argumentation in Artificial Intelligence*, pages 341–360. Springer, 2009. doi:10.1007/978-0-387-98197-0_17.

[110] B. Fazzinga, S. Flesca, and F. Furfaro. Reasoning over argument-incomplete AAFs in the presence of correlations. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI 2021)*, pages 189–195. ijcai.org, 2021. doi:10.24963/ijcai.2021/27.

[111] J. K. Fichte, M. Hecher, and A. Meier. Counting complexity for reasoning in abstract argumentation. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019)*, pages 2827–2834. AAAI Press, 2019. doi:10.1609/aaai.v33i01.33012827.

[112] S. A. Gaggl and S. Woltran. Strong equivalence for argumentation semantics based on conflict-free sets. In *Proceedings of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2011)*, volume 6717 of *Lecture Notes in Computer Science*, pages 38–49. Springer, 2011. doi:10.1007/978-3-642-22152-1_4.

[113] A. J. García and G. R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138, 2004. doi:10.1017/S1471068403001674.

[114] A. J. García and G. R. Simari. Argumentation based on logic programming. In *Handbook of Formal Argumentation*, chapter 8, pages 409–435. College Publications, 2018.

[115] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proceedings of the 5th International Conference and Symposium on Logic Programming (ICLP/SLP 1988)*, pages 1070–1080. MIT Press, 1988. ISBN 0-262-61056-6.

[116] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence*, 175(9-10): 1479–1497, 2011. doi:10.1016/j.artint.2010.12.003.

[117] A. Greßler. Argumentation frameworks with claims and collective attacks: Complexity results and answer-set programming encodings. Master's thesis, Vienna University of Technology, 2019. Stefan Woltran and Wolfgang Dvořák advisors.

[118] E. Hadjisoteriou and A. C. Kakas. Reasoning about actions and change in argumentation. *Argument and Computation*, 6(3):265–291, 2015. doi:10.1080/19462166.2015.1123774.

[119] T. Hamm, M. Lackner, and A. Rapberger. Computing kemeny rankings from d-euclidean preferences. In *Proceedings of the 7th International Conference on Algorithmic Decision Theory (ADT 2021)*, volume 13023 of *Lecture Notes in Computer Science*, pages 147–161, 2021. doi:10.1007/978-3-030-87756-9_10.

[120] J. F. Horty. Skepticism and floating conclusions. *Artificial Intelligence*, 135(1-2): 55–72, 2002. doi:10.1016/S0004-3702(01)00160-6.

[121] J. F. Horty, D. Touretzky, and R. Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI 1987)*, pages 476–482. Morgan Kaufmann Los Altos, CA, 1987. URL `http://ijcai.org/Proceedings/87-1/Papers/094.pdf`.

[122] J. F. Horty, R. H. Thomason, and D. S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial intelligence*, 42(2-3):311–348, 1990. doi:10.1016/0004-3702(90)90057-7.

[123] A. Hunter and M. Williams. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine*, 56(3):173–190, 2012. doi:10.1016/j.artmed.2012.09.004.

[124] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9(2):215–261, 1999. doi:10.1093/logcom/9.2.215.

[125] S. Kaci, L. W. N. van der Torre, S. Vesic, and S. Villata. Preference in abstract argumentation. In *Handbook of Formal Argumentation, Volume 2*, pages 211–248. College Publications, 2021.

[126] R. Kiesel and A. Rapberger. Concurrence for well-formed cafs: Naive semantics, 2021. URL `https://arxiv.org/abs/2112.10445`.

[127] M. König, A. Rapberger, and M. Ulbricht. Just a matter of perspective: Intertranslating expressive argumentation formalisms. In *Proceedings of the 10th International Conference on Computational Models of Argument (COMMA 2022)*, volume 353 of *Frontiers in Artificial Intelligence and Applications*, pages 212–223. IOS Press, 2022. doi:10.3233/FAIA220154.

[128] R. Kowalski. Algorithm = logic + control. *Communications of the ACM*, 22(7): 424–436, 1979. doi:10.1145/359131.359136.

[129] J. Lawrence and C. Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2019. doi:10.1162/coli_a_00364.

[130] H. Li, N. Oren, and T. J. Norman. Probabilistic argumentation frameworks. In *Proceedings of the 1st International Workshop on Theorie and Applications of Formal Argumentation (TAFA 2011), Revised Selected Papers*, volume 7132 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 2011. doi:10.1007/978-3-642-29184-5_1.

[131] B. S. Liao, L. Jin, and R. C. Koons. Dynamics of argumentation systems: A division-based method. *Artificial Intelligence*, 175(11):1790–1814, 2011. doi:10.1016/j.artint.2011.03.006.

[132] V. Lifschitz, D. Pearce, and A. Valverde. Strongly equivalent logic programs. *ACM Transactions on Computational Logic*, 2(4):526–541, 2001. doi:10.1145/383779.383783.

[133] F. Lin. Reducing strong equivalence of logic programs to entailment in classical propositional logic. In *Proceedings of the Eights International Conference on Principles and Knowledge Representation and Reasoning (KR 2002)*, pages 170–176. Morgan Kaufmann, 2002.

[134] F. Lin and Y. Chen. Discovering classes of strongly equivalent logic programs. *Journal of Artificial Intelligence Research*, 28:431–451, 2007. doi:10.1613/jair.2131.

[135] J. W. Lloyd. *Foundations of logic programming*. Springer, 1984. ISBN 3-540-13299-6.

[136] D. Makinson and K. Schlechta. Floating conclusions and zombie paths: two deep difficulties in the “directly skeptical” approach to defeasible inheritance nets. *Artificial intelligence*, 48(2):199–209, 1991. doi:10.1016/0004-3702(91)90061-N.

[137] S. Modgil and H. Prakken. Abstract rule-based argumentation. In *Handbook of Formal Argumentation*, chapter 6, pages 287–364. College Publications, 2018. also appears in IfCoLog Journal of Logics and their Applications 4(8):2319–2406.

[138] M. O. Moguillansky, N. D. Rotstein, M. A. Falappa, A. J. García, and G. R. Simari. Argument theory change applied to defeasible logic programming. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI 2008)*, pages 132–137. AAAI Press, 2008. URL `http://www.aaai.org/Library/AAAI/2008/aaai08-021.php`.

[139] S. H. Muggleton. Inductive logic programming. *New Gener. Comput.*, 8(4): 295–318, 1991. doi:10.1007/BF03037089. URL `https://doi.org/10.1007/BF03037089`.

[140] S. H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *3rd International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2006), Revised Selected and Invited Papers*, volume 4766 of *Lecture Notes in Computer Science*, pages 54–73. Springer, 2006. doi:10.1007/978-3-540-75526-5_4.

[141] S. H. Nielsen and S. Parsons. An application of formal argumentation: Fusing bayesian networks in multi-agent systems. *Artificial Intelligence*, 171(10-15):754–775, 2007. doi:10.1016/j.artint.2007.04.005.

[142] E. Oikarinen and S. Woltran. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence*, 175(14-15):1985–2009, 2011. doi:10.1016/j.artint.2011.06.003.

[143] M. Osorio, J. A. N. Pérez, and J. Arrazola. Equivalence in answer set programming. In *Proceedings of the 11th International Workshop on Logic Based Program Synthesis and Transformation (LOPSTR 2001)*, volume 2372 of *Lecture Notes in Computer Science*, pages 57–75. Springer, 2001. doi:10.1007/3-540-45607-4_4.

[144] S. Pandzic. Structured argumentation dynamics. *Ann. Math. Artif. Intell.*, 90(2-3): 297–337, 2022. doi:10.1007/s10472-021-09765-z.

[145] C. H. Papadimitriou. *Computational complexity*. Addison-Wesley, 1994. ISBN 978-0-201-53082-7.

[146] D. Pearce, H. Tompits, and S. Woltran. Encodings for equilibrium logic and logic programs with nested expressions. In *Progress in Artificial Intelligence, Knowledge Extraction, Multi-agent Systems, Logic Programming and Constraint Solving, Proceedings of the 10th Portuguese Conference on Artificial Intelligence (EPIA 2001)*, volume 2258 of *Lecture Notes in Computer Science*, pages 306–320. Springer, 2001. doi:10.1007/3-540-45329-6_31.

[147] P. Peppas. Belief revision. 3:317–359, 2008. doi:10.1016/S1574-6526(07)03008-8.

[148] S. Polberg. *Developing the abstract dialectical framework*. PhD thesis, TU Wien, Institute of Information Systems, 2017.

[149] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987. doi:10.1207/s15516709cog1104_4.

[150] H. Prakken. Intuitions and the modelling of defeasible reasoning: some case studies. pages 91–102, 2002. URL `https://nmr.cs.tu-dortmund.de/proceedings/NMR2002Proceedings.pdf`.

[151] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010. doi:10.1080/19462160903564592.

[152] M. Proietti and F. Toni. Learning assumption-based argumentation frameworks. In *Proceedings of the 31st International Conference on Inductive Logic Programming (ILP 2022)*, 2022. URL `http://hdl.handle.net/10044/1/98940`.

[153] T. C. Przymusinski. The well-founded semantics coincides with the three-valued stable semantics. *Fundamenta Informaticae*, 13(4):445–463, 1990.

[154] I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*. Springer, 2009. ISBN 0387981969, 9780387981963.

[155] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(4):343–375, 2003. doi:10.1017/S0269888904000098.

[156] I. Rahwan, M. I. Madakkatel, J. Bonnefon, R. N. Awan, and S. Abdallah. Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. volume 34, pages 1483–1502, 2010. doi:10.1111/j.1551-6709.2010.01123.x.

[157] A. Rapberger. Defining argumentation semantics under a claim-centric view. In *Proceedings of the 9th European Starting AI Researchers' Symposium (STAIRS 2020)*, volume 2655 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020. URL `http://ceur-ws.org/Vol-2655/paper2.pdf`.

[158] A. Rapberger and M. Ulbricht. On dynamics in structured argumentation formalisms. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022)*, pages 288–298, 2022. URL `https://proceedings.kr.org/2022/29/`.

[159] A. Rapberger, M. Ulbricht, and J. P. Wallner. Argumentation frameworks induced by assumption-based argumentation: Relating size and complexity. In *Proceedings of the 20th International Workshop on Non-Monotonic Reasoning (NMR 2022)*, volume 3197 of *CEUR Workshop Proceedings*, pages 92–103. CEUR-WS.org, 2022. URL `http://ceur-ws.org/Vol-3197/paper9.pdf`.

[160] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1-2):81–132, 1980. doi:10.1016/0004-3702(80)90014-4.

[161] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM (JACM)*, 12(1):23–41, 1965. doi:10.1145/321250.321253.

[162] V. H. N. Rocha and F. G. Cozman. A credal least undefined stable semantics for probabilistic logic programs and probabilistic argumentation. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022)*, 2022. URL `https://proceedings.kr.org/2022/31/`.

[163] V. H. N. Rocha and F. G. Cozman. Bipolar argumentation frameworks with explicit conclusions: Connecting argumentation and logic programming. In *Proceedings of the 20th International Workshop on Non-Monotonic Reasoning (NMR 2022)*, volume 3197 of *CEUR Workshop Proceedings*, pages 49–60. CEUR-WS.org, 2022. URL `http://ceur-ws.org/Vol-3197/paper5.pdf`.

[164] N. D. Rotstein, M. O. Moguillansky, M. A. Falappa, A. J. García, and G. R. Simari. Argument theory change: Revision upon warrant. In *Proceedings of the 2nd International Conference on Computational Models of Argument: Proceedings of (COMMA 2008)*, volume 172 of *Frontiers in Artificial Intelligence and Applications*, pages 336–347. IOS Press, 2008. URL `http://www.booksonline.iospress.nl/Content/View.aspx?piid=9292`.

[165] N. D. Rotstein, M. O. Moguillansky, A. J. García, and G. R. Simari. A dynamic argumentation framework. In *Proceedings of the 3rd International Conference on Computational Model of Argument (COMMA 2010)*, volume 216 of *FAIA*, pages 427–438. IOS Press, 2010. doi:10.3233/978-1-60750-619-5-427.

[166] K. Schlechta. Directly sceptical inheritance cannot capture the intersection of extensions. *Journal of Logic and Computation*, 3(5):455–467, 1993. doi:10.1093/logcom/3.5.455.

[167] U. Schöning. Graph isomorphism is in the low hierarchy. *Journal of Computer and System Sciences*, 37(3):312–323, 1988. doi:10.1016/0022-0000(88)90010-4.

[168] M. Snaith and C. Reed. Argument revision. *Journal of Logic and Computation*, 27 (7):2089–2134, 2017. doi:10.1093/logcom/exw028.

[169] L. A. Stein. *Skeptical Inheritance: Computing the Intersection of Credulous Extensions*. Morgan Kaufmann, 1989. URL `http://ijcai.org/Proceedings/89-2/Papers/049.pdf`.

[170] H. Strass. Approximating operators and semantics for abstract dialectical frameworks. *Artificial Intelligence*, 205:39–70, 2013. doi:10.1016/j.artint.2013.09.004.

[171] B. Testerink, D. Odekerken, and F. Bex. A method for efficient argument-based inquiry. In *Proceedings of the 13th International Conference on Flexible Query Answering Systems (FQAS 2019)*, pages 114–125. Springer, 2019. doi:10.1007/978-3-030-27629-4_13.

[172] P. M. Thang, P. M. Dung, and N. D. Hung. Towards a common framework for dialectical proof procedures in abstract argumentation. *J. Log. Comput.*, 19 (6):1071–1109, 2009. doi:10.1093/logcom/exp032. URL `https://doi.org/10.1093/logcom/exp032`.

[173] S. E. Toulmin. *The Uses of Argument, Updated Edition*. Cambridge University Press, 2008. ISBN 9780521534833.

[174] D. S. Touretzky. *The Mathematics of Inheritance Systems*. Morgan Kaufmann, 1986. ISBN 0-934613-06-0.

[175] L. van der Torre and S. Vesic. The principle-based approach to abstract argumentation semantics. *FLAP*, 4(8), 2017. URL `http://www.collegepublications.co.uk/downloads/ifcolog00017.pdf`.

[176] F. H. Van Eemeren, R. Grootendorst, R. H. Johnson, C. Plantin, and C. A. Willard. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Routledge, 1996. ISBN 9780805818628.

[177] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proceedings of the 8th Dutch Conference on Artificial Intelligence (NAIC'96)*, pages 357–368, 1996.

[178] J. P. Wallner. *Complexity Results and Algorithms for Argumentation - Dung's Frameworks and Beyond*. PhD thesis, Vienna University of Technology, Institute of Information Systems, 2014.

[179] J. P. Wallner. Structural constraints for dynamic operators in abstract argumentation. *Argument and Computation*, 11(1-2):151–190, 2020. doi:10.3233/AAC-190471.

[180] J. P. Wallner, A. Niskanen, and M. Järvisalo. Complexity results and algorithms for extension enforcement in abstract argumentation. *Journal of Artificial Intelligence Research*, 60:1–40, 2017. doi:10.1613/jair.5415.

[181] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, 2008. ISBN 9780521723749. URL `http://www.cambridge.org/us/academic/subjects/philosophy/logic/argumentation-schemes`.

[182] D. N. Walton. *Fundamentals of critical argumentation*. Critical reasoning and argumentation. Cambridge University Press, 2006. ISBN 978-0-521-53020-0.

[183] K. Wang and L. Zhou. Comparisons and computation of well-founded semantics for disjunctive logic programs. *ACM Transactions on Computational Logic*, 6(2): 295–327, 2005. doi:10.1145/1055686.1055690.

[184] Y. Wu, M. Caminada, and D. M. Gabbay. Complete extensions in argumentation coincide with 3-valued stable models in logic programming. *Studia Logica*, 93(2-3): 383–403, 2009. doi:10.1007/s11225-009-9210-5.