Reasoning, Arguing, Ranking, Aggregating

Habilitation à diriger des recherches

written & defended by

Srdjan Vesić

HABILITATION À DIRIGER DES RECHERCHES

presented by

Srdjan VESIC

March 18th 2021

at Univ. Artois, Lens, France

Reasoning, arguing, ranking and aggregating

Jury:

Pietro Baroni	Professor at Università degli Studi di Brescia	Rapporteur
Anthony Hunter	Professor at University College London	Rapporteur
Francesca Toni	Professor at Imperial College London	Rapporteur
Sébastien Konieczny	Research director at CNRS	Examinateur
Pierre Marquis	Professor at Université d'Artois	Examinateur
Nicolas Maudet	Professor at Sorbonne University	Examinateur

Contents

Many thanks to the committee, coauthors, friends, colleagues and family

1. Introduction

This habilitation thesis describes the research I conducted between July 2011 and now (October 2020). I summarise my motivations, research themes and main results in this section, and give more details in the rest of the document. This document is based on several papers I published with my colleagues and contains some parts of them, in order to provide an overview of my research activity from 2011 to 2020.

I obtained virtually all results together with my colleagues, the only exception being the work I did alone on maxi-consistent operators, published in ECAI 2012 and Journal of Artificial Intelligence Research (2013). I would like to thank my coauthors[1]: Madalina Croitoru (LIRMM, Montpellier), Bruno Yun (University of Aberdeen, Scotland), Leila Amgoud (IRIT, Toulouse), Sébastien Konieczny (CRIL, Lens), Pierre Bisquert (INRA, Montpellier), Leender van der Torre (University of Luxembourg), Jérôme Lang (LAMSADE, Paris), Marija Slavkovik (University of Bergen, Norway), Dragan Doder (University of Utrecht, The Netherlands), Pierre Marquis (CRIL, Lens), Jérôme Delobelle (LIPADE, Paris), Jonathan Ben-Naim (IRIT, Toulouse), Rallou Thomopoulos (INRA, Montpellier), Gabriella Pigozzi (LAMSADE, Paris), Martin Caminada (Cardiff University, Wales), Nir Oren (University of Aberdeen, Scotland), Wolfgang Dvořák (Vienna University of Technology, Autria), Philippe Besnard (IRIT, Toulouse), Abdallah Arioua (LIRIS, Lyon), François Schwarzentruber (IRISA, Rennes), Tjitze Rienstra (Universität Koblenz-Landau, Germany), Nicolas Maudet (LIP6, Paris), Elise Bonzon (LIPADE, Paris), João Leite (Universidade Nova de Lisboa, Portugal), Alexis Martin, (LIP6, Paris), Marco Correia, (Universidade Nova de Lisboa, Portugal), Jorge Cruz (Universidade Nova de Lisboa, Portugal).

In most of my publications, the authors' names are in alphabetical order. However, some of the papers follow the contribution order; the corresponding references in the bibliography section end with *(contr)*.

To keep this document within a reasonable page limit, I present my contributions briefly, leaving out the proofs and examples, which an interested reader can find in the corresponding papers I published. For similar reasons, I do not present the state of the art in detail; some works are mentioned, mostly to help in situating my contributions.

[1] ranked by the number of papers we wrote together since the end of my PhD thesis until now

1.1 Research questions

My field of research is *inferring from inconsistent knowledge*. When the knowledge base is inconsistent, classical logic is useless since it allows to infer everything. Another drastic stance is not to infer anything. Obviously, neither of the two options provides us with any information about the state of the world. The goal of my research is to allow to *infer as many conclusions as possible* while still staying consistent. I study different methods for inferring from inconsistent knowledge bases. One way to achieve this is to use argumentation, each argument being a consistent piece of information built from an inconsistent knowledge base. One constructs arguments and counter-arguments, as well as attacks (and possibly supports) between them. Then, we need to determine which arguments to accept. There are various ways to do this, as will be detailed later in the document. Considering the goal of increasing the amount of inferred conclusions, I worked on several ways to achieve this. Since, roughly speaking, the conclusions are inferred from the intersection of all extensions (repairs), one way to increase the amount of conclusions is to decrease the number of extensions (repairs). I also present other ways to do this in the corresponding chapter. Finally, I worked in the area of judgment aggregation, which studies the problems related to aggregating a finite set of individual judgments, cast on a collection of logically interrelated issues. Each individual provides a consistent judgment set, but combining them in an arbitrary way might result in an inconsistent result. Thus, the question is how to aggregate them in order to obtain a *consistent collective judgment set*, which should ideally represent the opinions of all the individuals, i.e. keep as much information as possible.

1.2 Methodology

My scientific approach is based on defining and studying the *principles*. Rather than defining an *ad hoc* solution and testing it on examples, I am interested in developing principles a system should satisfy. Such an approach comes with several benefits: principles allow to understand the underpinnings of the systems we study; they help us to evaluate and compare systems; finally, they can guide us when defining new systems.

The principle-based approach is a methodology that is successfully applied in many scientific disciplines. It can be used once a unique universal method is replaced by a variety of alternative methods, for example, once a variety of modal logics is used to represent knowledge instead of unique first order logic. The principle-based approach is also called the axiomatic approach, or the postulate based approach, for example in AGM theory change by Alchourrón et al. (1985). The principles are typically desirable, and desirable properties are sometimes called postulates. For the mathematical development of a principle-based theory, it may be less relevant whether principles are desirable or not.

Maybe the best known example of the principle-based approach is concerned with the variety of voting rules, a core challenge in democratic societies. Over the past two centuries many voting rules have been proposed, and researchers were wondering how we can know that the currently considered set of voting rules is sufficient or complete, and whether there is no better voting rule that has not been discovered yet. In voting theory, the principle-based approach was introduced by Nobel prize winner Kenneth Arrow (Arrow, 1951). The principle-based approach classifies existing approaches based on axiomatic principles, such that we can select a voting rule based on the set of requirements in an area. Moreover, there may be sets of principles for which no voting rule exists yet. Beyond voting theory, the principle-based approach has been applied in a large variety of domains, including abstract argumentation.

In *argumentation theory*, the key notion is that of a *semantics*, since it is used to determine which arguments to accept. I study the principles those semantics must satisfy, and also the non-mandatory (i.e. optional) principles that a semantics is not required to satisfy, but that can be used to discriminate and better understand semantics. My study of reducing the number of repairs and defining inference relations that allow to deduce more conclusions is also guided by the principles. In some works we prove that the relations we define satisfy existing principles, e.g. KLM (Kraus et al., 1990); in others we develop principles that are appropriate for the given setting. My research in judgment aggregation is also based on studying principles or the properties of judgment aggregation rules. For example, we studied and criticized existing principles (e.g. independence). We also made a taxonomy of virtually all judgment aggregation rules from the literature and the principles they satisfy in order to analyze and compare the rules with respect to the principles.

1.3 Summary of my work

This section contains a brief summary of my work. For coherence reasons, I do not detail all of my contributions.

1.3.1 Principles for argumentation semantics

One way to infer from an inconsistent knowledge base is by using argumentation. *Argumentation* is a study of how to construct arguments, counter-arguments and attacks between them in order to assess their acceptability degree. I study the computational models of argument (Dung, 1995; Besnard and Hunter, 2008; Rahwan and Simari, 2009; Baroni et al., 2018a), i.e. ways to model argumentation using tools from mathematics, logic and computer science. In order to determine which arguments are acceptable, different argumentation semantics were defined. They can be divided in *extension-based* semantics, where sets of arguments are accepted together in so-called extensions, and *ranking-based* semantics, where arguments are individually ranked from the most to the least acceptable. I am particularly interested in *principles* those semantics should satisfy.

The extension-based semantics have been studied for a longer period of time. Accordingly, many principles they should satisfy were proposed (Baroni and Giacomin, 2007; Baroni et al., 2011a; Caminada et al., 2012). My work in this area was not in defining new principles but rather studying the existing ones. For example, many new semantics were defined and we were interested in whether those semantics satisfy existing principles. We proposed a systematic study of the fifteen main alternatives for argumentation semantics using the twenty-seven main principles discussed in the literature on abstract argumentation (van der Torre and Vesic, 2018), thus extending Baroni and Giacomin's (2007) original classification with other semantics and principles proposed in the literature. This work lays the foundations for a study of representation and (im)possibility results for abstract argumentation, and for a principle-based approach for extended argumentation such as bipolar frameworks, preference-based frameworks, abstract dialectical frameworks, weighted frameworks, and input/output frameworks.

On the contrary, in the newer area of ranking-based semantics, our work consisted in proposing new principles, as well as studying the link between principles (Amgoud et al., 2017a). Namely, we show that some sets of principles are incompatible and that some sets of principles imply other principles. We also contributed by defining new semantics that satisfy those principles: weighted max-based semantics, weighted card-based semantics and weighted h-categorizer semantics. We studied those semantics and proved that the iterative algorithm that allows to calculate the arguments' acceptability degrees converges. We also applied the principle-based approach in our work on the compensation-based semantics, where several weaker attacks have the same effect on the target as one strong attack (Amgoud et al., 2016). We defined a broad class of argumentation semantics that allow the user to choose the degree of compensation by setting the value of a parameter and showed that all those semantics satisfy the desirable principles.

Another important question is that of the *link* between the ranking-based and extension-based semantics. One of the most obvious differences is the output format: when using a ranking-based semantics, the output is a *ranking* on the arguments; in the case of extension-based semantics, the output is a *set of extensions*. While ranking-based semantics provide a finer evaluation of individual arguments, the question “what are the points of view of the argumentation graph?” stays unanswered when using such a semantics. We defined a general and modular framework that allows the user to choose any ranking-based semantics and to customize the acceptability conditions on extensions. The framework then calculates possible points of view (i.e. extensions) with respect to the given ranking-based semantics (Yun et al., 2018d).

1.3.2 Comparing argument-based and non argument-based approaches

Using principles is not the only way to understand the reasoning approaches from an area. It is very important to *compare* different approaches. For example, when an argument-based approach gives the same result as another existing non argument-based approach, e.g. that based on the maximal consistent subsets of formulas of a knowledge base (Rescher and Manor, 1970), this helps to “validate” and better understand the argumentation approach. I identified the classes of attack relations and semantics that yield the same result as those of non argument-based approaches (Vesic, 2013). Another situation is the one when an argument-based approach returns a result different from all existing approaches. We identified two classes of approaches: the class of argumentation systems returning inconsistent results (Vesic, 2013), containing argumentation systems that were not well-defined; and the class of argumentation systems returning consistent results different from existing approaches (Vesic and van der Torre, 2012). Furthermore, we showed that the two classes are non empty.

Another area that uses the reasoning based on maximal consistent subsets of a knowledge base is ontology based data access (OBDA). We are given a set of facts, a set of rules and a set of negative constraints (Poggi et al., 2008). A set of facts is consistent if no negative constraint can be deduced from this set by applying rules. Maximal consistent subsets of the set of facts are called *repairs*. To infer from repairs, different semantics have been defined in the literature (e.g. AR, IAR, ICR). Since argumentation theory can also be used to draw conclusions under inconsistency our goal was to explore the similarities and differences of OBDA semantics and argumentation semantics (Croitoru and Vesic, 2013). In particular, we proved that

- skeptical acceptance under stable and preferred semantics corresponds to ICR semantics
- universal acceptance under stable and preferred semantics corresponds to AR semantics
- acceptance under grounded semantics corresponds to IAR semantics

We also showed that the argumentation framework we define satisfies the desirable principles from the literature (e.g. consistency, closure).

1.3.3 Allowing to infer more conclusions

Once the extensions are calculated, the common strategy is to accept the arguments that appear in all extensions (those are called skeptical arguments). An alternative is to accept all the arguments that belong to at least one extension, called credulous arguments. The problem with the second option is that the union of all credulous arguments is not conflict-free (i.e. there are some attacks between them). That is why the first option (using skeptical arguments) is the preferred one. In case of repairs, there are several options, for example close the repairs with respect to the rules and then infer from their intersection. A potential problems is that having too many extensions (resp. repairs) can yield few or no arguments (formulas) being accepted. Consequently, a significant part of my work was to allow for more conclusions to be drawn.

We proposed two different approaches for dealing with this problem in argumentation (Konieczny et al., 2015). The first idea is to reduce the number of extensions, the second is to define a new inference policy, which allows to infer more arguments than the skeptical policy while still staying conflict-free. There are several ways to realize the first idea (reducing the number of extensions). We can compare pairs of extensions and then choose the best extensions given the winners of the pairwise comparisons (e.g. by using the Copeland's method). Another method is to attach a score to each extension and then keep only those maximizing the scores. An example of the score of an extension is the number obtained by an aggregation operator by using the scores of arguments. As an example of argument score, consider the number of extensions an argument belongs to. The second idea is to define a new inference policy, that we call *supported inference*. The first phase is to construct *candidate sets*, which are conflict-free sets of arguments containing the best arguments. Then, an argument is *supportedly accepted* if it belongs to all candidate sets. We show that each skeptical argument is supported, and that each supported argument is credulous. Furthermore, the set of supportedly accepted arguments is guaranteed to be conflict-free.

We also studied the question of decreasing the number of repairs (Yun et al., 2018e). We used different ways to achieve this. For example, we might attach an inconsistency value (Hunter and Konieczny, 2010) to each formula and then use those values (and an aggregation function) to compare the repairs. We defined a modular framework and the principles it should satisfy. We identified the conditions on the framework's components that guarantee that the principles will be satisfied.

We also defined a *class of inference relations* that allow to infer more formulas by selecting only the best maximal consistent subsets of an inconsistent knowledge base (Konieczny et al., 2018, 2019). We studied the properties of those inference relations and showed that they satisfy the desirable principles from the literature, namely they are preferential (Kraus et al., 1990).

1.3.4 Properties of judgment aggregation rules

Another part of my research is on *judgment aggregation*, which studies the problems related to aggregating a finite set of individual judgments, cast on a collection of logically interrelated issues. While each individual has a consistent judgment set, the question is how to aggregate them in order to obtain a *consistent collective judgment set*, which should represent the opinions of the individuals. Several judgment aggregation operators were developed and analyzed. The suitability of an operator can be determined by the set of properties it satisfies. The goal of my research in judgment aggregation is also to study the *properties* of operators, since they allow to better understand operators and links between them.

The literature on judgment aggregation is moving from proving impossibility results to studying particular judgment aggregation rules. We decided to make a structured list of virtually all rules that have been proposed and studied in the literature recently, together with properties of such rules (Lang et al., 2017). We first focus on the majority preservation property, which generalizes Condorcet-consistency, and identify which of the rules satisfy it. We study the inclusion relationships that hold between the rules. We consider different properties: two forms of unanimity, monotonicity, homogeneity, separability and reinforcement and study which rules satisfy those properties.

2. Argumentation

In this chapter, we are interested in computational models of argumentation, namely in the formalization introduced by Dung (1995). For a survey of the area, the reader is referred to the existing handbooks (Besnard and Hunter, 2008; Rahwan and Simari, 2009; Baroni et al., 2018a).

Definition 2.0.1 — Argumentation graph. An *argumentation graph* is a couple $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ where $\mathscr{A}$ is a finite set and $\mathscr{R} \subseteq \mathscr{A} \times \mathscr{A}$. The elements of $\mathscr{A}$ are called *arguments* and $\mathscr{R}$ is called *attack relation*. We say that *a attacks b* if $(a,b) \in \mathscr{R}$; in that case we also write $a\mathscr{R}b$. For a set $S \subseteq \mathscr{A}$ and an argument $a \in \mathscr{A}$, we say that S attacks a if there exists $b \in S$ such that $b\mathscr{R}a$; we say that a attacks S if there exists $b \in S$ such that $a\mathscr{R}b$. We say that S attacks a set P if there exist $a \in S$, $b \in P$ such that a attacks b. We say that a indirectly attacks b if and only if there is an odd-length path from a to b with respect to the attack relation. We say that S is without indirect conflicts if and only if there exist no $x, y \in S$ such that x indirectly attacks y. The set of all argumentation graphs is denoted by $\mathscr{A}\mathscr{F}$.

In what follows, $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ stands for an argumentation graph.

Definition 2.0.2 — Projection, direct sum, subgraph. For an argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and a set $S \subseteq \mathscr{A}$, we define: $\mathscr{F} \downarrow_S = (S, \mathscr{R} \cap (S \times S))$. Let $\mathscr{F}_1 = (\mathscr{A}_1, \mathscr{R}_1)$ and $\mathscr{F}_2 = (\mathscr{A}_2, \mathscr{R}_2)$ be two argumentation graphs. We define $\mathscr{F}_1 \cup \mathscr{F}_2 = (\mathscr{A}_1 \cup \mathscr{A}_2, \mathscr{R}_1 \cup \mathscr{R}_2)$. We write $\mathscr{F}_1 \subseteq \mathscr{F}_2$ if and only if $\mathscr{A}_1 \subseteq \mathscr{A}_2$ and $\mathscr{R}_1 \subseteq \mathscr{R}_2$.

Now we define the notion of semantics. It is a function that, given an argumentation graph $(\mathscr{A}, \mathscr{R})$, returns a set of subsets of $\mathscr{A}$. For a set S, we denote its powerset by 2^S.

Definition 2.0.3 — Extension-based semantics. An extension-based semantics is a function σ such that for every argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, we have $\sigma(\mathscr{F}) \in 2^{2^{\mathscr{A}}}$. The elements of $\sigma(\mathscr{F})$ are called *extensions*.

Intuitively, each extension represents a coherent point of view. Our definition requires a semantics to satisfy universal domain, i.e. to be defined for every argumentation graph. We could give a more general definition, thus allowing a semantics to be defined only for some argumentation graphs; we do not do that in order to simplify the notations. All extension-based semantics from the literature are defined on all argument graphs[1] whereas some ranking-based semantics are defined for acyclic graphs only. However, they are not in the center of our attention in this habilitation, thus we do not introduce specific notations to deal with this issue.

2.1 The principle-based approach to extension-based semantics

Formal argumentation theory, following the methodology in non-monotonic logic, logic programming and belief revision, defines a diversity of semantics. This immediately raises the same questions that were raised before for voting rules, and in many other areas. How do we know that the currently considered set of semantics is sufficient or complete? May there be a better semantics that has not been discovered yet? If there are many semantics, then how to choose one semantics from this set of alternatives in a particular application? How to guide the search for new and hopefully better argumentation semantics?

The principle-based approach addresses those problems. For example, if one needs to exclude the possibility of multiple extensions, one may choose the grounded or ideal semantics. If it is important that at least some extension is available, then stable semantics should not be used. As another common example, consider the admissibility principle that if an argument in an extension is attacked, then it is defended against this attack by another argument in the extension. If one needs a semantics that is admissible, then for example CF2 or stage2 cannot be chosen.

Principles have also been used to guide the search for new semantics. For example, the principle of resolution was defined by Baroni and Giacomin (2007), well before resolution based semantics were defined and studied by Baroni et al. (2011b). Likewise it may be expected that the existing and new principles will guide further search for suitable argumentation semantics. For example, consider the conflict-freeness principle that says that an extension does not contain arguments attacking each other. All semantics we study in this section satisfy this property. If one needs to define new argumentation semantics that are para-consistent in the sense that its extensions are not necessarily conflict free (Arieli, 2015), then one can still adopt other principles such as admissibility in the search for such para-consistent semantics.

The principle-based approach consists of three steps. The first step is to define a general function, which will be the object of study. Kenneth Arrow defined social welfare functions from preference profiles to aggregated preference orders. For abstract argumentation, the obvious candidate is a function from graphs to sets of sets of nodes of the graph. We call the nodes of the graph arguments, we call sets of nodes extensions, we call the edges attacks, and we call the graphs themselves argumentation graphs. Moreover, we call the function an argumentation semantics. Obviously nothing hinges on this terminology, and in principle the developed theory could be used for other applications of graph theory as well. We call this function from argumentation graphs to sets of extensions a two valued function, as a node is either in the extension, or not. Also multi-valued functions are commonly used, in particular three-valued functions conventionally called labelings. For three-valued labelings, the values are usually called *in*, *out* and *und* (which stands for "undecided"). Other settings have been considered in abstract argumentation, for example in value based argumentation, bipolar argumentation, abstract dialectical frameworks, input/output frameworks, ranked semantics, and more. The principle-based approach can be applied to all of them.

The second step of the principle-based approach is to define the principles. The central relation of the principle-based approach is the relation between semantics and principles. In abstract argumentation a two valued relation is used, such that every semantics either satisfies a given property or not. In this case, principles can be defined also as sets of semantics, and they can be

[1] stable semantics is defined on all argumentation frameworks, it is just that in some cases the set of extensions is empty

represented by a constraint on the function from argumentation graphs to sets of extensions. An alternative approach would be to use a numerical value to represent to which degree a semantics satisfies a principle.

The third step of the principle-based approach is to classify and study sets of principles. For example, a set of principles may imply another one, or a set of principles may be satisfiable in the sense that there is a semantics that satisfies all of them. A particular useful challenge is to find a set of principles that characterises a semantics, in the sense that the semantics is the only one that satisfies all the principles. Such characterisations are sometimes called representation theorems.

The rest of the section is organised as follows. We first introduce the well-known semantics from the argumentation literature; then, we introduce the principles; finally, we study which semantics satisfies which principles. For more details, the reader is referred to the book chapter (van der Torre and Vesic, 2018).

2.1.1 Semantics

This section introduces different argumentation semantics we study in the rest of the chapter. Note that most of the properties from the literature, which we study in Subsection 2.1.2, can appear in two variants: extension-based and labelling-based. We present their versions for the extension-based approach.

We start by introducing the notions of conflict-freeness and admissibility.

Definition 2.1.1 — Conflict-freeness, admissibility, strong admissibility. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and $S \subseteq \mathscr{A}$. Set S is *conflict-free* in $\mathscr{F}$ if and only if for every $a, b \in S$, $(a, b) \notin \mathscr{R}$.

Argument $a \in \mathscr{A}$ is *defended* by set S if and only if for every $b \in \mathscr{A}$ such that $b\mathscr{R}a$ there exists $c \in S$ such that $c\mathscr{R}b$. Argument $a \in \mathscr{A}$ is *strongly defended* by set S if and only if for every $b \in \mathscr{A}$ such that $b\mathscr{R}a$ there exists $c \in S \setminus \{a\}$ such that $c\mathscr{R}b$ and c is strongly defended by $S \setminus \{a\}$. S is *admissible* in $\mathscr{F}$ if and only if it is conflict-free and it defends all its arguments. S is *strongly admissible* in $\mathscr{F}$ if and only if it is conflict-free and it strongly defends all its arguments. For a set $S \subseteq \mathscr{A}$, we denote by S^+ the set $S^+ = \{a \in \mathscr{A} \mid S \text{ attacks } a\}$.

Stable, complete, preferred and grounded semantics were introduced by Dung (1995):

Definition 2.1.2 — Complete, stable, grounded, preferred semantics. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and $S \subseteq \mathscr{A}$.

- Set S is a *complete extension* of $\mathscr{F}$ if and only if it is conflict-free, it defends all its arguments and it contains all the arguments it defends.
- Set S is a *stable extension* of $\mathscr{F}$ if and only if it is conflict-free and it attacks all the arguments of $\mathscr{A} \setminus S$.
- S is the *grounded extension* of $\mathscr{F}$ if and only if it is a minimal with respect to set inclusion complete extension of $\mathscr{F}$.
- S is a *preferred extension* of $\mathscr{F}$ if and only if it is a maximal with respect to set inclusion admissible set of $\mathscr{F}$.

Dung (1995) shows that each argumentation graph has a unique grounded extension. Stable extensions do not always exist, i.e. there exist argumentation graphs whose set of stable extensions is empty. Semi-stable semantics (Verheij, 1996; Caminada, 2006a) guarantees that every argumentation graph has an extension. Furthermore, semi-stable semantics coincides with stable semantics on argumentation graphs that have at least one stable extension.

Definition 2.1.3 — Semi-stable semantics. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and $S \subseteq \mathscr{A}$. Set S is a semi-stable extension of $\mathscr{F}$ if and only if it is a complete extension and $S \cup S^+$ is maximal with respect to set inclusion among complete extensions, i.e. there exists no complete extension S_1 such that $S \cup S^+ \subset S_1 \cup S_1^+$.

Ideal semantics (Dung et al., 2007) is an alternative to grounded semantics. Like grounded semantics, ideal semantics always returns a unique extension, which is also a complete extension (Dung et al., 2007). From the definition of the grounded semantics, we conclude that the ideal extension is a superset of the grounded extension. Ideal semantics is thus less skeptical than grounded semantics.

Definition 2.1.4 — Ideal semantics. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and $S \subseteq \mathscr{A}$. Set S is the ideal extension of $\mathscr{F}$ if and only if it is a maximal with respect to set inclusion admissible subset of every preferred extension.

We now introduce eager semantics (Caminada, 2007).

Definition 2.1.5 — Eager semantics. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and $S \subseteq \mathscr{A}$. Set S is the eager extension of $\mathscr{F}$ if and only if it is the maximal with respect to set inclusion admissible subset of every semi-stable extension.

Caminada (2007) shows that each argumentation graph has a unique eager extension and that the eager extension is also a complete extension. Note that eager semantics is similar to ideal semantics: the ideal extension is the unique biggest admissible subset of every preferred extension; the eager extension is the unique biggest admissible subset of each semi-stable extension. Since each semi-stable extension is a preferred extension (Caminada, 2006b), the eager extension is a superset of the ideal extension.

The definitions of other semantics can be found in the book chapter (van der Torre and Vesic, 2018).

We focus on the extension-based approach, which means that each semantics is defined by specifying the extensions it returns for a given argumentation graph. There exists an alternative, labelling-based approach. Instead of calculating extensions, it provides labellings, one labelling being a function that attaches to every argument a label *in*, *out* or *und*.

Definition 2.1.6 — Labelling-based semantics. Let $\Lambda = \{in, out, und\}$. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation graph. A labelling on $\mathscr{F}$ is a total function $\mathscr{L}ab : \mathscr{A} \to \Lambda$. A labelling-based semantics is a function λ such that for every argumentation graph $\mathscr{F}$, we have that $\lambda(\mathscr{F})$ is a set of labellings on $\mathscr{F}$.

To illustrate, let us provide a labelling-based definition of complete semantics.

Definition 2.1.7 — Complete labelling. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and $\mathscr{L}ab$ a labelling on $\mathscr{F}$. We say that $\mathscr{L}ab$ is a complete labelling if and only if for every $a \in \mathscr{A}$:

- if a is labelled in then all its attackers are labelled out
- if a is labelled out then none of its attackers is labelled in
- if a is labelled und then not all its attackers are labelled out and none of its attackers is labelled in.

We denote by $in(\mathscr{L}ab)$ (resp. $out(\mathscr{L}ab)$, $und(\mathscr{L}ab)$) the set of arguments labelled *in* (resp. *out*, *und*).

For every $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, the set of complete extensions under σ is exactly the set $\{in(\mathscr{L}ab) \mid \mathscr{L}ab$ is a complete labelling$\}$.

Moreover, there exists a general way that allows to obtain a labelling-based definition of a semantics given its extension-based definition, under the condition that the semantics returns conflict-free sets.

Definition 2.1.8 — Extension to labelling. Given an extension $\mathscr{E}$, labelling $\mathscr{L}ab_{\mathscr{E}}$ is defined as follows: $\mathscr{L}ab_{\mathscr{E}}(a) = in$ if $a \in \mathscr{E}$, $\mathscr{L}ab_{\mathscr{E}}(a) = out$ if $a \in \mathscr{E}^+$, $\mathscr{L}ab_{\mathscr{E}}(a) = und$ otherwise. Then, given a semantics σ, we say that $\mathscr{L}ab$ is a σ labelling of $\mathscr{F}$ if and only if there exists

$\mathscr{E} \in \sigma(\mathscr{F})$ such that $\mathscr{L}ab = \mathscr{L}ab_{\mathscr{E}}$.

Other ways to obtain a labelling from an extension are possible. For example, we could say that an argument is *out* if it is attacked by an argument in the extension, or it attacks an argument in the extension. This would make the definition of *out* more symmetric and more in line with naive based semantics[2]. However, it seems such alternatives have not been explored systematically in the literature. Moreover, even if extension and labelling based semantics are inter-translatable, it may affect other definitions such as equivalence of graphs. Finally, using Definition 2.1.8, every principle defined in terms of extension based semantics can be translated into labelings and vice versa, though one of the definitions may be more compact or intuitive than the other.

We saw an intuitive way to define complete labellings in Definition 2.1.7. Intuitive labelling-based definitions of other semantics also exist in the literature. For example: a grounded labelling is a complete labelling such that the set of arguments labelled *in* is minimal with respect to set inclusion among all complete labellings; a stable labelling is a complete labelling such that the set of undecided arguments is empty; a preferred labelling is a complete labelling such that the set of arguments labelled *in* is maximal with respect to set inclusion among all complete labellings. The reader interested in more details about the labelling-based approach is referred to the paper by Baroni et al. (2011a).

2.1.2 Principles

This section presents the properties from the literature and reviews all the semantics with respect to the properties. For completeness, the tables contain the results due to other authors. The reader interested to know the exact paper where each proof was presented is invited to consult detailed discussions and references in the book chapter (van der Torre and Vesic, 2018).

Definition 2.1.9 — Isomorphic argumentation graphs. Two argumentation graphs $\mathscr{F}_1 = (\mathscr{A}_1, \mathscr{R}_1)$ and $\mathscr{F}_2 = (\mathscr{A}_2, \mathscr{R}_2)$ are *isomorphic* if and only if there exists a bijective function $m : \mathscr{A}_1 \to \mathscr{A}_2$, such that $(a,b) \in \mathscr{R}_1$ if and only if $(m(a), m(b)) \in \mathscr{R}_2$. This is denoted by $\mathscr{F}_1 \doteq_m \mathscr{F}_2$.

The first property, called "language independence" by Baroni and Giacomin (2007) is an obvious requirement for argumentation semantics. It is sometimes called abstraction (Amgoud and Ben-Naim, 2013; Bonzon et al., 2016a) or anonymity (Amgoud et al., 2016).

Principle 1 — Language independence. A semantics σ satisfies the *language independence* principle if and only if for every two argumentation graphs $\mathscr{F}_1$ and $\mathscr{F}_2$, if $\mathscr{F}_1 \doteq_m \mathscr{F}_2$ then $\sigma(\mathscr{F}_2) = \{m(\mathscr{E}) \mid \mathscr{E} \in \sigma(\mathscr{F}_1)\}$.

It is immediate to see that all the semantics satisfy language independence, since the definitions of semantics take into account only the topology of the graph, and not the arguments' names.

Conflict-freeness is one of the basic principles. Introduced by Dung (1995) and stated as a principle by Baroni and Giacomin (2007), it is satisfied all the semantics we consider here. However, note that one can define a non conflict-free semantics (Arieli, 2015). As another example of relaxing conflict-freeness consider the work by Dunne et al. (2011), who introduce a framework where each attack is associated a weight; given an inconsistency budget β, they accept to disregard the set of attacks up to total weight of β.

Principle 2 — Conflict-freeness. A semantics σ satisfies the *conflict-freeness* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$, $\mathscr{E}$ is conflict-free set in $\mathscr{F}$.

Defence is a well-known property introduced by Dung (1995).

[2]see Principle 6 on page 18

Principle 3 — Defence. A semantics σ satisfies the *defence* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$, for every $a \in \mathscr{E}$, $\mathscr{E}$ defends a.

Baroni and Giacomin (2007) suppose that every extension is conflict-free. Thus an extension defends all it arguments if and only if it is admissible. However, if conflict-freeness is seen as an optional criterion, we can distinguish between the principles of admissibility and defence.

Principle 4 — Admissibility. A semantics σ satisfies the *admissibility* principle if and only if for every argumentation graph $\mathscr{F}$, every $\mathscr{E} \in \sigma(\mathscr{F})$ is admissible in $\mathscr{F}$.

Observation 1 If a semantics σ satisfies admissibility it also satisfies conflict-freeness and defence.

We now study the notion of strong admissibility (Baroni and Giacomin, 2007).

Principle 5 — Strong admissibility. A semantics σ satisfies the *strong admissibility* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$ it holds that $a \in \mathscr{E}$ implies that $\mathscr{E}$ strongly defends a.

Observation 2 If a semantics σ satisfies strong admissibility then it satisfies admissibility.

Another principle, which we call *naivety*, says that every extension under semantics σ is a naive extension.

Principle 6 — Naivety. A semantics σ satisfies the *naivety* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$, $\mathscr{E}$ is a maximal for set inclusion conflict-free set in $\mathscr{F}$.

Coste-Marquis et al. (2005) introduced prudent semantics, which are based on the notion of indirect conflict-freeness. The definition of those semantics can be found in the book chapter (van der Torre and Vesic, 2018).We now introduce the corresponding principle.

Principle 7 — Indirect conflict-freeness. A semantics σ satisfies the *indirect conflict-freeness* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$, $\mathscr{E}$ is without indirect conflicts in $\mathscr{F}$.

Observation 3 If a semantics σ satisfies indirect conflict-freeness then it satisfies conflict-freeness.

	Defence	Admiss.	Strong admiss.	Naivety	Indirect CF	Reinst.	Weak reinst.	CF--reinst.
complete	✓	✓	×	×	×	✓	✓	✓
grounded	✓	✓	✓	×	×	✓	✓	✓
preferred	✓	✓	×	×	×	✓	✓	✓
stable	✓	✓	×	✓	×	✓	✓	✓
semi-stable	✓	✓	×	×	×	✓	✓	✓
ideal	✓	✓	×	×	×	✓	✓	✓
eager	✓	✓	×	×	×	✓	✓	✓
p-complete	✓	✓	×	×	✓	×	×	×
p-grounded	✓	✓	✓	×	✓	×	×	×
p-preferred	✓	✓	×	×	✓	×	×	×
p-stable	✓	✓	×	✓	✓	✓	✓	✓
naive	×	×	×	✓	×	×	×	✓
CF2	×	×	×	✓	×	×	✓	✓
stage	×	×	×	✓	×	×	×	✓
stage2	×	×	×	✓	×	×	✓	✓

Table 2.1: Properties of semantics: basic properties, admissibility and reinstatement

Defence says that an extension must defend all the arguments it contains. Reinstatement can be seen as its counterpart, since it says that an extension must contain all the arguments it defends.

This principle was first studied in a systematic way by Baroni and Giacomin (2007).

Principle 8 — Reinstatement. A semantics σ satisfies the *reinstatement* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$, for every $a \in \mathscr{A}$ it holds that if $\mathscr{E}$ defends a then $a \in \mathscr{E}$.

Baroni and Giacomin (2007) study another property called weak reinstatement.

Principle 9 — Weak reinstatement. A semantics σ satisfies the *weak reinstatement* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$ it holds that

$\mathscr{E}$ strongly defends a implies $a \in \mathscr{E}$.

Observation 4 If a semantics σ satisfies reinstatement then it satisfies weak reinstatement.

The reinstatement principle makes sure that as soon as an argument a is defended by an extension $\mathscr{E}$, a should belong to $\mathscr{E}$—without specifying that a is not in conflict with arguments of $\mathscr{E}$. To take this into account, another principle was defined by Baroni and Giacomin (2007).

Principle 10 — $\mathscr{CF}$-reinstatement. A semantics σ satisfies the *$\mathscr{CF}$-reinstatement* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E} \in \sigma(\mathscr{F})$, for every $a \in \mathscr{A}$ it holds that if $\mathscr{E}$ defends a and $\mathscr{E} \cup \{a\}$ is conflict-free then $a \in \mathscr{E}$.

Observation 5 If a semantics σ satisfies reinstatement then it satisfies $\mathscr{CF}$-reinstatement.

The next principle was first considered by Baroni and Giacomin (2007). It says that an extension cannot contain another extension.

Principle 11 — I-maximality. A semantics σ satisfies the *I-maximality* principle if and only if for every argumentation graph $\mathscr{F}$, for every $\mathscr{E}_1, \mathscr{E}_2 \in \sigma(\mathscr{F})$, if $\mathscr{E}_1 \subseteq \mathscr{E}_2$ then $\mathscr{E}_1 = \mathscr{E}_2$.

We next consider the allowing abstention principle (Baroni et al., 2011a). Roughly speaking, it says that if there exists a labelling where an argument is labelled *in* and there exists a labelling where the same argument is labelled *out*, there should be a labelling where that argument is labelled *und*.

Principle 12 — Allowing abstention. A semantics σ satisfies the *allowing abstention* principle if and only if for every argumentation graph $\mathscr{F}$, for every $a \in \mathscr{A}$, if there exist two extensions $\mathscr{E}_1, \mathscr{E}_2 \in \sigma(\mathscr{F})$ such that $a \in \mathscr{E}_1$ and $a \in \mathscr{E}_2^+$ then there exists an extension $\mathscr{E}_3 \in \sigma(\mathscr{F})$ such that $a \notin (\mathscr{E}_3 \cup \mathscr{E}_3^+)$.

Observe that unique status semantics trivially satisfy this principle. Allowing abstention is thus satisfied by grounded, ideal, eager and p-grounded semantics.[3]

To define crash resistance (Caminada et al., 2012), we first need to introduce the following two definitions.

Definition 2.1.10 — Disjoint argumentation graphs. Two argumentation graphs $\mathscr{F}_1 = (\mathscr{A}_1, \mathscr{R}_1)$ and $\mathscr{F}_2 = (\mathscr{A}_2, \mathscr{R}_2)$ are disjoint if and only if $\mathscr{A}_1 \cap \mathscr{A}_2 = \emptyset$.

A graph $\mathscr{F}^\star$ is contaminating if joining $\mathscr{F}^\star$ with an arbitrary disjoint graph $\mathscr{F}$ results in a graph $\mathscr{F} \cup \mathscr{F}^\star$ having the same extensions as $\mathscr{F}^\star$. The intuition behind this definition is that $\mathscr{F}^\star$ contaminates every graph.

Definition 2.1.11 — Contaminating. An argumentation graph $\mathscr{F}^\star$ is contaminating for a semantics σ if and only if for every argumentation graph $\mathscr{F}$ disjoint from $\mathscr{F}^\star$ it holds that

[3]Note that Table 2 by Baroni et al. (2011a) specifies that grounded semantics does not satisfy dilemma abstaining. The reason is that Baroni et al. consider the property as being "non-applicable" to unique status semantics (personal communication, 2016).

$\sigma(\mathscr{F}^\star \cup \mathscr{F}) = \sigma(\mathscr{F}^\star)$.

A semantics is crash resistant if and only if there are no contaminating graphs. The intuition behind this name is that a contaminating graph causes the system to crash.

Principle 13 — Crash resistance. A semantics σ satisfies the *crash resistance* principle if and only if there are no contaminating argumentation graphs for σ.

Crash resistance forbids only the most extreme form of interferences between disjoint subgraphs. A stronger property, non-interference, was defined by Caminada et al. (2012). We first need to define a notion of isolated set, i.e. a set that neither attacks outside arguments nor is attacked by them.

Definition 2.1.12 — Isolated set of arguments. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation graph. A set $S \subseteq \mathscr{A}$ is isolated in $\mathscr{F}$ if and only if

$$((S \times (\mathscr{A} \setminus S)) \cup ((\mathscr{A} \setminus S) \times S)) \cap \mathscr{R} = \emptyset.$$

A semantics satisfies non-interference principle if for every isolated set S, the intersections of the extensions with set S coincide with the extensions of the restriction of the graph on S.

Principle 14 — Non-interference. A semantics σ satisfies the *non-interference* principle if and only if for every argumentation graph $\mathscr{F}$, for every set of arguments S isolated in $\mathscr{F}$ it holds that $\sigma(\mathscr{F}\downarrow_S) = \{\mathscr{E} \cap S \mid \mathscr{E} \in \sigma(\mathscr{F})\}$.

The previous principle can be made even stronger by considering the case when the set S is not attacked by the rest of the graph, but can attack the rest of the graph. Let us formalize the notion of an unattacked set.

Definition 2.1.13 — Unattacked arguments. Given an argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, a set U is *unattacked* if and only if there exists no $a \in \mathscr{A} \setminus U$ such that a attacks U. The set of unattacked sets in $\mathscr{F}$ is denoted $\mathtt{US}(\mathscr{F})$.

We can now define the principle of directionality, introduced by Baroni and Giacomin (2007).

Principle 15 — Directionality. A semantics σ satisfies the *directionality* principle if and only if for every argumentation graph $\mathscr{F}$, for every $U \in \mathtt{US}(\mathscr{F})$, it holds that $\sigma(\mathscr{F}\downarrow_U) = \{\mathscr{E} \cap U \mid \mathscr{E} \in \sigma(\mathscr{F})\}$.

Baroni et al. (2011a) show the following dependencies between directionality, interference and crash resistance.

Observation 6 Directionality implies non-interference, and non-interference implies crash resistance.

Let us now consider two variants of directionality, called weak directionality and semi-directionality and suggested by M. Giacomin (personal communication, 2016).

Principle 16 — Weak directionality. A semantics σ satisfies the *weak directionality* principle if and only if for every argumentation graph $\mathscr{F}$, for every $U \in \mathtt{US}(\mathscr{F})$, it holds that $\sigma(\mathscr{F}\downarrow_U) \supseteq \{\mathscr{E} \cap U \mid \mathscr{E} \in \sigma(\mathscr{F})\}$.

Principle 17 — Semi-directionality. A semantics σ satisfies the *semi-directionality* principle if and only if for every argumentation graph $\mathscr{F}$, for every $U \in \mathtt{US}(\mathscr{F})$, it holds that $\sigma(\mathscr{F}\downarrow_U) \subseteq \{\mathscr{E} \cap U \mid \mathscr{E} \in \sigma(\mathscr{F})\}$.

Observation 7 A semantics σ satisfies directionality if and only if σ satisfies both weak directionality and semi-directionality.

We now consider the six properties related to skepticism and resolution adequacy (Baroni and Giacomin, 2007). The first definition says that a set of extensions $\mathtt{Ext}_1$ is more skeptical than $\mathtt{Ext}_2$

	I-max.	Allowing abstention	Crash resistance	Non- -interference	Direct.	Weak- -direct.	Semi- -direct.
complete	×	✓	✓	✓	✓	✓	✓
grounded	✓	✓	✓	✓	✓	✓	✓
preferred	✓	×	✓	✓	✓	✓	✓
stable	✓	×	×	×	×	✓	×
semi-stable	✓	×	✓	✓	×	×	×
ideal	✓	✓	✓	✓	✓	✓	✓
eager	✓	✓	✓	✓	✓	✓	✓
p-complete	×	✓	✓	✓	×	×	✓
p-grounded	✓	✓	✓	✓	✓	✓	✓
p-preferred	✓	×	✓	✓	×	×	✓
p-stable	✓	×	×	×	×	✓	×
naive	✓	×	✓	✓	×	×	✓
CF2	✓	×	✓	✓	✓	✓	✓
stage	✓	×	✓	✓	×	×	×
stage2	✓	×	✓	✓	✓	✓	✓

Table 2.2: Properties of semantics, part 2

if the set of skeptically accepted arguments with respect to $\mathtt{Ext}_1$ is a subset of the set of skeptically accepted arguments with respect to $\mathtt{Ext}_2$.

Definition 2.1.14 — $\preceq^E_\cap$**.** Let $\mathtt{Ext}_1$ and $\mathtt{Ext}_2$ be two sets of sets of arguments. We say that $\mathtt{Ext}_1 \preceq^E_\cap \mathtt{Ext}_2$ if and only if

$$\bigcap_{\mathscr{E}_1 \in \mathtt{Ext}_1} \mathscr{E}_1 \subseteq \bigcap_{\mathscr{E}_2 \in \mathtt{Ext}_2} \mathscr{E}_2.$$

The previous definition compares only the intersections of extensions. A finer criterion was introduced by Baroni et al. (2004).

Definition 2.1.15 — $\preceq^E_W$**.** Let $\mathtt{Ext}_1$ and $\mathtt{Ext}_2$ be two sets of sets of arguments. We say that $\mathtt{Ext}_1 \preceq^E_W \mathtt{Ext}_2$ if and only if

$$\text{for every } \mathscr{E}_2 \in \mathtt{Ext}_2, \text{ there exists } \mathscr{E}_1 \in \mathtt{Ext}_1 \text{ such that } \mathscr{E}_1 \subseteq \mathscr{E}_2.$$

Baroni and Giacomin (2007) refine the previous relation by introducing the following definition.

Definition 2.1.16 — $\preceq^E_S$**.** Let $\mathtt{Ext}_1$ and $\mathtt{Ext}_2$ be two sets of sets of arguments. We say that $\mathtt{Ext}_1 \preceq^E_S \mathtt{Ext}_2$ if and only if $\mathtt{Ext}_1 \preceq^E_W \mathtt{Ext}_2$ and

$$\text{for every } \mathscr{E}_1 \in \mathtt{Ext}_1, \text{ there exists } \mathscr{E}_2 \in \mathtt{Ext}_2 \text{ such that } \mathscr{E}_1 \subseteq \mathscr{E}_2.$$

Letters W and S in the previous definitions stand for *weak* and *strong*. Baroni and Giacomin (2007) showed that the three relations are reflexive and transitive and that they are also in strict order of implication. Namely, given two sets of sets of arguments $\mathtt{Ext}_1$ and $\mathtt{Ext}_2$, we have

Observation 8

$$\mathtt{Ext}_1 \preceq^E_S \mathtt{Ext}_2 \text{ implies } \mathtt{Ext}_1 \preceq^E_W \mathtt{Ext}_2$$

$$\mathtt{Ext}_1 \preceq^E_W \mathtt{Ext}_2 \text{ implies } \mathtt{Ext}_1 \preceq^E_\cap \mathtt{Ext}_2$$

We now define a skepticism relation $\preceq^A$ between argumentation graphs. It says that $\mathscr{F}_1 \preceq^A \mathscr{F}_2$ if $\mathscr{F}_1$ may have some symmetric attacks where $\mathscr{F}_2$ has a directed attack.

Definition 2.1.17 — $\preceq^A$. Given an argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, the conflict set is defined as $\mathtt{CONF}(\mathscr{F}) = \{(a,b) \in \mathscr{A} \times \mathscr{A} \mid (a,b) \in \mathscr{R}$ or $(b,a) \in \mathscr{R}\}$. Given two argumentation graphs $\mathscr{F}_1 = (\mathscr{A}_1, \mathscr{R}_1)$ and $\mathscr{F}_2 = (\mathscr{A}_2, \mathscr{R}_2)$, we say that $\mathscr{F}_1 \preceq^A \mathscr{F}_2$ if and only if $\mathtt{CONF}(\mathscr{F}_1) = \mathtt{CONF}(\mathscr{F}_2)$ and $\mathscr{R}_2 \subseteq \mathscr{R}_1$.

Observe that $\preceq^A$ is a partial order, as it consists of an equality and a set inclusion relation (Baroni and Giacomin, 2007). Note that within the set of argumentation graphs comparable with a given argumentation graph $\mathscr{F}$, there might be several maximal elements with respect to $\preceq^A$, since there might be several ways to replace all symmetric attacks by asymmetric ones.

We can now introduce the skepticism adequacy principle. Its idea is that if $\mathscr{F}$ is more skeptical than $\mathscr{F}'$ then the set of extensions of $\mathscr{F}$ is more skeptical than that of $\mathscr{F}'$.

Principle 18 — Skepticism adequacy. Given a skepticism relation $\prec^E$ between sets of sets of arguments, a semantics σ satisfies the $\preceq^E$-*skepticism adequacy* principle if and only if for every two argumentation graphs $\mathscr{F}$ and $\mathscr{F}'$ such that $\mathscr{F} \preceq^A \mathscr{F}'$ it holds that $\sigma(\mathscr{F}) \preceq^E \sigma(\mathscr{F}')$.

For example, if $\mathscr{F}$ consists of two arguments a and b attacking each other and $\mathscr{F}'$ has only an attack from a to b, then the intersection of the extensions of $\mathscr{F}$ ($\emptyset$ for all semantics) is a subset of extensions of $\mathscr{F}'$, typically $\{a\}$. Roughly speaking: the more symmetric attacks we replace, the more we know, but we do not loose any accepted arguments.

Observation 9

- If σ satisfies $\preceq^E_S$-sk. adequacy then it satisfies $\preceq^E_W$-sk. adequacy
- If σ satisfies $\preceq^E_W$-sk. adequacy then it satisfies $\preceq^E_\cap$-sk. adequacy

	$\preceq^E_\cap$-sk. ad.	$\preceq^E_W$-sk. ad.	$\preceq^E_S$-sk. ad.	$\preceq^E_\cap$-res. ad.	$\preceq^E_W$-res. ad.	$\preceq^E_S$-res. ad.
complete	✓	✓	×	×	×	×
grounded	✓	✓	✓	×	×	×
preferred	×	×	×	✓	✓	✓
stable	✓	✓	×	✓	✓	✓
semi-stable	×	×	×	✓	✓	×
ideal	×	×	×	×	×	×
eager	×	×	×	×	×	×
p-complete	×	×	×	×	×	×
p-grounded	×	×	×	✓	×	×
p-preferred	×	×	×	×	×	×
p-stable	×	×	×	✓	✓	×
naive	✓	✓	✓	✓	✓	✓
CF2	✓	✓	×	×	×	×
stage	×	×	×	✓	✓	×
stage2	×	×	×	×	×	×

Table 2.3: Properties of semantics, skepticism and resolution adequacy

Let us now consider resolution adequacy (Baroni and Giacomin, 2007). The idea is, roughly speaking, to consider the set RES, which consists of all argumentation graphs where all the conflicts were "resolved".

Definition 2.1.18 — RES. We denote by $RES(\mathscr{F})$ the set of all argumentation graphs comparable with $\mathscr{F}$ and maximal with respect to $\preceq^A$.

Then, we are interested in the set UR, which contains the union of extensions of all the graphs from RES.

Definition 2.1.19 — UR. Given an argumentation graph $\mathscr{F}$ and a semantics σ, we define $UR(\mathscr{F},\sigma) = \bigcup_{\mathscr{F}' \in RES(\mathscr{F})} \sigma(\mathscr{F}')$.

Resolution adequacy (Baroni and Giacomin, 2007) aims at comparing the set of extensions of an argumentation graph with the corresponding UR. For this, any relation that compares two sets of extensions may be used.

Principle 19 — Resolution adequacy.
Given a skepticism relation $\preceq^E$ between sets of sets of arguments, a semantics σ satisfies the $\preceq^E$-*resolution adequacy* principle if and only if for every argumentation graph $\mathscr{F}$ we have $UR(\mathscr{F},\sigma) \preceq^E \sigma(\mathscr{F})$.

We consider three variants of the resolution adequacy principle: $\preceq^E_\cap$-resolution adequacy, $\preceq^E_W$-resolution adequacy and $\preceq^E_S$-resolution adequacy.

Observation 10

- If σ satisfies $\preceq^E_S$-res. adequacy then it satisfies $\preceq^E_W$-res. adequacy
- If σ satisfies $\preceq^E_W$-res. adequacy then it satisfies $\preceq^E_\cap$-res. adequacy

	Succinctness	Tightness	Conflict- -sensitiveness	Com- -closure	SCC- -recursiveness	Cardinality
complete	×	×	×	✓	✓	1+
grounded	×	✓	✓	✓	✓	1
preferred	×	×	✓	✓	✓	1+
stable	×	✓	✓	✓	✓	0+
semi-stable	×	×	✓	✓	×	1+
ideal	×	✓	✓	✓	×	1
eager	×	✓	✓	✓	×	1+
p-complete	×	×	×	×	×	1+
p-grounded	×	✓	✓	✓	×	1
p-preferred	×	✓	✓	✓	×	1+
p-stable	×	✓	✓	✓	×	0+
naive	×	✓	✓	✓	×	1+
CF2	✓	✓	✓	✓	✓	1+
stage	×	✓	✓	✓	×	1+
stage2	✓	✓	✓	✓	✓	1+

Table 2.4: Properties of semantics, part 4

Baroni et al. (2011b) introduce resolution-based family of semantics, which are developed to satisfy the resolution properties. Let us now consider the last group of properties listed in Table 2.4. We first need to define the notion of strong equivalence (Oikarinen and Woltran, 2010). Basic equivalence means that the two graphs have the same extensions. Strong equivalence is a dynamic notion. It says that whatever new arguments arrive to both graphs, they will still be equivalent. More formally, two graphs $\mathscr{F}_1$ and $\mathscr{F}_2$ are strongly equivalent if for every argumentation graph $\mathscr{F}_3$, we have that $\mathscr{F}_1 \cup \mathscr{F}_3$ has the same extensions as $\mathscr{F}_2 \cup \mathscr{F}_3$.

Definition 2.1.20 — Strong equivalence. Two argumentation graphs $\mathscr{F}_1$ and $\mathscr{F}_2$ are strongly equivalent with respect to semantics σ, in symbols $\mathscr{F}_1 \equiv^\sigma_S \mathscr{F}_2$ if and only if for each argumentation graph $\mathscr{F}_3$, $\sigma(\mathscr{F}_1 \cup \mathscr{F}_3) = \sigma(\mathscr{F}_2 \cup \mathscr{F}_3)$.

An attack is redundant in $\mathscr{F}$ if removing it does not change the extensions of any $\mathscr{F}'$ that contains $\mathscr{F}$.

Definition 2.1.21 — Redundant attack. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation graph and σ a semantics. Attack $(a,b) \in \mathscr{R}$ is said to be redundant in $\mathscr{F}$ with respect to σ if and only if for all argumentation graphs $\mathscr{F}'$ such that $\mathscr{F} \subseteq \mathscr{F}'$ we have $\sigma(\mathscr{F}') = \sigma(\mathscr{F}' \setminus (a,b))$.

We can now define the succinctness principle (Gaggl and Woltran, 2013).

Principle 20 — Succinctness. A semantics σ satisfies the *succinctness* principle if and only if no argumentation graph contains a redundant attack with respect to σ.

Gaggl and Woltran (2013) show that a semantics σ satisfies succinctness if and only if for every two argumentation graphs $\mathscr{F}_1$ and $\mathscr{F}_2$ strong equivalence under σ coincides with $\mathscr{F}_1 = \mathscr{F}_2$. Note that this shows that succinctness is an extremely strong property. Namely, it can only be satisfied by a semantics σ if strong equivalence under σ is trivial, i.e. strong equivalence coincides with syntactic equivalence. It is not a surprise that out of all the semantics we consider, only two very particular semantics satisfy succinctness, namely CF2 and stage2.

The next principle we consider is tightness. Let us first define the notion of pairs. A couple (a,b) is in $\mathscr{P}airs$ if there is an extension containing both a and b.

Definition 2.1.22 — $\mathscr{P}airs$. Given a set of extensions $\mathscr{S} = \{\mathscr{E}_1, \ldots, \mathscr{E}_n\}$, we define

$$\mathscr{P}airs(\mathscr{S}) = \{(a,b) \mid \text{ there exists } \mathscr{E}_i \in \mathscr{S} \text{ such that } \{a,b\} \subseteq \mathscr{E}_i\}.$$

Tightness was introduced by Dunne et al. (2015). Roughly speaking, it says that if argument a does not belong to extension $\mathscr{E}$, then there must be an argument $b \in \mathscr{E}$ that is somehow incompatible with a.

Principle 21 — Tightness. A set of extensions $\mathscr{S} = \{\mathscr{E}_1, \ldots, \mathscr{E}_n\}$ is tight if and only if for every extension $\mathscr{E}_i$ and for every $a \in \mathscr{A}$ that appears in at least one extension from $\mathscr{S}$ it holds that if $\mathscr{E}_i \cup \{a\} \notin \mathscr{S}$ then there exists $b \in \mathscr{E}_i$ such that $(a,b) \notin \mathscr{P}airs(\mathscr{S})$.

A semantics σ satisfies the *tightness* principle if and only if for every argumentation graph $\mathscr{F}$, $\sigma(\mathscr{F})$ is tight.

Directly from the definition of tightness, we conclude that unique status semantics satisfy this principle.

Observation 11 If σ is a semantics that returns exactly one extension for every argumentation graph then σ satisfies tightness.

It can also be shown that if every extension under σ is a maximal conflict-free set, σ satisfies tightness.

We now study the notion of conflict-sensitiveness (Dunne et al., 2015). Note that an equivalent principle was called adm-closure in some papers. The idea is that if the union of two extensions does not form an extension, that is because there are two arguments – one from each extension – that never appear together in an extension. Thus, the semantics does not allow for the big extension since it is sensitive to the "conflict" between those arguments.

Principle 22 — Conflict-sensitiveness. A set of extensions $\mathscr{S} = \{\mathscr{E}_1, \ldots, \mathscr{E}_n\}$ is conflict-sensitive if and only if for every two extensions $\mathscr{E}_i$, $\mathscr{E}_j$ such that $\mathscr{E}_i \cup \mathscr{E}_j \notin \mathscr{S}$ it holds that there exist $a, b \in \mathscr{E}_i \cup \mathscr{E}_j$ such that $(a,b) \notin \mathscr{P}airs_{\mathscr{S}}$.

A semantics σ satisfies the *conflict-sensitiveness* principle if and only if for every argumentation graph $\mathscr{F}$, $\sigma(\mathscr{F})$ is conflict-sensitive.

Let us now turn to com-closure (Dunne et al., 2015). To define this principle, we first need to introduce the notion of completion set. Completion sets are the smallest extensions that contain a given set.

Definition 2.1.23 — Completion set. Given a set of extensions $\mathscr{S} = \{\mathscr{E}_1, \ldots, \mathscr{E}_n\}$ and a set of arguments $\mathscr{E}$, set $\mathscr{E}'$ is a completion set of $\mathscr{E}$ in $\mathscr{S}$ if and only if $\mathscr{E}'$ is a minimal for $\subseteq$ set such that $\mathscr{E}' \in \mathscr{S}$ and $\mathscr{E} \subseteq \mathscr{E}'$.

Roughly speaking, com-closure says that, given a set of extensions $\mathscr{S}$, if for every $\mathscr{T} \subseteq \mathscr{S}$ each two arguments from sets of $\mathscr{T}$ appear in some extension of $\mathscr{S}$, then $\mathscr{T}$ can be extended to an extension in a unique way.

Principle 23 — Com-closure. A set of extensions $\mathscr{S} = \{\mathscr{E}_1, \ldots, \mathscr{E}_n\}$ is com-closed if and only if for every $\mathscr{T} \subseteq \mathscr{S}$ the following holds: if $(a,b) \in \mathscr{P}airs_{\mathscr{S}}$ for each $a, b \in \cup_{\mathscr{E}_i \in \mathscr{T}} \mathscr{E}_i$, then $\cup_{\mathscr{E}_i \in \mathscr{T}} \mathscr{E}_i$ has a unique completion set in $\mathscr{S}$.

A semantics σ satisfies the *com-closure* principle if and only if for every argumentation graph $\mathscr{F}$, $\sigma(\mathscr{F})$ is com-closed.

We drop the definition and the study of the notion of SCC-recursiveness (Baroni et al., 2005) as well as of some properties that are not satisfied by any of the studied semantics. The interested reader is referred to the book cahpter (van der Torre and Vesic, 2018).

In what follows, we consider applying a principle-based approach to ranking-based semantics. We summarize and give perspectives for future work regarding principle-based approach to both extension-based and ranking-based semantics in Subsection 2.2.6.

2.2 The principle-based approach to ranking-based semantics

The previous section was devoted to the principle-based approach to extension-based semantics. In this section we apply the same methodology to ranking-based semantics. Furthermore, in order to generalise the framework, we take into account the initial weights (intrinsic or basic strengths of arguments). This basic strength may come from various sources like certainty degrees of its premises (Benferhat et al., 1993b), votes provided by users (Leite and Martins, 2011), importance degree of a value it promotes (Bench-Capon, 2003), trustworthiness of its source (da Costa Pereira et al., 2011). In all these disparate cases, the basic strength may be expressed by a numerical value, leading to *weighted argumentation graphs*. The question of evaluating the *overall strength* or *overall acceptability* of an argument in such graphs raises naturally.

Amgoud and Ben-Naim (2013) proposed the first set of principles for ranking-based semantics. That set was later extended and refined by the same authors (Amgoud and Ben-Naim, 2016) by decomposing some principles into more elementary ones. Baroni et al. (2019) analyse the principles from the literature (including those that deal with bipolar argumentation frameworks). They analyse the deep motivations behind them, and then define a small number of more abstract principles that are then shown to be their generalisations. Each of their general principles implies several existing principles from the literature. The generalizations make it easier to check whether a group of principles is satisfied by a semantics. Roughly speaking, our work is more concerned with proposing the concrete principles and studying whether they should be satisfied whereas the goal of Baroni et al. (2019) is not primarily to define novel principles but rather to show how to represent existing ones within a generalized framework in order to formally show their underpinnings.

Our work in this area that we present in this section (Amgoud et al., 2017a) consists of three main contributions. Our first contribution is to extend the set of principles to account for basic strengths of arguments. We also introduce four novel principles. The second contribution consists of providing the formal analysis and the thorough comparison of the semantics. These shed light on underpinnings, strengths and weaknesses of each semantics, as well as similarities and differences between pairs of semantics. The results also reveal the limitations in the literature, namely there are no semantics that satisfy certain desirable principles. Our third contribution consists of filling the previous gaps by introducing three novel semantics, each of them satisfying different principles.

Some principles being incompatible, we need several semantics, each of them satisfying different groups of principles.

In Section 2.1, we first presented the definitions of well-known existing extension-based semantics, since they represent the state of the art in computational argumentation. Later, we introduced the principles, which were defined after the semantics. In this section, we start by introducing our principles. We later present three novel semantics in order to illustrate the principles.

To take into account *initial weights* of arguments, we add the third component, which is a function w attaching to each argument a value in $[0,1]$. A weighted argumentation graph is then defined as follows.

Definition 2.2.1 — WAG. A *weighted argumentation graph* (WAG) is a tuple $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, where $\mathcal{A}$ is a finite set, w is a function from $\mathcal{A}$ to $[0,1]$, and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$.

As opposed to an extension-based semantics, which returns the set of extensions, a ranking-based semantics returns a ranking on arguments. The ranking-based approach follows the idea that arguments can be evaluated individually. This approach also allows for more fine graded evaluation, as opposed to the "classical" three value status: skeptically accepted, credulously accepted and rejected, which is used to evaluate individual arguments by using an extension-based semantics. Some ranking-based semantics define the ranking on the arguments directly. Some other (indeed, the majority of them) attach a score to each argument; often, but not always, this score is in the interval $[0,1]$. Those scores then naturally induce a ranking on the arguments. The value assigned to an argument by a ranking-based semantics is called *acceptability degree*. This value represents the *overall strength* of an argument, and is issued from the *aggregation* of the basic strength of the argument and the overall strengths of its attackers. The greater this value, the more acceptable the argument. Since in this section we consider only the semantics that attach scores to arguments, we provide the corresponding definition.

Definition 2.2.2 — Weighted ranking-based semantics. A *weighted ranking-based semantics* is a function σ such that for every WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, for every $a \in \mathcal{A}$, $\sigma_{\mathcal{G}}(a) \in [0,1]$. Number $\sigma_{\mathcal{G}}(a)$ is called *acceptability degree* of a.

Let us comment on the choice of the interval $[0,1]$, which we use both for the initial weights and the acceptability degrees. The weakest possible argument is the one having the value 0, it is "worthless" and, intuitively, does not impact the other arguments even if it attacks or supports them[4]. The degree 1 represents a "perfect" argument, e.g. the argument that has the maximal initial value and is not attacked might obtain this acceptability degree. Some scholars discuss the use of the scale $[-1,1]$, or the set of all real numbers. The main argument in favor of this choice is to have a neutral value (e.g. 0). They justify the need for the neutral element since, for instance, an argument on a social network might have neither positive nor negative votes, so it might be natural to put its initial weight to 0. Another feature they advocate is the possibility to assign negative votes to very bad arguments. Mossakowski and Neuhaus (2018) define a semantics such that a support of a worthless argument acts like an attack of a strong argument. According to them: "if an argument a is supported by an argument b with the acceptability degree of -1, then this support has the same effect as if b would attack a with an acceptability degree of $+1$". Vice, versa, in their semantics, "an attack by an unacceptable argument will strengthen the argument that is attacked". We do not find this property desirable, since, the worse an argument is, less it should influence our opinion. This should hold for rational agents, and that is what we want to model. This is why we use the interval $[0,1]$, where the score 0 is attached to a worthless argument, i.e. an argument that should not influence the acceptability status of the other arguments.

[4]Note that in this habilitation thesis we do not study the supports between arguments.

2.2.1 Principles

This subsection presents the principles for ranking-based semantics. We propose 15 principles, which describe the role and the impact of attacks and basic strengths in the evaluation of arguments, and how these two elements are aggregated. We start by introducing the notations.

Notations: Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG and $a \in \mathcal{A}$. $\mathtt{Att}_{\mathcal{G}}(a)$ denotes the set of all *attackers* of a in $\mathcal{G}$ (i.e. $\mathtt{Att}_{\mathcal{G}}(a) = \{b \in \mathcal{A} \mid b\mathcal{R}a\}$). For $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ and $\mathcal{G}' = \langle \mathcal{A}', w', \mathcal{R}' \rangle$ such that $\mathcal{A} \cap \mathcal{A}' = \emptyset$, we define $\mathcal{G} \oplus \mathcal{G}' = \langle \mathcal{A} \cup \mathcal{A}', w'', \mathcal{R} \cup \mathcal{R}' \rangle$, where w'' is such that for every $x \in \mathcal{A}$ (resp. $x \in \mathcal{A}'$), $w''(x) = w(x)$ (resp. $w''(x) = w'(x)$).

The first principle, called *anonymity*, can be found in different domains, e.g. in game theory (Shapley, 1953). In the argumentation literature, anonymity is sometimes called abstraction (Amgoud and Ben-Naim, 2013; Bonzon et al., 2016a) or language independence (Baroni and Giacomin, 2007). We first have to define an isomorphism between WAGs.

Definition 2.2.3 — Isomorphism. Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ and $\mathcal{G}' = \langle \mathcal{A}', w', \mathcal{R}' \rangle$ be two WAGs. An *isomorphism* from $\mathcal{G}$ to $\mathcal{G}'$ is a bijective function f from $\mathcal{A}$ to $\mathcal{A}'$ such that:
- $\forall a \in \mathcal{A}, w(a) = w'(f(a))$
- $\forall a, b \in \mathcal{A}, a\mathcal{R}b$ iff $f(a)\mathcal{R}'f(b)$.

We can now define the principle itself.

Principle 24 — Anonymity. A semantics σ satisfies *anonymity* iff, for any two WAGs $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ and $\mathcal{G}' = \langle \mathcal{A}', w', \mathcal{R}' \rangle$, for any isomorphism f from $\mathcal{G}$ to $\mathcal{G}'$, $\forall a \in \mathcal{A}$, $\sigma_{\mathcal{G}}(a) = \sigma_{\mathcal{G}'}(f(a))$.

Anonymity could be violated by an approach where the name of the arguments, and not only the structure of the attack graph, would be taken into account during the evaluation process. The second principle, called *independence*, states that the acceptability degree of an argument should be independent of any argument that is not connected to it.

Principle 25 — Independence. A semantics σ satisfies *independence* iff, for any two WAGs $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ and $\mathcal{G}' = \langle \mathcal{A}', w', \mathcal{R}' \rangle$ such that $\mathcal{A} \cap \mathcal{A}' = \emptyset$, the following holds: $\forall a \in \mathcal{A}$, $\sigma_{\mathcal{G}}(a) = \sigma_{\mathcal{G} \oplus \mathcal{G}'}(a)$.

Independence could be violated by a semantics that takes into account the graph as a whole, and not only the connected component in question. This reminds remotely of the stable semantics. The next principle states that the acceptability degree of an argument a in a graph can depend on argument b only if there is a *path* from b to a, where a path is a finite non-empty sequence $\langle x_1, \ldots, x_n \rangle$ s.t. $x_1 = b$, $x_n = a$ and $\forall i < n$, $x_i \mathcal{R} x_{i+1}$. Note that this principle is more general than the circumscription (Amgoud and Ben-Naim, 2016) even when the arguments have the same basic strengths.

Principle 26 — Directionality. A semantics σ satisfies *directionality* iff, for any two WAGs $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\mathcal{G}' = \langle \mathcal{A}, w, \mathcal{R}' \rangle$ s.t. $\mathcal{R}' = \mathcal{R} \cup \{(a,b)\}$, it holds that: $\forall x \in \mathcal{A}$, if there is no path from b to x, then $\sigma_{\mathcal{G}}(x) = \sigma_{\mathcal{G}'}(x)$.

The next principle, called *neutrality*, states that an argument whose acceptability degree is 0 has no impact on the arguments it attacks.

Principle 27 — Neutrality. A semantics σ satisfies *neutrality* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a, b \in \mathcal{A}$, if
- $w(a) = w(b)$
- $\mathtt{Att}_{\mathcal{G}}(b) = \mathtt{Att}_{\mathcal{G}}(a) \cup \{x\}$ with $x \in \mathcal{A} \setminus \mathtt{Att}_{\mathcal{G}}(a)$ and $\sigma_{\mathcal{G}}(x) = 0$

then $\sigma_{\mathcal{G}}(a) = \sigma_{\mathcal{G}}(b)$.

Equivalence principle ensures that the overall strength of an argument depends only on the basic strength of the argument and the overall strengths of its attackers.

Principle 28 — Equivalence. A semantics σ satisfies *equivalence* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a, b \in \mathcal{A}$, if

- $w(a) = w(b)$
- there exists a bijective function $f : \mathtt{Att}_{\mathcal{G}}(a) \to \mathtt{Att}_{\mathcal{G}}(b)$ s.t. $\forall x \in \mathtt{Att}_{\mathcal{G}}(a)$, $\sigma_{\mathcal{G}}(x) = \sigma_{\mathcal{G}}(f(x))$

then $\sigma_{\mathcal{G}}(a) = \sigma_{\mathcal{G}}(b)$.

Equivalence requires that all the information that impacts the acceptability degree of an argument is expressed by the acceptability degrees of its direct attackers. Maximality principle states that the degree of an unattacked argument is equal to its basic strength.

Principle 29 — Maximality. A semantics σ satisfies *maximality* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a \in \mathcal{A}$, if $\mathtt{Att}_{\mathcal{G}}(a) = \emptyset$, then $\sigma_{\mathcal{G}}(a) = w(a)$.

The role of attacks is to *weaken* their targets. Indeed, when an argument receives an attack, its overall strength decreases. The only exception is when the attacker's acceptability degree is zero.

Principle 30 — Weakening. A semantics σ satisfies *weakening* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a \in \mathcal{A}$, if

- $w(a) > 0$
- $\exists b \in \mathtt{Att}_{\mathcal{G}}(a)$ s.t. $\sigma_{\mathcal{G}}(b) > 0$

then $\sigma_{\mathcal{G}}(a) < w(a)$.

The counting principle states that each non zero degree attacker has an impact on the overall strength of the argument. Thus, the more numerous the attackers of an argument, the weaker the argument. Even some reasonable semantics may violate this principle, namely those that look for particular attackers (e.g. the strongest ones). It is however very useful for formal comparisons of semantics.

Principle 31 — Counting. A semantics σ satisfies *counting* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a, b \in \mathcal{A}$, if

- $w(a) = w(b)$
- $\sigma_{\mathcal{G}}(a) > 0$
- $\mathtt{Att}_{\mathcal{G}}(b) = \mathtt{Att}_{\mathcal{G}}(a) \cup \{y\}$ with $y \in \mathcal{A} \setminus \mathtt{Att}_{\mathcal{G}}(a)$ and $\sigma_{\mathcal{G}}(y) > 0$

then $\sigma_{\mathcal{G}}(a) > \sigma_{\mathcal{G}}(b)$.

Weakening soundness principle goes further than weakening by stating that attacks are the *only* source of strength loss.

Principle 32 — Weakening Soundness. A semantics σ satisfies *weakening soundness* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a \in \mathcal{A}$ s.t. $w(a) > 0$, if $\sigma_{\mathcal{G}}(a) < w(a)$, then $\exists b \in \mathtt{Att}_{\mathcal{G}}(a)$ s.t. $\sigma_{\mathcal{G}}(b) > 0$.

The next two principles are about the *intensity* of an attack. Intensity depends on the strength of the source of the attack as well as that of the target. *Reinforcement* principle states that the stronger the source of an attack, the greater its intensity.

Principle 33 — Reinforcement. A semantics σ satisfies *reinforcement* iff, for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a, b \in \mathcal{A}$, if

- $w(a) = w(b)$,
- $\sigma_{\mathcal{G}}(a) > 0$ or $\sigma_{\mathcal{G}}(b) > 0$
- $\mathtt{Att}_{\mathcal{G}}(a) \setminus \mathtt{Att}_{\mathcal{G}}(b) = \{x\}$
- $\mathtt{Att}_{\mathcal{G}}(b) \setminus \mathtt{Att}_{\mathcal{G}}(a) = \{y\}$
- $\sigma_{\mathcal{G}}(y) > \sigma_{\mathcal{G}}(x)$

then $\sigma_{\mathcal{G}}(a) > \sigma_{\mathcal{G}}(b)$.

From now on, when graphically representing the argumentation graphs, the numbers next to the arguments represent their initial weights.

■ **Example 2.1** Consider the WAG $\mathcal{G}$ depicted in Figure 2.1. Suppose $\sigma_{\mathcal{G}}$ satisfies maximality, so $\sigma_{\mathcal{G}}(a) = 0.5$ and $\sigma_{\mathcal{G}}(c) = 0.9$. If $\sigma_{\mathcal{G}}$ satisfies reinforcement, $\sigma_{\mathcal{G}}(b) > \sigma_{\mathcal{G}}(d)$.

Figure 2.1: If $\sigma_{\mathcal{G}}$ satisfies maximality and reinforcement, $\sigma_{\mathcal{G}}(b) > \sigma_{\mathcal{G}}(d)$. (The numbers represent initial weights of arguments.)

■

Resilience states that an attack cannot completely destroy an argument. In case of non-weighted argumentation graphs, it is argued by Amgoud and Ben-Naim (2016) that resilience is one of the main principles which separates Dung's semantics and existing graded semantics proposed by Amgoud and Ben-Naim (2013) and Besnard and Hunter (2001). Resilience is not a mandatory principle. There might be situations where one wants to allow downgrading arguments' strengths to zero. On the contrary, in debates deprived of formal rules, like those on societal issues (e.g. capital punishment, life without parole), an argument cannot use a formal rule to completely destroy another argument, which is perhaps why it is difficult (impossible?) to reach a consensus on those issues.

Principle 34 — Resilience. A semantics σ satisfies *resilience* iff, for any WAG $\mathcal{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$, $\forall a \in \mathscr{A}$, if $w(a) > 0$, then $\sigma_{\mathcal{G}}(a) > 0$.

Proportionality states that the stronger the target of an attack, the weaker its intensity.

Principle 35 — Proportionality. A semantics $\sigma_{\mathcal{G}}$ satisfies *proportionality* iff, for any WAG $\mathcal{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$, $\forall a, b \in \mathscr{A}$ s.t.

- $\mathtt{Att}_{\mathcal{G}}(a) = \mathtt{Att}_{\mathcal{G}}(b)$
- $w(a) > w(b)$
- $\sigma_{\mathcal{G}}(a) > 0$ or $\sigma_{\mathcal{G}}(b) > 0$

then $\sigma_{\mathcal{G}}(a) > \sigma_{\mathcal{G}}(b)$.

■ **Example 2.2** Consider the WAG $\mathcal{G}$ depicted in Figure 2.2. If $\sigma_{\mathcal{G}}$ satisfies resilience, $\sigma_{\mathcal{G}}(b) > 0$. If $\sigma_{\mathcal{G}}$ also satisfies proportionality, we have $\sigma_{\mathcal{G}}(b) > \sigma_{\mathcal{G}}(c)$.

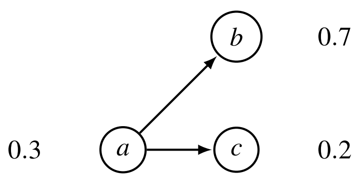

Figure 2.2: If $\sigma_{\mathcal{G}}$ satisfies resilience and proportionality, $\sigma_{\mathcal{G}}(b) > \sigma_{\mathcal{G}}(c)$. (The numbers represent initial weights of arguments.)

■

The three last principles concern possible choices offered to a semantics when it faces a conflict between the quality and the number of attackers as shown by the following example.

■ **Example 2.3** Consider the WAG $\mathscr{G}$ depicted in Figure 2.3, whose arguments are all assigned the same basic strength 1. Argument a has two weak attackers (each attacker is attacked). The argument b has only one attacker, but a strong one. The question is which of a and b is more acceptable?

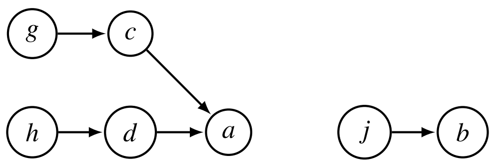

Figure 2.3: Which argument is stronger: a or b? Argument a has several weak attackers whereas b has one strong attacker.

■

The answer to the previous question depends on which of quantity and quality is more important. *Cardinality precedence* principle states that several attackers have more effect on an argument than just few of them.

Principle 36 — Cardinality precedence. A semantics σ satisfies *cardinality precedence* (CP) iff, for any WAG $\mathscr{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$, $\forall a, b \in \mathscr{A}$, if

- $w(a) = w(b)$
- $\sigma_{\mathscr{G}}(b) > 0$
- $|\{x \in \mathtt{Att}_{\mathscr{G}}(a) \mid \sigma_{\mathscr{G}}(x) > 0\}| > |\{y \in \mathtt{Att}_{\mathscr{G}}(b) \mid \sigma_{\mathscr{G}}(y) > 0\}|$

then $\sigma_{\mathscr{G}}(a) < \sigma_{\mathscr{G}}(b)$.

Quality precedence gives more importance to the quality. It is important in e.g. debates requiring expertise. If a Fields medalist says P, whilst three students say $\neg P$, one probably believes P.

Principle 37 — Quality precedence. A semantics σ satisfies *quality precedence* (QP) iff, for any WAG $\mathscr{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$, $\forall a, b \in \mathscr{A}$, if

- $w(a) = w(b)$
- $\sigma_{\mathscr{G}}(a) > 0$
- $\exists y \in \mathtt{Att}_{\mathscr{G}}(b)$ s.t. $\forall x \in \mathtt{Att}_{\mathscr{G}}(a)$, $\sigma_{\mathscr{G}}(y) > \sigma_{\mathscr{G}}(x)$

then $\sigma_{\mathscr{G}}(a) > \sigma_{\mathscr{G}}(b)$.

Compensation says that several weak attacks may compensate the quality. In Figure 2.3, the two attackers of a may compensate the strong attacker of b, thus a might be as acceptable as b.

Principle 38 — Compensation. A semantics σ satisfies *compensation* iff, there exists a WAG $\mathscr{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$ such that for two arguments $a, b \in \mathscr{A}$,

- $w(a) = w(b)$,
- $\sigma_{\mathscr{G}}(a) > 0$,
- $|\{x \in \mathtt{Att}_{\mathscr{G}}(a) \mid \sigma_{\mathscr{G}}(x) > 0\}| > |\{y \in \mathtt{Att}_{\mathscr{G}}(b) \mid \sigma_{\mathscr{G}}(y) > 0\}|$
- $\exists y \in \mathtt{Att}_{\mathscr{G}}(b)$ s.t. $\forall x \in \mathtt{Att}_{\mathscr{G}}(a)$, $\sigma_{\mathscr{G}}(y) > \sigma_{\mathscr{G}}(x)$

and $\sigma_{\mathscr{G}}(a) = \sigma_{\mathscr{G}}(b)$.

Note how weak the compensation principle is. Namely, it is sufficient to find one argumentation graph that satisfies the given conditions in order to satisfy this principle.

2.2.2 Links between the principles

Some of the principles are incompatible; some groups of principles imply other principles etc. In this subsection, we study those links, without restraining us to any particular semantics.

Intuitively, Cardinality precedence and quality precedence represent opposed ideas. We can formally show that they are incompatible in presence of maximality and resilience.

Proposition 2.2.1 If a semantics satisfies maximality and resilience then CP, QP and compensation are pairwise incompatible (i.e. the semantics can satisfy only one of them).

Another principle that is also opposed to quality precedence is counting. Indeed, no semantics can satisfy both of them in presence of several intuitive principles, as shown below.

Proposition 2.2.2 Counting, quality precedence, independence, directionality, equivalence, resilience, reinforcement, maximality and weakening are incompatible (i.e. no semantics can satisfy all of them).

The following result shows some dependencies between the principles. Namely, weakening, weakening soundness, and counting follow from other principles.

Proposition 2.2.3 Let σ be a semantics which satisfies directionality, independence, maximality and neutrality. Then:

- σ satisfies weakening soundness;
- if σ satisfies reinforcement, then it also satisfies counting and weakening.

Arguments that are attacked only by the zero degree attackers keep their basic strength in case the semantics satisfies independence, directionality, neutrality and maximality.

Theorem 2.2.4 Let σ satisfy independence, directionality, neutrality and maximality. For any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, for any $a \in \mathcal{A}$, if for every $x \in \mathtt{Att}_{\mathcal{G}}(a)$ we have $\sigma_{\mathcal{G}}(x) = 0$, then

$$\sigma_{\mathcal{G}}(a) = w(a).$$

A semantics satisfying independence, directionality, proportionality, neutrality, weakening and maximality assigns to each argument a degree between 0 and its basic strength.

Theorem 2.2.5 If a semantics σ satisfies independence, directionality, neutrality, proportionality, weakening and maximality, then for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, for any argument $a \in \mathcal{A}$, we have

$$\sigma_{\mathcal{G}}(a) \in [0, w(a)].$$

Another property which follows from our principles is counter transitivity. It was introduced by Amgoud and Ben-Naim (2013) for ranking semantics in case of non-weighted graphs. It states that if the attackers of an argument b are at least as numerous and strong as those of an argument a, then a is at least as strong as b.

Theorem 2.2.6 If a semantics σ satisfies independence, directionality, equivalence, reinforcement, maximality and neutrality then for any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, $\forall a, b \in \mathcal{A}$, if

- $w(a) = w(b)$
- there exists an injective function f from $\mathtt{Att}_{\mathcal{G}}(a)$ to $\mathtt{Att}_{\mathcal{G}}(b)$ such that $\forall x \in \mathtt{Att}_{\mathcal{G}}(a)$, $\sigma_{\mathcal{G}}(x) \leq \sigma_{\mathcal{G}}(f(x))$

then $\sigma_{\mathcal{G}}(a) \geq \sigma_{\mathcal{G}}(b)$.

2.2.3 Three novel semantics

This section presents three novel semantics, each of them satisfying a different principle (compensation, quality precedence, cardinality precedence). We first present the weighted h-categorizer.

This semantics extends h-categorizer, initially proposed by Besnard and Hunter (2001) for *non-weighted* and *acyclic* graphs. It was shown by Pu et al. (2014) that the original definition can be used for all graphs. Our generalization consists in allowing for the weights to be taken into account. We also show that the semantics we define satisfies a maximal number of principles from Subsection 2.2.1.

The semantics follows a multiple steps process. In the initial step, it assigns to every argument its basic strength. Then, in each step, all the scores are simultaneously recomputed on the basis of the attackers' scores in the previous step.

Definition 2.2.4 — $\mathtt{f_h}$. Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG. We define $\mathtt{f}_{h,\mathcal{G}}: \mathcal{A} \to [0, +\infty)$ as follows: for any argument $a \in \mathcal{A}$, $\mathtt{f}^0_{h,\mathcal{G}}(a) = w(a)$ and for $i \in \{1, 2, \ldots\}$,

$$\mathtt{f}^i_{h,\mathcal{G}}(a) = \frac{w(a)}{1 + \sum_{b_i \in \mathtt{Att}_\mathcal{G}(a)} \mathtt{f}^{i-1}_{h,\mathcal{G}}(b_i)}$$

By convention, if $\mathtt{Att}_\mathcal{G}(a) = \emptyset$, $\sum_{b_i \in \mathtt{Att}_\mathcal{G}(a)} \mathtt{f}^{i-1}_{h,\mathcal{G}}(b_i) = 0$.

Note the idea behind this function: to generalize h-categorizer by multiplying the value with the initial weight of an argument. We can show that the function $\mathtt{f}^i_{h,\mathcal{G}}$ converges.

Theorem 2.2.7 For every weighted argumentation graph, the function $\mathtt{f}^i_{h,\mathcal{G}}$ converges as i approaches infinity.

The acceptability degree of each argument in a weighted graph can thus be defined as the limit reached using the function $\mathtt{f}^i_{h,\mathcal{G}}$.

Definition 2.2.5 — $\mathtt{Hbs}$. The weighted h-categorizer semantics $\mathtt{Hbs}$ is defined as follows: for every WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, for every $a \in \mathcal{A}$,

$$\mathtt{Hbs}(a) = \lim_{i \to +\infty} \mathtt{f}^i_{h,\mathcal{G}}(a).$$

Note that from now on, we drop the graph $\mathcal{G}$ from the index of the semantics when there is no danger of confusion. We thus write $\mathtt{Hbs}$ instead of $\mathtt{Hbs}_\mathcal{G}$. We now present a characterization of weighted h-categorizer semantics.

Theorem 2.2.8 Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG, and let $D : \mathcal{A} \to [0, +\infty)$. If for every $a \in \mathcal{A}$,

$$D(a) = \frac{w(a)}{1 + \sum_{b \in \mathtt{Att}_\mathcal{G}(a)} D(b)},$$

then $D \equiv \mathtt{Hbs}$.

Finally, we show that this semantics satisfies compensation as well as all the principles that are compatible with it.

Theorem 2.2.9 Weighted h-categorizer semantics satisfies all the principles from Subsection 2.2.1 except CP and QP.

We can show that the degree of an argument is never greater than its weight.

Proposition 2.2.10 Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG. For any $a \in \mathcal{A}$, $\mathtt{Hbs}(a) \in [0, w(a)]$.

Weighted h-categorizer satisfies compensation but not CP and QP.

Let us now introduce weighted max-based semantics, which satisfies QP.

Definition 2.2.6 — $\mathtt{f_m}$. Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG. We define $\mathtt{f}_{\mathtt{m},\mathcal{G}}$:$\mathcal{A} \to [0, +\infty)$ as follows: for any argument $a \in \mathcal{A}$, $\mathtt{f}^0_{m,\mathcal{G}}(a) = w(a)$ and for $i \in \{1, 2, \ldots\}$,

$$\mathtt{f}^i_{m,\mathcal{G}}(a) = \frac{w(a)}{1 + \max_{b \in \mathtt{Att}_{\mathcal{G}}(a)} \mathtt{f}^{\mathtt{i}-1}_{\mathtt{m}}(b)}$$

By convention, if $\mathtt{Att}_{\mathcal{G}}(a) = \emptyset$, $\max_{b \in \mathtt{Att}_{\mathcal{G}}(a)} \mathtt{f}^{\mathtt{i}}_{\mathtt{m}}(b) = 0$

The idea is to take into account the initial weight of an argument as well as the strength of the strongest attacker. The function $\mathtt{f}^i_{m,\mathcal{G}}$ converges.

Theorem 2.2.11 For every weighted argumentation graph, the function $\mathtt{f}^i_{m,\mathcal{G}}$ converges as i approaches infinity.

Like in the case of $\mathtt{f}^i_{h,\mathcal{G}}$, the acceptability degree of each argument in a weighted graph can thus be defined as the limit reached using the function $\mathtt{f}^i_{m,\mathcal{G}}$.

Definition 2.2.7 — $\mathtt{Mbs}$. The weighted max-based semantics $\mathtt{Mbs}$ is defined as follows: for every WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, for every $a \in \mathcal{A}$,

$$\mathtt{Mbs}(a) = \lim_{i \to +\infty} \mathtt{f}^i_{m,\mathcal{G}}(a).$$

As in the case of weighted h-categorizer, we have the following characterization.

Theorem 2.2.12 Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG, and let $D : \mathcal{A} \to [0, +\infty)$. If for every $a \in \mathcal{A}$,

$$D(a) = \frac{w(a)}{1 + \max_{b \in \mathtt{Att}(a)} D(b)},$$

then $D \equiv \mathtt{Mbs}$.

Theorem 2.2.13 Weighted max-based semantics violates cardinality precedence, compensation, counting and reinforcement. It satisfies all the remaining principles from Subsection 2.2.1.

We show next that an attacked argument cannot lose more than half of its basic strength with this semantics. This is not very surprising since only one attacker has an effect on the argument.

Proposition 2.2.14 For any WAG $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$, for any $a \in \mathcal{A}$, $\mathtt{Mbs}(a) \in \left[\frac{w(a)}{2}, w(a)\right]$.

The last semantics we present satisfies cardinality precedence. It favors the number of arguments over their quality. However, it considers only the arguments with strictly positive degree. This choice is due to the view that the zero degree arguments do not impact the arguments they attack.

Definition 2.2.8 — Founded argument. Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG and $a \in \mathcal{A}$. The argument a is *founded* iff $w(a) > 0$. Let $\mathtt{AttF}_{\mathcal{G}}(a)$ denote the set of founded attackers of a.

Definition 2.2.9 — $\mathtt{f_c}$. Let $\mathcal{G} = \langle \mathcal{A}, w, \mathcal{R} \rangle$ be a WAG. We define $\mathtt{f}_{\mathtt{c},\mathcal{G}}$:$\mathcal{A} \to [0, +\infty)$ as follows:

for any argument $a \in \mathscr{A}$, $\mathtt{f}^0_{c,\mathscr{G}}(a) = w(a)$ and for $i \in \{1,2,\ldots\}$,

$$\mathtt{f}^i_{c,\mathscr{G}}(a) = \frac{w(a)}{1 + |\mathtt{AttF}_{\mathscr{G}}(a)| + \frac{\sum_{b \in \mathtt{AttF}_{\mathscr{G}}(a)} \mathtt{f}_{\mathtt{c}}^{\mathtt{i-1}}(b)}{|\mathtt{AttF}_{\mathscr{G}}(a)|}}$$

By convention, if $\mathtt{Att}_{\mathscr{G}}(a) = \emptyset$, $\max_{b \in \mathtt{Att}_{\mathscr{G}}(a)} \mathtt{f}_{\mathtt{c}}^{\mathtt{i}}(b) = 0$

The idea behind the previous function is to take into account the initial weight of an argument as well as the number of founded attackers. The expression $\frac{\sum_{b \in \mathtt{AttF}_{\mathscr{G}}(a)} \mathtt{f}_{\mathtt{c}}^{\mathtt{i-1}}(b)}{|\mathtt{AttF}_{\mathscr{G}}(a)|}$ is added in order to be able to take into account the scores of attackers in case two arguments have the same number of attackers.

Theorem 2.2.15 For every weighted argumentation graph, the function $\mathtt{f}^i_{c,\mathscr{G}}$ converges as i approaches infinity.

Definition 2.2.10 — Cbs. The weighted card-based semantics Cbs is defined as follows: for every WAG $\mathscr{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$, for every $a \in \mathscr{A}$,

$$\mathtt{Cbs}(a) = \lim_{i \to +\infty} \mathtt{f}^i_{c,\mathscr{G}}(a).$$

Theorem 2.2.16 Let $\mathscr{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$ be a WAG, and let $D : \mathscr{A} \to [0, +\infty)$. If for every $a \in \mathscr{A}$,

$$D(a) = \frac{w(a)}{1 + |\mathtt{AttF}_{\mathscr{G}}(a)| + \frac{\sum_{b \in \mathtt{AttF}_{\mathscr{G}}(a)} D(b)}{|\mathtt{AttF}_{\mathscr{G}}(a)|}}$$

then $D \equiv \mathtt{Cbs}$.

Theorem 2.2.17 Weighted card-based semantics satisfies all the principles except quality precedence and compensation.

The acceptability degree of an argument is never greater than its initial weight.

Proposition 2.2.18 For any WAG $\mathscr{G} = \langle \mathscr{A}, w, \mathscr{R} \rangle$, for any $a \in \mathscr{A}$, $\mathtt{Cbs}(a) \in [0, w(a)]$.

We implemented the three semantics from this subsection and conducted numerous experiments. The degrees can be calculated very fast (a couple of seconds even for complex, large graphs, consisting of hundreds of arguments).

2.2.4 Formal analysis of existing semantics

What about the other ranking-based semantics? Do they satisfy the postulates? Are our postulates only suited for ranking-based semantics? What is the essential difference between ranking-based and extension-based semantics? We now study those questions.

It is straightforward to check whether existing ranking-based semantics satisfy our postulates, and we do it, as expected. Furthermore, we extend our analysis to extension-based semantics in order to check whether they satisfy the postulates, too. Recall that in all approaches for extension-based semantics that model different argument strengths (Amgoud and Cayrol, 2002a; Bench-Capon, 2003; Modgil, 2009), the idea is to ignore the attacks whose source is weaker than the target before computing the extensions and then apply a given extension-based semantics. Thus, we do the same thing:

1. Given a WAG $\mathscr{G}$, we say that a is preferred to b iff $w(a) \geq w(b)$. Then, a new graph

	Ground.	Stable	Pref.	Compl.	IS	(DF-)QuAD	TB	Mbs	Cbs	Hbs
Anonymity	•	•	•	•	•	•	•	•	•	•
Independence	•	×	•	•	•	•	•	•	•	•
Directionality	•	×	•	•	•	•	•	•	•	•
Neutrality	•	•	×	×	•	•	•	•	•	•
Equivalence	×	×	×	×	•	•	•	•	•	•
Maximality	×	×	×	×	×	•	•	•	•	•
Weakening	×	×	×	×	×	•	×	•	•	•
Counting	×	×	×	×	×	•	×	×	•	•
Weakening soundness	•	×	×	×	•	•	•	•	•	•
Proportionality	×	×	×	×	×	•	×	•	•	•
Reinforcement	×	×	×	×	×	•	×	×	•	•
Resilience	×	×	×	×	×	×	×	•	•	•
Cardinality Precedence	×	×	×	×	×	×	×	×	•	×
Quality Precedence	×	×	×	×	•	×	×	•	×	×
Compensation	•	•	•	•	×	•	•	×	×	•

Table 2.5: The symbol • (resp. $\times$) stands for the principle is satisfied (resp. violated) by the semantics.

$\mathscr{F}' = \langle \mathscr{A}, \mathscr{R}' \rangle$ is generated, where $a\mathscr{R}'b$ iff a attacks b and a attacks b, and b is not strictly preferred to a. Formally $\mathscr{R}' = \{(a,b) \mid a\mathscr{R}b \text{ and } w(b) \leq w(a)\}$.

2. Given extension-based semantics is applied to $\mathscr{F}'$.
3. The previous procedure allows us to attach acceptability degrees to arguments of a given weighted graph using an extension-based semantics. An acceptability degree is then assigned to each argument as proposed in the literature (Amgoud and Ben-Naim, 2016). Namely, an argument gets acceptability degree 1 iff it belongs to all extensions; value 0.5 if it belongs to some but not all extensions; value 0.3 if it does not belong to and is not attacked by any extension; value 0 if it does not belong to any extension and is attacked by an extension.

Table 2.5 shows the results for grounded, stable, preferred and complete semantics enriched with preferences. It is not surprising that many postulates are not satisfied by those semantics. Of course, one of the reasons for this is that extension-based semantics were not defined with our postulates in mind, and, of course, not satisfying some of the postulates does not mean that a semantics is useless, corrupted or anything similar. However, we think checking whether existing extension-based semantics satisfy our postulates provides us with some interesting insights in the underpinnings of argumentation semantics and conceptual differences between extension-based and ranking-based semantics. Note that the only two postulates satisfied by all four Dung's semantics are anonymity (which is not surprising) and compensation (which is, indeed, very easy to satisfy, since it is sufficient to find one graph where the corresponding degrees coincide).

To illustrate, observe that weakening is not satisfied by the four semantics. This is because those postulates are based on the hypothesis that, roughly speaking, each argument is evaluated "individually" and that the influences of all the arguments on a given argument x are "summarized" in the degree of x. Thus, one can determine the degree of argument x only by knowing the degrees of its direct attackers. This hypothesis is not a part of the intuition behind extension-based semantics, where knowing the acceptance status (skeptical, credulous, rejected) of direct attackers is not enough to determine the status of a given argument.

We now study whether existing ranking-based semantics satisfy the postulates. Gabbay and Rodrigues (2015) developed "Iterative Schema" (IS). Basic strengths are used as initial *labels* of arguments. Value 1 corresponds to label `in`, value 0 to `out`, and any other value to `und`. If this labeling is legal (i.e. if it satisfies some conditions defined by the authors), IS returns a single extension, consisting of the arguments having the value 1. Otherwise, it modifies the values using an iterative schema until reaching a legal labeling. Note, however, that IS is not a "real" ranking-based semantics, since the set of resulting scores (i.e. acceptability degrees) is $\{0, 0.5, 1\}$, which means that an argument can have only three different degrees. Baroni et al. (2015) developed QuAD, which assumes that arguments can both be attacked and supported. QuAD was extended to DF-QuAD by Rago et al. (2016). The two versions coincide when the support relation is empty. They both consider only *acyclic* graphs. To analyze them against our principles, we assume acyclic graphs and empty support relations. We also analyze TB (da Costa Pereira et al., 2011), where basic strengths express degrees of trustworthiness of arguments' sources. We do not study the framework of Leite and Martins (2011) since that system is not guaranteed to assign a unique numerical value to each argument Amgoud et al. (2017c).

Let us comment on some interesting findings. Note that IS is the only semantics apart from our Mbs that satisfies QP principle. However, IS violates Maximality. Thus, if we are given a graph with two arguments a and b, not attacking each other and having initial weights 0.01 and 1, IS returns a single extension, $\{a, b\}$, declaring thus both arguments "perfectly" acceptable. This is related with the fact that IS assigns only three different statuses to the arguments.

(DF-)QuAD satisfies the same properties as Hbs except Resilience. Indeed, Hbs does not allow an argument to lose all its basic weight (i.e. if its weight is strictly positive, its degree is strictly positive). On the contrary, QuAD sets an argument's degree to zero if it is attacked by a "perfectly" acceptable argument. Observe that the two semantics are different for many other reasons, e.g. QuAD can treat the supports but is defined only for acyclic graphs.

TB satisfies several postulates but violates *weakening*. Indeed, an argument may not lose weight even if it has a strong attacker. This is a consequence of the modeling choice of this semantics, where the degree of the argument never greater than its initial weight. Thus, an argument having initial weight 0.1 will have the degree 0.1 both in the case when it is not attacked at all and in the case when it is attacked by ten arguments, each of them having the degree 0.9.

Observe that Cbs is the first semantics in the literature that satisfies CP. Even if several semantics satisfy compensation, Hbs is the only one that satisfies all the principles compatible with compensation, and that for any graph structure.

Let us mention the paper by Bonzon et al. (2016a), where the authors study the semantics and principles for graphs without initial weights.

2.2.5 Compensation-based semantics

One of the key ideas behind the ranking-based semantics is that an attack does not completely destroy its target. Instead, the attacked argument is weakened to a certain degree. The principles states that

- more attackers there are, lower the acceptability of the attacked argument is
- stronger the attackers are, lower the acceptability of the attacked argument is.

However, nothing is said about the link between the number and the strength of the attackers. This section studies the question: can a large number of weak attackers have the same effect as the small number of strong attackers?

In order to simplify the framework and more easily concentrate on the question of compensation, we suppose that all the arguments have the same initial weight. The case when we are also given the function w is left for future work.

Observe now the example from Figure 2.4. The argument a is attacked by one strong argument, whereas b is attacked by three weak arguments. Very roughly speaking, quantity precedence would imply that a is better than b and quality precedence that b is better than a. However, if a semantics in question satisfies compensation, we do not know how many weak arguments (e.g. r, s, t) have the same effect on the attacked argument as one strong argument (e.g. p)? How can a semantics decide on this?

Note first that each ranking-based semantics *must* decide on this, since it has to produce an

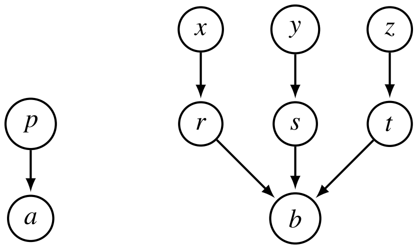

Figure 2.4: Compensation between the strength and the number of attackers

order on the arguments. In this particular example,

- a can be stronger than b, meaning that the semantics, roughly speaking, gives precedence to the *number* of attacks
- b can be stronger than a, meaning that the semantics gives precedence to the *quality* of attacks
- a and b can be equally good, meaning that the attacks of three "weak" arguments *compensate* the attack of one "strong argument".

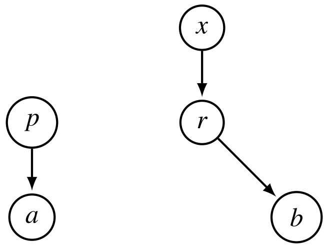

Figure 2.5: Argument b is expected to be stronger than a for most of the semantics

For example, h-categorizer (Besnard and Hunter, 2001) declares a better than b. If we remove the arguments z and t (or just the attack from t to b), arguments a and b will have the same acceptability degree with respect to h-categorizer. This means that this semantics compensates one attack by a strong argument with two attacks from weak arguments (where a strong argument is an unattacked argument and a weak argument is an argument attacked by exactly one strong argument).

If we now go a step further and consider the argumentation graph from Figure 2.5, h-categorizer declares b better than a. Indeed, for most of the ranking-based semantics, one might expect a behavior similar to that of h-categorizer:

- in the situation like in Figure 2.5, b is stronger than a
- if one keeps adding weak attackers of b
  - at some point a will be come as good as b
  - depending on the semantics, a might become strictly better than b.

Let us emphasize that there exist reasonable and useful semantics that do not exhibit this behavior (Bonzon et al., 2016c). One might want b to be strictly better than a no matter how many attackers we add, since it is defended from all its attackers. Such a semantics strictly privileges quality of attackers over their quantity. However, as illustrated by examining h-categorizer, there are ranking-based semantics that allow for some kind of compensation between a small number of strong attackers and a large number of weak attackers.

The initial question that motivates our research reported in the rest of this section is: why does h-categorizer compensate at the degree of exactly *two* weak arguments. More generally, the first goal of our work is to better understand this compensation mechanism by defining a principle that allows to formally examine existing and new semantics with respect to how they deal with compensation. The second goal is to define a parametrized semantics which allows a user to choose the degree of compensation by changing the value of this parameter.

Recall how the condition from the compensation principle (Principle 38, page 30) is easy to satisfy. Namely a semantics satisfies compensation if there is at least one graph where

- an argument a is attacked by more arguments than b
- the strongest attacker of b is stronger than the strongest attacker of a
- arguments a and b are equally strong.

We now introduce another principle, called (n,i)-compensation, which measures to which degree a semantics satisfies compensation. This measure is based on the number of "weak" arguments (n) necessary to compensate one attack from a "strong" argument, where a strong argument is an unattacked argument and a weak argument is an argument attacked by i strong arguments. Let us define $\mathscr{C}^i(\mathscr{F})$ as the set of all arguments that are attacked by exactly i unattacked arguments (and only by those arguments).

Definition 2.2.11 For every argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$,

$$\mathscr{C}^i(\mathscr{F}) = \{a \in \mathscr{A} \text{ such that } |\texttt{Att}(a)| = i \text{ and } \forall b \in \texttt{Att}(a), \texttt{Att}(b) = \emptyset\}.$$

Now, (n,i)-compensation can be defined as follows.

Definition 2.2.12 — (n,i)-compensation. Let $n,i \in \{1,2,3,\ldots\}$. A ranking semantics σ satisfies *compensation* at degree (n,i) iff for every argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, for all $a,b \in \mathscr{A}$, the following holds: if

- $|\texttt{Att}(a)| = n$, $\texttt{Att}(a) \subseteq \mathscr{C}^i(\mathscr{F})$, and
- $|\texttt{Att}(b)| = 1$, $\texttt{Att}(b) \subseteq \mathscr{C}^0(\mathscr{F})$

then $\sigma_{\mathscr{F}}(a) = \sigma_{\mathscr{F}}(b)$.

We now show that h-categorizer satisfies $(k+1,k)$-compensation for every natural number k.

Proposition 2.2.19 *h-categoriser* satisfies (n,i)-compensation if and only $n = i+1$ and $i \geq 0$.

Let us now define a parametric semantics that allows the user to choose the compensation degree.

Definition 2.2.13 — s_α. The compensation-based semantics s_α is defined as follows: for every argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, for every $\alpha \in (0, +\infty)$:

$$s_\alpha(a) = \frac{1}{1 + \left(\sum_{b \in \texttt{Att}(a)} (s_\alpha(b))^\alpha\right)^{1/\alpha}}$$

The previous semantics is a generalization of h-categorizer, inspired by p-norms in L^p spaces.

■ **Example 2.4** Reconsider the argumentation graph from Figure 2.4. There exists $\alpha' \approx 1.585$ such that for $\alpha < \alpha'$, $s_\alpha(a) > s_\alpha(b)$, for $\alpha > \alpha'$, $s_\alpha(a) < s_\alpha(b)$ and for $\alpha = \alpha'$, $s_\alpha(a) = s_\alpha(b)$. ■

For every argumentation graph and for every value of the parameter α, our compensation-based semantics attributes a unique score to each argument.

Theorem 2.2.20 For every argumentation graph $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ and for every $\alpha \in (0, +\infty)$, compensation-based semantics s_α is well defined, i.e. for every argument $a \in \mathscr{A}$, there exists a unique number $s_\alpha(a) \in [0,1]$ such that the system of equations from Definition 2.2.13 (one for each argument) is satisfied.

In order to check compliance with the principles from Section 2.2.1, we suppose that the initial weight of all arguments is equal to 1 and that the semantics s_α is applied on a weighted graph by ignoring the initial weights of arguments. We conclude that our parametrized semantics satisfies all the principles from Subsection 2.2.1 except cardinality precedence and quality precedence[5].

Theorem 2.2.21 For every $\alpha \in (0, +\infty)$, compensation-based semantics s_α satisfies anonymity, independence, directionality, neutrality, equivalence, maximality, weakening, counting, weakening soundness, reinforcement and resilience on the class of graphs where all the arguments have initial weight 1.

We can also show that s_α satisfies (n,i)-compensation for any n and i such that $n > i$.

Theorem 2.2.22 For every $n, i \in \mathbb{N}$ such that $n > i$ there exists a unique $\alpha \in (0, +\infty)$ s.t. s_α satisfies (n,i)-compensation.

Whether and how we can define a semantics that satisfies (n,i)-compensation for all values of n and i, i.e. when $n > i$ and $n \leq i$ is the subject of current and future research.

To calculate the scores with respect to s_α, we can apply an iterative algorithm. It is sufficient to assign the value 1 to all the arguments and then update this value iteratively, by using the formula from Definition 2.2.13. We proved that this algorithm converges and calculates the correct values. We also implemented and tested the algorithm, which calculates the scores of the arguments very quickly.

2.2.6 Summary and outlook

The principle-based approach has developed over the past ten years into a cornerstone of formal argumentation theory, because it allows for a more systematic study and comparison of argumentation semantics. We give an analysis of the main extension-based semantics using the principles discussed in the literature. We also introduced and studied the principles for ranking-based semantics. Furthermore, we introduced three new ranking-based semantics and analyzed them with respect to the principles. There are other works in this area, for example Caminada (2018) discusses the principles used in structured argumentation, which he calls rationality postulates, and Dung (2016) analyses *prioritised* argumentation using a principle-based or axiomatic approach.

It may be expected that the principle-based approach will play an even more prominent role in the future of formal argumentation, as the number of alternatives for argumentation semantics increases, new argumentation principles are introduced, and more requirements of actual applications are expressed in terms of such principles.

Finally, the principle-based approach to formal argumentation may lead to the study of impossibility and possibility results, as well as the development of representation theorems characterizing sets of argumentation semantics. The use of the principle-based approach in other areas of reasoning, such as voting theory or AGM theory change, may inspire such further formal investigations.

The principle-based approach has also been used to provide a more systematic study and analysis of the semantics of extended argumentation frameworks, of the aggregation of argumentation graphs, and of the dynamics of argumentation graphs. Likewise, we expect a further study of input/output frameworks, abstract dialectical frameworks, and so on.

I recently worked on two other principle-based approaches. One consists of defining the principles for semantics where the attack relation is generalized in order to allow for sets of arguments to attack an argument (Yun et al., 2020b). The other defines principles for ranking-based semantics for bipolar argumentation frameworks, where one relation represents attacks between arguments and another represents the necessities between them (Doder et al., 2020). I do not detail those works due to space constraints, the interested reader is referred to the recently appeared conference publications.

[5]Strictly speaking, proportionality is also satisfied, although in a vacuous way.

Note that Bonzon et al. (2016b) also study the principles for ranking-based semantics, however, their work does not take into account the weights of arguments.

Let us now mention the increasing interest in probabilistic argumentation whose aim is handling uncertainty in an argumentation context (Dung and Thang, 2010; Li et al., 2011; Thimm, 2012; Hunter, 2013; Hunter and Thimm, 2017; Polberg and Hunter, 2018; Hunter et al., 2020; Hunter, 2020). Those works look similar to our approach, but the intuitions are quite different. There are many different papers and it is impossible to give a quick comparison. Roughly speaking, our approach is centred around the graph, for example if two parts of graph are not connected, they will not influence each other. In (many of the) probabilistic approaches this is not true. The acceptability of arguments in probabilistic approaches can be influenced by constraints on arguments' probabilities, without necessarily being visible in the argument graph.

2.3 Links between argumentation and other approaches

In this section, we are interested by the links between argument-based reasoning and other approaches for reasoning under inconsistent knowledge. We present some of those links as well as the differences between the two classes of approaches. We also study the links between extension-based and ranking-based semantics. Also, at the very end of this section, we briefly mention some other works in argumentation. However, we do not present them in detail in order to keep this habilitation thesis within a reasonable page limit.

2.3.1 Identifying the class of maxi-consistent operators in argumentation

This subsection shows how to identify the large class of logic-based instantiations of Dung's theory that correspond to the maxi-consistent operator, i.e. to the function which returns maximal consistent subsets of an inconsistent knowledge base. In other words, we study the class of instantiations where every extension of the argumentation system corresponds to exactly one maximal consistent subset of the knowledge base. We show that an attack relation belonging to this class must be conflict-dependent, must not be valid, must not be conflict-complete, must not be symmetric etc. Then, we show some inclusion results, e.g. if an attack relation contains *canonical undercut* then it is not a member of this class. By using our results, we show for all existing attack relations whether or not they belong to this class. We also define *new* attack relations which are members of this class. Finally, we interpret our results and discuss more general questions, like: what is the added value of argumentation in such a setting? We believe that this work can help to understand the role of argumentation and, particularly, the expressivity of logic-based instantiations of Dung-style argumentation frameworks.

In order to keep the manuscript succinct, we present only some of the results. For a throughout discussion about different classes of operators and their role in reasoning under inconsistency, the reader is referred to the journal publication (Vesic, 2013).

The starting point of the research we present in this subsection is the question whether Dung's (1995) abstract theory can be used as a general framework for non-monotonic reasoning. This question has drawn a particular amount of attention among researchers in artificial intelligence. More precisely: can existing or new approaches to reasoning be seen as instantiations of Dung's theory? This is certainly a very general question. Furthermore, different approaches suppose that the available knowledge is represented in different form. We focus on the problem setting when one is given a finite inconsistent set of classical propositional logic formulae, which we refer to as a *knowledge base*. There are a number of approaches for dealing with inconsistent information: a notable example are paraconsistent logics (Priest, 2002) where one is able to draw some (but not all) conclusions from an inconsistent set of formulae. Indeed, each paraconsistent logic allows for a subset of the inferences that could be obtained using classical logic with the same knowledge.

Other examples of dealing with inconsistent knowledge include belief revision (Gardenfors, 1988), belief merging (Konieczny and Pérez, 2011) or voting (Arrow et al., 2002). To be completely precise, note that there are approaches where one is given a *multiset* instead of a set, for example where several voters can express their knowledge or preferences and the number of agents stating / voting for a proposition is important. However, in this study, we suppose that the information is represented in form of a set.

We call an *operator* a function which provides a way to go from an inconsistent knowledge base to a set of subsets of that knowledge base. Examples of operators are: a function returning maximal for set inclusion consistent subsets of a knowledge base, called *maxi-consistent operator*, a function returning maximal for cardinality consistent subsets of a knowledge base, called *maxi-card operator*, a function returning *all* consistent subsets of a knowledge base...

To understand how and to which extent Dung's theory can be used as a general framework for reasoning, it is essential to study the link between the result obtained by applying an operator to a knowledge base Σ and the extensions of the argumentation graph $\mathscr{F} = (\texttt{Arg}(\Sigma), \mathscr{R})$, where for a set $S \subseteq \Sigma$, we denote by $\texttt{Arg}(S)$ the set of all arguments that can be built from S, and by $\mathscr{R}$ the attack relation.

We now aim at giving a broad overview of the class of operators that can be viewed as instantiation of Dung's theory. First note that, interestingly, there is a rather big class of instantiations of Dung's theory returning inconsistent results, as showed by Caminada and Amgoud (2007). However, one would normally prefer to avoid this type of behaviour, and to study the class of instantiations returning consistent results. Thus, our long-term goal is to identify the whole class of instantiations of Dung's theory that yield a consistent result. However, that is certainly a hard task. We start by noticing that, given a set Σ, the most common and a well-known way to deal with inconsistent information is to use the maxi-consistent operator, i.e. to select maximal consistent subsets of Σ. Also, we conjecture that most of the existing instantiations fall into this class. That is why we start by studying this class.

As already mentioned, we suppose that one is given a set of classical propositional logic formulae Σ. We use the well-known (Besnard and Hunter, 2001; Amgoud and Cayrol, 2002b; Gorogiannis and Hunter, 2011) *logic-based* approach for instantiating Dung's theory. $\mathscr{L}$ denotes the set of well-formed formulae, $\vdash$ stands for classical entailment, and $\equiv$ for logical equivalence. We use the notation $\texttt{MC}(\Sigma)$ for the set of all maximal consistent subsets of Σ.

A logical argument is defined as a pair $(support, conclusion)$.

Definition 2.3.1 — Logic-based argument. An argument is a pair (Φ, α) such that $\Phi \subseteq \Sigma$ is a minimal (for set inclusion) consistent set of formulae such that $\Phi \vdash \alpha$.

For an argument $a = (\Phi, \alpha)$, we use the function $\texttt{Supp}(a) = \Phi$ to denote its support and $\texttt{Conc}(a) = \alpha$ to denote its conclusion. For a given set of formulae S, we denote by $\texttt{Arg}(S)$ the set of arguments constructed from S. Formally, $\texttt{Arg}(S) = \{a \mid a \text{ is an argument and } \texttt{Supp}(a) \subseteq S\}$. Let $\texttt{Arg}(\mathscr{L})$ denote the set of all argument can be constructed from the language of propositional logic. For a given set of arguments $\mathscr{E}$, we denote $\texttt{Base}(\mathscr{E}) = \bigcup_{a \in \mathscr{E}} \texttt{Supp}(a)$. We suppose that function $\texttt{Arg}$ is defined on $\mathscr{L}$ and that function $\texttt{Base}$ is defined on $\texttt{Arg}(\mathscr{L})$; we sometimes write $\texttt{Arg}$ (respectively $\texttt{Base}$) for the restriction of these functions on any set of formulae (respectively arguments).

Given a knowledge base Σ, we construct an argumentation system $\mathscr{F} = (\texttt{Arg}(\Sigma), \mathscr{R})$, and then, using a chosen semantics, calculate extensions. Since all the components of the system except a semantics and an attack relation are fixed, then whether an instantiation corresponds to maxi-consistent operator depends exclusively on those two components. The next definition provides a formal definition of what we mean by saying that an instantiation of Dung's framework "corresponds" to maxi-consistent operator. The idea is that the function $\texttt{Arg}$ should be a bijection between the maximal consistent subsets of the knowledge base, denoted $\mathtt{MC}(\Sigma)$, and the extensions of the corresponding argumentation system.

Definition 2.3.2 — MC $\leftrightarrow$ Ext. Let σ be an argumentation semantics. We say that attack relation $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$ if and only if for every finite set of propositional formulae Σ we have that

$$\mathtt{Arg} \text{ is a bijection between } \mathtt{MC}(\Sigma) \text{ and } \sigma((\mathtt{Arg}(\Sigma), \mathscr{R}))$$

We now show the link between the existing properties of attack relations, like conflict-dependence (Amgoud and Besnard, 2009) and validity (Amgoud and Besnard, 2010), and satisfying $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$.

Definition 2.3.3 — Conflict-dependency. Let $\mathscr{R} \subseteq \mathtt{Arg}(\mathscr{L}) \times \mathtt{Arg}(\mathscr{L})$ be an attack relation. We say that $\mathscr{R}$ is conflict-dependent if and only if for every $a, b \in \mathtt{Arg}(\mathscr{L})$

$$(a,b) \in \mathscr{R} \text{ implies } \mathtt{Supp}(a) \cup \mathtt{Supp}(b) \vdash \bot.$$

We now prove that conflict-dependence is a necessary condition for satisfying $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$. To be completely precise, we here specify that we say that a *semantics σ returns conflict-free sets* if and only if for every argumentation system $(\mathscr{A}, \mathscr{R})$, for every $\mathscr{E} \in \sigma(\mathscr{A}, \mathscr{R})$, it holds that $\mathscr{E}$ is conflict-free with respect to $\mathscr{R}$.

Proposition 2.3.1 Let $\mathscr{R}$ be an attack relation and σ a semantics returning conflict-free sets. If $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$, then $\mathscr{R}$ is conflict-dependent.

Having proved this, we know that a relation satisfying $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ enjoys all the properties of conflict-dependent relations. For example, it was shown that if an attack relation is conflict-dependent, then there are no self-attacking arguments (Amgoud and Besnard, 2009).

Corollary 2.3.2 Let $\mathscr{R}$ be an attack relation and σ a semantics returning conflict-free sets. If $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$ then for every argument $a \in \mathtt{Arg}(\mathscr{L})$, we have that $(a,a) \notin \mathscr{R}$.

This means that we have another way to identify (some of the) attack relations not satisfying $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$. Namely, if for an attack relation there exists a self-attacking argument, then the given attack relation falsifies $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ for all semantics returning conflict-free sets. Let us now study the notion of validity (Amgoud and Besnard, 2010).

Definition 2.3.4 — Validity. Let $\mathscr{R} \subseteq \mathtt{Arg}(\mathscr{L}) \times \mathtt{Arg}(\mathscr{L})$ be an attack relation. We say that $\mathscr{R}$ is valid if and only if for every $\mathscr{E} \subseteq \mathtt{Arg}(\mathscr{L})$ it holds that if $\mathscr{E}$ is conflict-free, then $\mathtt{Base}(\mathscr{E})$ is consistent.

Let us now show that this property is incompatible with conflict-dependence.

Proposition 2.3.3 There exists no attack relation which is both conflict-dependent and valid.

This means that if an attack relation $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ then there must exist a set $\mathscr{E}$ which is conflict-free with respect to $\mathscr{R}$ but whose base is inconsistent.

Corollary 2.3.4 Let $\mathscr{R}$ be an attack relation, σ be an acceptability semantics returning conflict-free sets and let $\mathscr{R}$ satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$. Then, $\mathscr{R}$ is not valid.

The previous result is useful since if an attack relation is valid, we can immediately conclude that it violates $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$ for all (possible) semantics returning conflict-free sets.

On the more general level, we see that asking for every conflict-free set to have a consistent base is very demanding. Roughly speaking, this is due to the fact that attacks are binary whereas minimal conflicts may be ternary (or of a greater cardinality). Some authors argue that to obtain a consistent result, one should concentrate on admissibility and not on conflict-freeness. For example, Caminada and Vesic (2012) claim that n-ary attacks, for $n \geq 3$, are "simulated" in Dung's framework throughout the notion of admissibility. Thus, an idea for future work could be to study an alternative condition, which is that every *admissible* set has a consistent base.

We now study the properties related to particular semantics. We use abbreviations c, p, s, ss, g and i for respectively complete, preferred, stable, semi-stable, grounded and ideal semantics. For example, $p(\mathscr{F})$ denotes the set of preferred extensions argumentation system $\mathscr{F}$.

We show that if an attack relation satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ for stable semantics, then it satisfies it for semi-stable semantics also. Then we identify conditions under which an attack relation satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ for stable semantics. We provide a similar result for preferred semantics. We also identify a sufficient condition so that an attack relation falsifies $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ under complete semantics. Then, we discuss the case of single-extension semantics, like grounded and ideal.

First, suppose that $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$. This means that for every finite set of formulae Σ, function $\mathtt{Arg}$ is a bijection between $\mathtt{MC}(\Sigma)$ and $\sigma(\mathtt{Arg}(\Sigma), \mathscr{R})$. Since every finite set of formulae has at least one maximal consistent subset (even if that is the empty set) then for every Σ, it must be that $(\mathtt{Arg}(\Sigma), \mathscr{R})$ has at least one stable extension. Since there are stable extensions, then stable and semi-stable semantics coincide (Caminada, 2006a). Thus, we obtain the following proposition.

Proposition 2.3.5 Let $\mathscr{R}$ be an attack relation. If $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_s)$ then:

- for every finite set of formulae Σ and $\mathscr{F} = (\mathtt{Arg}(\Sigma), \mathscr{R})$, we have that $s(\mathscr{F}) = ss(\mathscr{F})$
- $\mathscr{R}$ satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_{ss})$.

The next result identifies two sufficient conditions that imply that a given attack relation satisfies $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ for stable and preferred semantics.

Proposition 2.3.6 Let $\mathscr{R}$ be an attack relation. If for every set of formulae Σ and $\mathscr{F} = (\mathtt{Arg}(\Sigma), \mathscr{R})$

- for all $S \in \mathtt{MC}(\Sigma)$, $\mathtt{Arg}(S) \in s(\mathscr{F})$, and
- for all $\mathscr{E} \in p(\mathscr{F})$, $\mathtt{Base}(\mathscr{E})$ is consistent

then $\mathscr{R}$ satisfies both $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_s)$ and $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_p)$.

Note that if an attack relation returns a stable extension having an inconsistent base, then it violates $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ for stable, semi-stable, preferred and complete semantics.

Proposition 2.3.7 Let $\mathscr{R}$ be an attack relation. If there exists a finite set of formulae Σ such that $\mathscr{F} = (\mathtt{Arg}(\Sigma), \mathscr{R})$ has a stable extension $\mathscr{E}$ such that $\mathtt{Base}(\mathscr{E})$ is inconsistent, then $\mathscr{R}$ falsifies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$ for $\sigma \in \{s, ss, p, c\}$.

Under a very mild condition, we can show that it is not possible for an attack relation to satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_c)$. Namely, the only condition we use in our result is that for every argument a, if a has a formula φ in its support, and $\neg\varphi \in \Sigma$, then there exists an argument $b \in \mathtt{Arg}(\Sigma)$ such that b attacks a.

Proposition 2.3.8 Let $\mathscr{R}$ be an attack relation such that for every finite set of formulae Σ, for every $a \in \mathtt{Arg}(\Sigma)$, for every $\varphi \in \mathtt{Supp}(a)$, if there exists $\psi \in \Sigma$ such that $\psi \equiv \neg\varphi$ then there exists $b \in \mathtt{Arg}(\Sigma)$ such that $(b, a) \in \mathscr{R}$. Then, $\mathscr{R}$ does not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_c)$.

It should be clear that each unique-extension semantics violates $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$. Note, however, that the sufficient conditions for $\mathscr{R}$ were identified (Gorogiannis and Hunter, 2011) so that for every finite set Σ and $\mathscr{F} = (\mathtt{Arg}(\Sigma), \mathscr{R})$ the grounded and the ideal semantics coincide and that the extension is exactly $\mathtt{Arg}(\Sigma \setminus (\Phi_1 \cup \ldots \cup \Phi_k))$ where $\{\Phi_1, \ldots, \Phi_k\}$ is the set of all minimal (for set inclusion) inconsistent subsets of Σ.

We showed how to identify properties that an attack relation satisfying $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ must satisfy. Also, we provided several results closely related to the choice of a specific acceptability semantics. We now identify classes of attack relations which satisfy, do not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$, or serve as lower (upper) bounds (with respect to set inclusion) for (non-)satisfying $(\texttt{MC} \leftrightarrow \texttt{Ext})$.

We first show that the whole class of *symmetric* attack relations *violates* $(\texttt{MC} \leftrightarrow \texttt{Ext})$ under stable, semi-stable, preferred, complete, grounded and ideal semantics.

Proposition 2.3.9 If $\mathscr{R}$ is a symmetric attack relation, then for every $\sigma \in \{s, ss, p, c, g, i\}$, $\mathscr{R}$ falsifies $(\texttt{MC} \leftrightarrow \texttt{Ext}_\sigma)$.

We now identify another class of attack relations that do not satisfy $(\texttt{MC} \leftrightarrow \texttt{Ext})$. Namely, we show that every (possible) attack generating "too many attacks" falsifies $(\texttt{MC} \leftrightarrow \texttt{Ext})$. First, we need to formally define what we mean by "too many attacks". We do this by introducing the notion of conflict-completeness.

Definition 2.3.5 — Conflict-complete. Let $\mathscr{R} \subseteq \texttt{Arg}(\mathscr{L}) \times \texttt{Arg}(\mathscr{L})$ be an attack relation. We say that $\mathscr{R}$ is conflict-complete if and only if for every minimal conflict $C \subseteq \mathscr{L}$ (i.e. for every inconsistent set whose every proper subset is consistent), for every $C_1, C_2 \subseteq C$ such that $C_1 \neq \emptyset$, $C_2 \neq \emptyset$, $C_1 \cup C_2 = C$, for every argument a_1 such that $\texttt{Supp}(a_1) = C_1$, there exists an argument a_2 such that $\texttt{Supp}(a_2) = C_2$ and $(a_2, a_1) \in \mathscr{R}$.

Intuitively, an attack relation is *conflict-complete* if when two sets form a minimal conflict, then every argument built from one of the two sets can be attacked by an argument from the other set. This notion is inspired by the desire to describe properties of a class of existing (and new) attack relations. For example, *canonical undercut*[6] is conflict-complete.

We can show that if an attack relation is *conflict-complete*, then it *falsifies* $(\texttt{MC} \leftrightarrow \texttt{Ext})$ for stable, semi-stable, preferred and complete semantics.

Proposition 2.3.10 Let $\mathscr{R}$ be an attack relation. If $\mathscr{R}$ is conflict-complete then $\mathscr{R}$ does not satisfy $(\texttt{MC} \leftrightarrow \texttt{Ext}_\sigma)$ for $\sigma \in \{s, ss, p, c\}$.

Until now, we provided some general results on classes of attack relations. We now study some particular attack relations. If $\Phi = \{\varphi_1, \dots, \varphi_k\}$ is a set of formulae, notation $\bigwedge \Phi$ stands for $\varphi_1 \wedge \dots \wedge \varphi_k$.

Definition 2.3.6 — Attack relations. Let $a, b \in \texttt{Arg}(\mathscr{L})$. We define the following attack relations:

- *defeat*: $a\mathscr{R}_d b$ if and only if $\texttt{Conc}(a) \vdash \bigwedge \neg\texttt{Supp}(b)$
- *direct defeat*: $a\mathscr{R}_{dd} b$ if and only if there exists $\varphi \in \texttt{Supp}(b)$ such that $\texttt{Conc}(a) \vdash \neg\varphi$
- *undercut*: $a\mathscr{R}_u b$ if and only if there exists $\Phi \subseteq \texttt{Supp}(b)$ such that $\texttt{Conc}(a) \equiv \neg \bigwedge \Phi$
- *direct undercut*: $a\mathscr{R}_{du} b$ if and only if there exists $\varphi \in \texttt{Supp}(b)$ such that $\texttt{Conc}(a) \equiv \neg\varphi$
- *canonical undercut*: $a\mathscr{R}_{cu} b$ if and only if $\texttt{Conc}(a) \equiv \neg \bigwedge \texttt{Supp}(b)$
- *rebut*: $a\mathscr{R}_r b$ if and only if $\texttt{Conc}(a) \equiv \neg\texttt{Conc}(b)$
- *defeating rebut*: $a\mathscr{R}_{dr} b$ if and only if $\texttt{Conc}(a) \vdash \neg\texttt{Conc}(b)$
- *conflicting attack*: $a\mathscr{R}_c b$ if and only if $\texttt{Supp}(a) \cup \texttt{Supp}(b) \vdash \bot$
- *rebut + direct undercut*: $a\mathscr{R}_{rdu} b$ if and only if $a\mathscr{R}_r b$ or $a\mathscr{R}_{du} b$
- *big argument attack*: $a\mathscr{R}_{ba} b$ if and only if there exists $\varphi \in \texttt{Supp}(b)$ s.t. $\texttt{Supp}(a) \vdash \neg\varphi$.

For more details about each of the attack relations, the reader is invited to read the discussion in the journal paper (Vesic, 2013).

By using our results, we can now show that every attack relation containing[7] *canonical undercut* falsifies $(\texttt{MC} \leftrightarrow \texttt{Ext})$ for stable, semi-stable, preferred and complete semantics.

Proposition 2.3.11 Let $\mathscr{R}$ be an attack relation. If $\mathscr{R}_{cu} \subseteq \mathscr{R}$, then $\mathscr{R}$ does not satisfy $(\texttt{MC} \leftrightarrow \texttt{Ext}_\sigma)$ for $\sigma \in \{s, ss, p, c\}$.

[6] defined very soon, in Definition 2.3.6
[7] in the set-theoretical sense

Therefore, we can identify several attack relations that do not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ under those semantics.

Corollary 2.3.12 Let $\mathscr{R}$ be an attack relation. If $\mathscr{R}_u \subseteq \mathscr{R}$, or $\mathscr{R}_d \subseteq \mathscr{R}$ or $\mathscr{R}_c \subseteq \mathscr{R}$ then $\mathscr{R}$ falsifies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_\sigma)$ for $\sigma \in \{s, ss, p, c\}$.

Hence, there is a whole class of attack relations based on undercutting which do not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$. We also identified another class of attack relations, this time based on rebutting, which do not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_s)$. Namely, every attack relation contained in defeating rebut falsifies $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_s)$.

Proposition 2.3.13 Let $\mathscr{R}$ be an attack relation. If $\mathscr{R} \subseteq \mathscr{R}_{dr}$ then $\mathscr{R}$ does not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext}_s)$.

Since $\mathscr{R}_r \subseteq \mathscr{R}_{dr}$, the previous conclusion holds for every relation contained in $\mathscr{R}_r$.

Let us now summarise the results regarding all particular attack relations[8].

Proposition 2.3.14 Attack relations $\mathscr{R}_{du}$, $\mathscr{R}_{dd}$ and $\mathscr{R}_{ba}$ satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ under stable, semi-stable and preferred semantics.

Proposition 2.3.15 Attack relations $\mathscr{R}_d$, $\mathscr{R}_u$, $\mathscr{R}_{cu}$, $\mathscr{R}_r$, $\mathscr{R}_{dr}$, $\mathscr{R}_{rdu}$, $\mathscr{R}_c$ falsify $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ under stable, semi-stable and preferred semantics.

Proposition 2.3.16 All the attack relations from Definition 2.3.6 falsify $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ under complete, grounded and ideal semantics.

We showed several properties regarding satisfying $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$, namely that attack relations belonging to this class must be conflict-dependent, must not be valid, must not be conflict-complete, must not be symmetric etc. We also showed some inclusion results (e.g. if an attack relation contains *canonical undercut* then it is not a member of this class).

In the next section, our goal is to tackle exploring the space of attack relations that always return consistent results but do not belong to $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$.

2.3.2 Beyond maxi-consistent operators

After considering virtually all the attack relations from the literature, we can observe that there are no instantiations of Dung's framework where extensions are always consistent (more precisely their bases are consistent sets of formulas) but that return results different from $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$. Why is this? Is it even possible to define an attack relation that returns a consistent result different from the maxi-consistent sets? More formally, is there an attack relation $\mathscr{R}$ that does not satisfy $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ and such that the base of each extension is a consistent set? In this section, we show that there are large classes of such attack relations.[9]

The main idea behind the class of instantiations we propose is that the arguments made from "less inconsistent" formulae have "more chance" to be in extensions. This means that we need a tool for indicating how inconsistent a set or a formulae is. In this section, we use Shapley Inconsistency Values (Hunter and Konieczny, 2010) to obtain that measure. This concept for measuring inconsistency is inspired by Shapley Value (Shapley, 1953).

There is an important remark that has to be done here. An argument's support is supposed to be consistent throughout this habilitation thesis. How can then one talk about arguments made from "more" or "less" consistent formulas? We will define those notions soon, but the intuition is that a formula φ is "more" inconsistent than ψ if φ contributes more to the inconsistency of the whole knowledge base than ψ.

[8] The interested reader is again referred to the journal paper (Vesic, 2013) for more details.

[9] This section summarizes mostly the work originally published by Vesic and van der Torre (2012). The interested reader is referred to that publication for more details.

Another important remark is about the attack relations. Note that all the attack relations from Definition 2.3.6 are defined on $\mathtt{Arg}(\mathscr{L}) \times \mathtt{Arg}(\mathscr{L})$. For a given Σ, one can just use the restriction of the relation from $\mathtt{Arg}(\mathscr{L}) \times \mathtt{Arg}(\mathscr{L})$ to $\mathtt{Arg}(\Sigma) \times \mathtt{Arg}(\Sigma)$. This is not the case with the attack relations we use in this section. Namely, for some attack relations we use, there exist arguments $a, b \in \mathtt{Arg}(\mathscr{L})$, such that whether a attacks b or not depends on the knowledge base Σ. Formally, the more general case is when an attack relation is defined by specifying its behaviour on every $\mathtt{Arg}(\Sigma) \times \mathtt{Arg}(\Sigma)$ for every finite $\Sigma \subseteq \mathscr{L}$. In the rest of this section, when we use the term "attack relation", we refer to the more general case. Formally, one should write $(a, b, \Sigma) \in \mathscr{R}$. However, since it is always clear to which Σ we refer to, there is no danger of confusion and in order to simplify the notation we simply write $(a, b) \in \mathscr{R}$ or $a\mathscr{R}b$ throughout the chapter.

The main idea behind the class of instantiations we propose is that a user is free to choose a basic inconsistency measure, under the condition that it satisfies the four properties stated in the following definition.

Definition 2.3.7 — Basic inconsistency measure (Hunter and Konieczny, 2010). A basic inconsistency measure I is a function that for every finite set of formulae returns a real number and satisfies the following properties for all finite sets $\Sigma, \Sigma' \subseteq \mathscr{L}$ and all formulae $\varphi, \psi \in \mathscr{L}$:

- $I(\Sigma) = 0$ if and only if Σ is a consistent set (Consistency)
- $I(\Sigma \cup \Sigma') \geq I(\Sigma)$ (Monotony)
- If φ is a free formula of $\Sigma \cup \varphi$, then $I(\Sigma \cup \varphi) = I(\Sigma)$ (Free Formula Independence)
- If $\varphi \vdash \psi$ and $\varphi \nvdash \bot$, then $I(\Sigma \cup \{\varphi\}) \geq I(\Sigma \cup \{\psi\})$ (Dominance)

The corresponding Shapley Inconsistency Value can then be calculated automatically. Different basic inconsistency measures induce different Shapley Inconsistency Values.

Note that, originally, Shapley's idea was to measure the merit of an individual in a coalition. Here, the idea is to use it to measure the "blame" of a formula for the inconsistency of a knowledge base. To do that, the identical mathematical expression as in the Shapley's work is used, but with different interpretation. The next definition shows how to, roughly speaking, distribute the "inconsistency blame" between the formulas.

Definition 2.3.8 — Shapley Inconsistency Value (Hunter and Konieczny, 2010). Let $\Sigma \subseteq \mathscr{L}$ and let I be a basic inconsistency measure. We define the corresponding Shapley Inconsistency Value (SIV), noted S^I, as follows. For every $\varphi \in \Sigma$:

$$S^I_\varphi(\Sigma) = \sum_{S \subseteq \Sigma} \frac{(|S|-1)!(|\Sigma|-|S|)!}{|\Sigma|!}(I(S) - I(S \setminus \{\varphi\})).$$

It has been shown that the previous formula is the only one that satisfies a set of intuitive axioms for measuring inconsistency (Hunter and Konieczny, 2010). This SIV gives a value for each formula of the base Σ. Thus, the previous definition allows us to define to what extent a formula is concerned with the inconsistencies. Note that for a formula φ, SIV depends essentially on the sum of differences of inconsistencies of all subsets of Σ together and without φ. Those values are then just multiplied with coefficients which depend only on the cardinalities of the corresponding sets. So, the main intuition can be resumed in: "How much does inconsistency decrease when φ is removed?"

We now use SIV to define an instantiation of Dung's abstract argumentation theory. Namely, once we are given a basic inconsistency measure, we can obtain the corresponding SIV, and use it to compare the formulas of the knowledge base. We first define how to construct a stratified version of a knowledge base, where the least inconsistent formulae (according to a given measure) are put in Σ_0 and the most inconsistent ones in Σ_n.

Definition 2.3.9 — Inconsistency ordered version of a knowledge base. Let I be a basic inconsistency measure, and S^I the corresponding SIV. Let $\Sigma \subseteq \mathscr{L}$ be a knowledge base. The inconsistency ordered version of Σ (with respect to I) is an n-tuple $(\Sigma_0, \ldots, \Sigma_n)$ such that

- $\Sigma_0 \cup \ldots \cup \Sigma_n = \Sigma$,
- for every $i, j \in \{0, \ldots, n\}$, if $i \neq j$ then $\Sigma_i \cap \Sigma_j = \emptyset$,
- for any two formulae $\varphi, \psi \in \Sigma$ such that $\varphi \in \Sigma_i$ and $\psi \in \Sigma_j$, we have

$$S^I_\varphi(\Sigma) \geq S^I_\psi(\Sigma) \quad \text{if and only if} \quad i \geq j.$$

Note that such a stratification was also used by Konieczny and Roussel (2013) to build a reasoning platform based on an inconsistency value. This order induces a preference on Σ, which can be used to define a preference relation on $\mathtt{Arg}(\Sigma)$. Let us first define a level of a formula and of an argument.

Definition 2.3.10 — Level of formulas and arguments. Let I be a basic inconsistency measure, S^I the corresponding SIV, let $\Sigma \subseteq \mathscr{L}$ be a knowledge base and $(\Sigma_0, \ldots, \Sigma_n)$ its inconsistency ordered version with respect to I. For a formula $\varphi \in \Sigma$,

$$\mathtt{level}(\varphi) = i \text{ if and only if } \varphi \in \Sigma_i.$$

For an argument $a \in \mathtt{Arg}(\Sigma)$,

$$\mathtt{level}(a) = max_{\varphi \in \mathtt{Supp}(a)} \mathtt{level}(\varphi).$$

We can now define an attack relation taking into account the level of formulas.

Definition 2.3.11 — Direct undercut on the ordered knowledge base. Direct undercut on the ordered knowledge base $(\Sigma_0, \ldots, \Sigma_n)$ is a relation $\mathscr{R}_{duo}$ defined as: $a \mathscr{R}_{duo} b$ if and only if $(a \mathscr{R}_{du} b$ and $\mathtt{level}(a) \leq \mathtt{level}(b))$ or $(b \mathscr{R}_{du} a$ and $\mathtt{level}(a) < \mathtt{level}(b))$.

We can now show the two main results of our study. The first one is that the extensions are consistent, closed for $\vdash$ and for sub-argument relation[10].

For a set of arguments $\mathscr{E}$, let us denote by $\mathtt{Concs}(\mathscr{E})$ the set of conclusions of all the arguments from $\mathscr{E}$. Formally, $\mathtt{Concs}(\mathscr{E}) = \{\mathtt{Conc}(a) \mid a \in \mathscr{E}\}$.

Proposition 2.3.17 Let I be a basic inconsistency measure and S^I the corresponding Shapley inconsistency measure. Let $\Sigma \subseteq \mathscr{L}$ be a knowledge base and $(\Sigma_0, \ldots, \Sigma_n)$ its inconsistency ordered version. Let $\mathscr{E}$ be a stable extension of $(\mathtt{Arg}(\Sigma), \mathscr{R}_{duo})$. Then:

- $\mathtt{Base}(\mathscr{E})$ and $\mathtt{Concs}(\mathscr{E})$ are consistent sets
- $\mathtt{Concs}(\mathscr{E})$ is closed for $\vdash$, i.e. for every $\varphi \in \mathscr{L}$, if $\mathtt{Concs}(\mathscr{E}) \vdash \varphi$ then $\varphi \in \mathtt{Concs}(\mathscr{E})$,
- $\mathscr{E}$ is closed for sub-arguments, i.e. if $a \in \mathscr{E}$ and b is an argument such that $\mathtt{Supp}(b) \subseteq \mathtt{Supp}(a)$, then $b \in \mathscr{E}$.

The second result shows that, by following the approach we describe in this section, one obtains a refinement of the approach that returns the extensions corresponding to the maximal consistent subsets of the knowledge base. Namely, if a basic inconsistency measure is used to order the knowledge base, and $\mathscr{R}_{duo}$ is then applied to calculate the extensions under stable semantics, every extension corresponds to exactly one maximal consistent subset of Σ, but there might be some maximal consistent subsets of Σ which do not correspond to any extensions. The next proposition shows that for every extension, there exists a maximal consistent subset of Σ corresponding to that extension. Example 11 from the paper by Vesic and van der Torre (2012) shows that there can exist maximal consistent sets which do not correspond to any stable extensions.

[10] We suppose the definition of sub-argument by Gorogiannis and Hunter (2011).

Proposition 2.3.18 Let I be a basic inconsistency measure and S^I the corresponding SIV. Let $\Sigma \subseteq \mathscr{L}$ be a knowledge base and $(\Sigma_0, \dots, \Sigma_n)$ its inconsistency ordered version. Then[11]:

$$s((\mathtt{Arg}(\Sigma), \mathscr{R})) \subseteq \{\mathtt{Arg}(S) \mid S \in \mathtt{MC}(\Sigma)\}$$

What are the maxi-consistent subsets of the knowledge base that are left out, i.e. that do not correspond to any stable extensions? While the exact result depends on the particular inconsistency measure, the general rule is that the ones that are kept are the least inconsistent ones. So, our system uses argumentation and inconsistency values to filter out only the extensions built from the formulas that have low inconsistency scores.

2.3.3 On the link between argumentation and OBDA semantics

In this section, we show the link between semantics that were defined in the context of ontology-based data access (Poggi et al., 2008) and the semantics from argumentation theory. The area of ontology-based data access (OBDA) studies the problem of inferring from an inconsistent ontology. To deal with such a situation, different semantics have been defined (e.g. AR, IAR, ICR). In this section we show that:

- sceptical acceptance under stable and preferred semantics corresponds to ICR semantics
- universal acceptance under stable and preferred semantics corresponds to AR semantics
- acceptance under grounded semantics corresponds to IAR semantics.

We keep this section brief and informal. For more information, the interested reader is referred to our papers (Croitoru and Vesic, 2013; Yun et al., 2018e).

The existential rules language is composed of formulas built with the usual quantifiers $(\exists, \forall)$ and *only* two connectors: implication $(\rightarrow)$ and conjunction $(\wedge)$ and is composed of facts, rules and negative constraints. A *fact* is a ground atom of the form $p(t_1, \dots, t_k)$ where p is a predicate of arity k and t_i, with $i \in [1, \dots, k]$, constants.

An existential *rule* is of the form $\forall \vec{X}, \vec{Y}\ H[\vec{X}, \vec{Y}] \rightarrow \exists \vec{Z} C[\vec{Z}, \vec{X}]$ where H and C are existentially closed atoms or conjunctions of existentially closed atoms and $\vec{X}, \vec{Y}, \vec{Z}$ their respective vectors of variables. A *rule is applicable* on a set of facts $\mathscr{F}$ iff there exists a homomorphism (Baget et al., 2016) from H to $\mathscr{F}$. Applying a rule to a set of facts (also called *chase*) consists of adding the set of atoms of C to the facts according to the application homomorphism. Different *chase* mechanisms use different restrictions that prevent infinite redundancies (Baget et al., 2011). Here, we use recognisable classes of existential rules where the chase is guaranteed to stop (Baget et al., 2011).

A *negative constraint* is a rule of the form $\forall \vec{X}, \vec{Y}\ H[\vec{X}, \vec{Y}] \rightarrow \perp$ where H is an existentially closed atom or a conjunction of existentially closed atoms, $\vec{X}, \vec{Y}$, their respective vectors of variables and $\perp$ is *absurdum*. Please note that the number of atoms in H is not bounded and that negative constraints generalise simple binary conflicts that can easily be translated between the two representations: $\neg p(\vec{X})$ is transformed into $np(\vec{X})$ and the negative constraint $p(\vec{X}) \wedge np(\vec{X}) \rightarrow \perp$ is added to the rules.

We say that F_1 *entails* F_2 denoted by $F_1 \models F_2$ iff there is a homomorphism from the set of atoms in F_2 to the set of atoms in F_1 where F_1 and F_2 are two existentially closed conjunctions of atoms. A conjunctive query is an existentially quantified conjunction of atoms. For readability, we restrict ourselves to Boolean conjunctive queries, which are closed formulas (the framework and the obtained results can be extended to general conjunctive queries).

A *knowledge base* $\mathscr{K}$ is a tuple $\mathscr{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$ where $\mathscr{F}$ is a finite set of facts, $\mathscr{R}$ a set of existential rules and $\mathscr{N}$ a set of negative constraints.

[11] recall that notation $s(\mathscr{F})$ stands for the set of stable extensions of $\mathscr{F}$

The *closure* of $\mathscr{F}$ by $\mathscr{R}$ is the set of all possible atoms and conjunctions of atoms that are entailed, after using all possible rule applications from $\mathscr{R}$ over $\mathscr{F}$ until a fixed point. The output of this process is called the closure and is denoted by $\mathtt{Cl}_{\mathscr{R}}(\mathscr{F})$. A set $\mathscr{F}$ is said to be *$\mathscr{R}$-consistent* if no negative constraint hypothesis can be entailed, i.e. $\mathtt{Cl}_{\mathscr{R}}(\mathscr{F}) \not\models \bot$. Otherwise, $\mathscr{F}$ is said to be *$\mathscr{R}$-inconsistent*.

Maximal for set inclusion *$\mathscr{R}$-consistent* sets (repairs) of $\mathscr{K}$ are defined as

$$\mathscr{R}epairs(\mathscr{K}) = \{X \subseteq \mathscr{F} \mid \mathtt{Cl}_{\mathscr{R}}(X) \not\models \bot \text{ and for every } X', X \subset X' \text{ implies } \mathtt{Cl}_{\mathscr{R}}(X) \models \bot\}$$

Once the repairs calculated, there are different ways to calculate the set of facts that follow from an inconsistent knowledge base. For example, we may want to accept a query if it is entailed in *all repairs* (AR semantics).

Definition 2.3.12 Let $\mathscr{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$ be a knowledge base and let α be a query. Then α is **AR-entailed** from $\mathscr{K}$, written $\mathscr{K} \models_{AR} \alpha$ iff for every repair $A' \in \mathscr{R}epairs(\mathtt{K})$, it holds that $\mathtt{Cl}_{\mathscr{R}}(A') \models \alpha$.

Another possibility is to check whether the query is entailed from the *intersection of closed repairs* (ICR semantics).

Definition 2.3.13 Let $\mathscr{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$ be a knowledge base and let α be a query. Then α is **ICR-entailed** from $\mathscr{K}$, written $\mathscr{K} \models_{ICR} \alpha$ iff $\bigcap_{A' \in \mathscr{R}epairs(\mathtt{K})} \mathtt{Cl}_{\mathscr{R}}(A') \models \alpha$.

One could also consider the *intersection of all repairs* and then close this intersection under the rules (IAR semantics).

Definition 2.3.14 Let $\mathscr{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$ be a knowledge base and let α be a query. Then α is **IAR-entailed** from $\mathscr{K}$, written $\mathscr{K} \models_{IAR} \alpha$ iff $\mathtt{Cl}_{\mathscr{R}}(\bigcap_{A' \in \mathscr{R}epairs(\mathtt{K})}) \models \alpha$.

The three semantics are different (Lembo et al., 2010; Bienvenu, 2012).

We now define an instantation of Dung's theory to reason with an inconsistent ontological knowledge base. We first define the notion of an argument.

Definition 2.3.15 Let $\mathtt{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$ be a knowledge base. The corresponding argumentation graph is the pair $\mathtt{AF}_{\mathscr{K}} = (\mathscr{A}, \mathscr{C})$, where $\mathscr{A}$ is the set of arguments and $\mathscr{C}$ the set of attacks. An argument is a pair (H, C) such that

- $H \subseteq \mathscr{F}$
- $H \neq \emptyset$
- H is $\mathscr{R}$-consistent
- $C \subseteq \mathtt{Cl}_{\mathscr{R}}(H)$
- H is minimal, i.e. there exists no H' such that $H \subsetneq H'$ and H' satisfies all the above conditions.

We say that $a = (H, C)$ attacks $b = (H', C')$ denoted by $(a, b) \in \mathscr{C}$ if and only if there exists $\varphi \in H'$ such that $C \cup \{\varphi\}$ is $\mathscr{R}$-inconsistent.

As usual, we call H support and C conclusion of argument (H, C); notation $\mathtt{Arg}(S)$ denotes the set of arguments constructed from the set of facts $S \subseteq \mathscr{F}$; for a set of arguments $\mathscr{E}$, we denote $\mathtt{Base}(\mathscr{E}) = \bigcup_{a \in \mathscr{E}} \mathtt{Supp}(a)$. We omit $\mathtt{K}$, $\mathscr{F}$, $\mathscr{R}$ and $\mathscr{N}$ when they are clear from the context.

The output of an argumentation graph is usually defined as the set of conclusions that appear in all the extensions (Caminada and Amgoud, 2007, Definition 12):

Definition 2.3.16 — Output of an argumentation graph. Let $\mathtt{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$ be a knowledge base and $\mathtt{AF}_{\mathscr{K}}$ the corresponding argumentation graph. The output of $\mathtt{AF}_{\mathscr{K}}$ under semantics

σ is defined as:

$$\texttt{Output}_\sigma(\texttt{AF}_\mathscr{K}) = \bigcap_{\mathscr{E}\in\sigma(\texttt{AF}_\mathscr{K})} \texttt{Concs}(\mathscr{E}).$$

When $\sigma(\texttt{AF}_\mathscr{K}) = \emptyset$, we define $\texttt{Output}(\texttt{AF}_\mathscr{K}) = \emptyset$ by convention.

Note that the previous definition asks for existence of a conclusion in every extension. This kind of acceptance is usually referred to as *sceptical* acceptance. We say that a query α is sceptically accepted if it is a logical consequence of the output of $\texttt{AF}_\mathscr{K}$:

Definition 2.3.17 — Sceptical acceptance of a query. Let $\texttt{K} = (\mathscr{F},\mathscr{R},\mathscr{N})$ be a knowledge base and $\texttt{AF}_\mathscr{K}$ the corresponding argumentation graph. A query α is sceptically accepted under semantics x if and only if $\texttt{Output}_\sigma(\texttt{AF}_\mathscr{K}) \models \alpha$.

It is possible to make an alternative definition, which uses the notion of universal acceptance instead of sceptical one. According to universal criteria, a query α is accepted if it is a logical consequence of conclusions of every extension:

Definition 2.3.18 — Universal acceptance of a query. Let $\texttt{K} = (\mathscr{F},\mathscr{R},\mathscr{N})$ be a knowledge base and $\texttt{AF}_\mathscr{K}$ the corresponding argumentation framework. A query α is universally accepted under semantics x if and only if for every extension $\mathscr{E} \in \sigma(\texttt{AF}_\mathscr{K})$, it holds that $\texttt{Concs}(\mathscr{E}) \models \alpha$.

In general, universal and sceptical acceptance of a query do not coincide. Note that for single-extension semantics (e.g. grounded), the notions of sceptical and universal acceptance coincide. So we simply use the word "accepted" in this context.

Definition 2.3.19 — Acceptance of a query. Let $\texttt{K} = (\mathscr{F},\mathscr{R},\mathscr{N})$ be a knowledge base, $\texttt{AF}_\mathscr{K}$ the corresponding argumentation framework, σ a single-extension semantics and $\mathscr{E}$ the unique extension of $\texttt{AF}_\mathscr{K}$ under σ. A query α is accepted under σ if and only if $\texttt{Concs}(\mathscr{E}) \models \alpha$.

We now prove that the repairs of the knowledge base correspond exactly to the stable (and preferred, since in this instantiation the stable and the preferred semantics coincide) extensions of the argumentation framework.

Theorem 2.3.19 Let $\texttt{K} = (\mathscr{F},\mathscr{R},\mathscr{N})$ be a knowledge base, $\texttt{AF}_\mathscr{K}$ the corresponding argumentation framework and $\sigma \in \{s,p\}$. Then:

$$\sigma(\texttt{AF}_\mathscr{K}) = \{\texttt{Arg}(A') \mid A' \in \mathscr{R}epairs(\texttt{K})\}$$

We can also show that the intersection of all the repairs of the knowledge base corresponds to the grounded extension.

Theorem 2.3.20 Let $\texttt{K} = (\mathscr{F},\mathscr{R},\mathscr{N})$ be a knowledge base and $\texttt{AF}_\mathscr{K}$ the corresponding argumentation framework. Denote the grounded extension of $\texttt{AF}_\mathscr{K}$ by $\texttt{GE}$. Then:

$$\texttt{GE} = \texttt{Arg}\left(\bigcap_{A'\in\mathscr{R}epairs(\texttt{K})} A'\right)$$

We are now ready to prove the links between argumentation-based semantics and OBDA semantics.

Theorem 2.3.21 Let $\mathscr{K} = (\mathscr{F},\mathscr{R},\mathscr{N})$ be a knowledge base, $\texttt{AF}_\mathscr{K}$ the corresponding argumentation framework and α a query. Then:

- $\mathtt{K} \models_{ICR} \alpha$ iff α is sceptically accepted under stable/preferred semantics
- $\mathtt{K} \models_{AR} \alpha$ iff α is universally accepted under stable/preferred semantics
- $\mathtt{K} \models_{IAR} \alpha$ iff α is accepted under grounded semantics

Note that when the argumentation graph is generated from a knowledge base $\mathscr{K} = (\mathscr{F}, \mathscr{R}, \mathscr{N})$, stable and preferred semantics coincide.

We presented several links between argument-based and non argument-based approaches. In what follows, we present the links between extension-based and ranking-based semantics.

2.3.4 On the link between extension-based and ranking-based semantics

Until now, the reader certainly noticed the differences between ranking-based an extension-based semantics. One of the main differences is the form of the output: in case of a ranking-based semantics we are given an order (i.e. a ranking) and in case of extension-based semantics we are given a collection of sets (i.e. extensions). While the ranking and the scores might allow to better assess the acceptability degree of each individual argument, the question *"what are the points of view of the argumentation framework?"* stays unanswered when using a ranking-based semantics. We now describe our contribution in answering that question. We define a new modular framework that generates viewpoints (i.e. extensions) based on ranking argumentation semantics by considering a selection function, a ranking on arguments and a lifting function as its input parameters.

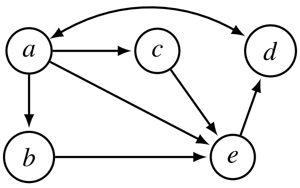

Figure 2.6: Deciding whether to prefer one strong argument (a) or several weaker arguments together (b, c and d)

Consider the argumentation framework from Figure 2.6. Let us use h-categorizer. We obtain $a \succ b, c \succ d \succ e$. What are the possible points of view? Clearly $\mathscr{E}_1 = \{a\}$ is admissible and strong, but should we also accept $\mathscr{E}_2 = \{b, c, d\}$? Since $\mathscr{E}_2$ contains three arguments, is it (even) better than $\mathscr{E}_1$? It seems that both $\mathscr{E}_1$ and $\mathscr{E}_2$ are acceptable points of view, but which one is stronger: $\mathscr{E}_1$ having only one highly ranked argument or $\mathscr{E}_2$ having three arguments of medium strength?

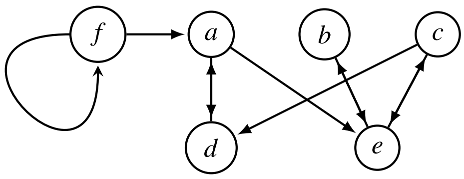

Figure 2.7: Deciding whether to prefer a non admissible set containing highly ranked arguments (a, b, c) or an admissible set containing weaker arguments (d, e)

Consider now the AF from Figure 2.7. Let us use h-categorizer ranking semantics. The three strongest arguments are a, b and c with scores approximately 0.76, 0.71 and 0.71, respectively. Arguments d and e are much weaker with scores approximately 0.40 and 0.31, respectively. If one looks for conflict-free sets containing highly ranked arguments, two potential candidates are

$\mathscr{E}_1 = \{a,b,c\}$ and $\mathscr{E}_2 = \{d,e\}$. Should one accept $\mathscr{E}_1$, which contains the three strongest arguments but is not admissible (it is attacked by a self-attacking argument)?

The two previous examples show that the question *"how to generate points of view when arguments are evaluated by a ranking-based semantics?"* is not easy. In particular, the strongest arguments with respect to the ranking-based semantics do not always form an admissible (or not even a conflict-free) set.

Of course, the previous question depends of the content of the arguments and no fully abstract solution can give the definitive answer[12]. In a real-life situation (especially if it is about a high stake decision, e.g. choosing a treatment plan for a patient) one should further explore the arguments contents, hypothesis, try to fetch more information about the topic (e.g. maybe the patient can be examined again to get more details, another doctor's opinion can be taken into account to help in decision making). One could also more closely examine the reason why the set in question is not admissible, the reason that there is a self-attacking argument etc by asking: did we forget some arguments?, is the knowledge represented correctly? However, there are cases when this is not possible. Take, for instance, the case of an on-line debate where thousands of arguments are stated by different users. We must be prepared for errors in arguments, reasoning, representation, duplicate arguments etc. In such cases, no one expects the argumentation semantics to model the perfect reasoner, since already the input is fuzzy and might contain errors. Under such circumstances, our goal is to make a robust framework for estimating the strengths of arguments. Of course, the output of such a framework is not used to take an automated decision. It is rather shown to the humans in order to help them to make further reasoning.

Given all those considerations, it is clear that there is no unique or "best" answer to the aforementioned question. Instead, *we propose to define a general and modular framework able to suit different needs*. Our framework is composed of three layers.

1. First, a *selection function* selects a set of subsets of arguments. This can be any function; some popular choices would probably be all maximal conflict-free sets or all admissible sets.
2. Second, a *ranking-based semantics* is used to calculate the ranking on the set of arguments. In the previous example we mentioned h-categorizer, but any function returning an order on the set of arguments can be used.
3. Third, we need a *lifting operator*, i.e. a function that compares the set of arguments returned by the selection function and produces a ranking on those sets, based on the individual scores of arguments. A simple criterion would be to compare the strongest arguments of each set. A generalisation of this criterion is so-called *leximax*, which in case that the best arguments are equally strong proceeds to compare the second best arguments of each set, and so on. If a ranking-based semantics returns numerical scores, one could also compare the sums of scores of all arguments.

Out of all sets returned by the selection function, only the best ones (w.r.t. the ranking function and the lifting operator) are kept and they represent the output of our framework.

Let us give a quick preview of the kind of results our framework can return. In the example from Figure 2.6 the scores of arguments with respect to h-categorizer ranking semantics are approximately 0.67 for a, 0.60 for b and c, 0.50 for d and 0.35 for e. Let us select all maximal conflict-free sets; we obtain $\mathscr{E}_1 = \{a\}$, $\mathscr{E}_2 = \{b,c,d\}$ and $\mathscr{E}_3 = \{e\}$. Now, if we use leximax or max as lifting, we will obtain $\mathscr{E}_1$ as the output of our framework. If we use the sum of arguments' scores as the lifting function, we obtain $\mathscr{E}_2$. So, according to the user's choice of the lifting function, one of the two results is obtained.

Consider now the example from Figure 2.7. If admissibility is important for the user, they will select only (a subset of) admissible sets, thus $\{a,b,c\}$ will not be a part of the output. On the

[12]like in the case of compensation-based semantics between a small number of strong and a large number of weak attacks, which we studied in Subsection 2.2.5

contrary, if we select all the maximal conflict-free sets, we obtain $\mathscr{E}_1 = \{a,b,c\}$ and $\mathscr{E}_2 = \{d,e\}$. For all reasonable lifting operators $\mathscr{E}_1$ is preferred to $\mathscr{E}_2$, thus the output of the framework is $\mathscr{E}_1$ in this case. To keep this habilitation thesis within the expected page limit, we drop the formal details of our approach. The interested reader is referred to the corresponding publication (Yun et al., 2018d). That paper also contains:

- the comparison between different combinations of parameters (in the three above-mentioned levels) and the corresponding (non-)inclusion results
- an analysis of how the properties of the underlying functions (e.g. selection function) impact the properties of the system as a whole.

At the end of this chapter, I now briefly mention some other works in argumentation that I consider important, but that I will not recall in more details. I provide the references to the published papers.

- We studied the question: if an argument is attacked by several other arguments, to which extent each of those attacks contributes to the loss of the strength of the attacked argument? We introduced the novel concept of contribution measure for evaluating those contributions. We defined a set of principles such a measure should satisfy and identified a unique measure that satisfies those principles (Amgoud et al., 2017b).
- We studied the problem of aggregation of argumentation graphs (Delobelle et al., 2018). Under the hypothesis that an argumentation graph represents the beliefs of an agent, the aggregation aims to represent the beliefs of the group of agents. Some aggregation operators as well as some rationality properties were defined for aggregation of argumentation graphs. We studied the existing operators as well as new ones, which we defined in light of the proposed properties. We highlight the fact that existing operators do not satisfy a lot of properties. The conclusions are that on one hand none of the existing operators seem fully satisfactory, but on the other hand some of the properties proposed so far seem too demanding.
- We studied the properties of argumentation graphs generated from inconsistent knowledge bases expressed using existential rules (Yun et al., 2018b). We developed techniques to optimize the generation of argument graphs in order to reduce the number of arguments by dropping some non-essential arguments or in order to generate the graph faster (Yun et al., 2018c, 2020a). We also showed how to generate a benchmark of logical argumentation graphs (Yun et al., 2018a).

2.3.5 Summary and outlook

This section studied the links between argumentation and non-argumentation approaches. We first identified and studied the large class of instantiations of Dung's abstract theory corresponding to the *maxi-consistent operator*. In other words, we studied the instantiations where every extension of the argumentation system corresponds to exactly one maximal consistent subset of the knowledge base. We proved properties of attack relations belonging to this class: they must be conflict-dependent, must not be valid, must not be conflict-complete, must not be symmetric etc. We also identified some attack relations serving as lower or upper bounds of the class. By using our results, we showed for all existing attack relations from the argumentation literature whether or not they belong to this class. We also showed that an attack relation not depending on arguments' conclusions can return reasonable results. Furthermore, we showed that such a relation is a member of $(\mathtt{MC} \leftrightarrow \mathtt{Ext})$ class. We now discuss the potential benefits of this and the similar studies.

(I) A case when an instantiation of Dung's theory is shown to correspond to an existing operator.

First, such a work can help to "validate" an argumentation-based approach by showing in which cases it returns a result comparable with that of a non argumentation-based approach. The possible criticism of such an instantiation is that it is useless, since one can obtain the same result without using argumentation. But, this is far from being true; namely, argumentation can be used for explanatory purposes. For example, if one wants to know why a certain conclusion is accepted, an argument having that conclusion can be presented. That argument can be attacked by other arguments and so on. Also, it might be possible to construct only a *part of the argumentation graph* related to the argument in question, thus having a better knowledge representation (i.e. ignoring the parts of the knowledge base unrelated to the argument one wants to concentrate on).

The second benefit of this type of work is that it can help to *reduce computational complexity* by using the simpler approach in the cases when the result obtained by an argumentation-based approaches and a non argumentation-based approaches is the same. Observe that the work in this category (capturing an operator with an instantiation of Dung's theory) is far from being limited to the case of the maxi-consistent operator, as we saw in Subsection 2.3.2 that there exists a large class of instantiations of the abstract argumentation theory returning a consistent result substantially different from the one returned by the maxi-consistent operator.

(II) A case when an instantiation of Dung's theory does not correspond to any existing operator.

Working on the links between instantiations of Dung's theory and operators can be even more beneficial in the case when an instantiation of the abstract argumentation theory not corresponding to any known operator happens to be found. We distinguish three possible situations.

(a) A case when an instantiation calculates a "useful" result which can be obtained by an operator, but that operator was unknown until now. In such a case, a new operator is discovered thanks to argumentation. The question is then, in which situations to use argumentative approach, and when to apply the operator? The answer depends on the balance between the need for computational efficiency (which we conjecture is often on the side of the approach directly applying the operator) and the need to represent knowledge in a format that is easy to grasp, argue and justify an accepted piece of knowledge, which are the usual advantages of argumentation.

(b) A case when an instantiation of Dung's abstract theory returns a "useful" result which cannot be obtained by any operator. Recall that an operator is a function that, for every finite knowledge base, returns a set of its subsets. But, an argumentative approach could return a result that cannot be represented in that form, for instance, if an argument (Φ, α) is in an extension, whereas (Φ, β) is not, with $\alpha \neq \beta$. Thus, the expressive power of the operator-based approach might be not enough to distinguish those subtleties. A very important question of how to define such an instantiation is still open. Another relevant issue is to see in which context such instantiations make sense and how they can be applied.

(c) A case when an instantiation returns a "bad" result. This class regroups a set of instantiations representing a behaviour one would like to avoid. The general question: "how to distinguish *useful* from *bad* instantiations?" is certainly a hard one. Apart from a scientific debate, evaluation can include tests on a set of benchmark examples. However, the limits of testing a reasoning formalism on a set of benchmark examples have been pointed out by Vreeswijk (1995). Another, more principled (and more demanding) way to proceed is to define a set of postulates to be satisfied by an argumentation formalism (Caminada and Amgoud, 2007; Caminada et al., 2012).

As a remark, note that the fact that an instantiation may return an inconsistent result, does not mean that it is completely useless. Namely, there might be cases when arguments are constructed from an inconsistent knowledge base, and when one resolves just some of the existing inconsistencies by an argumentative approach, and *then* applies another inconsistency-tolerant approach. Also, inconsistency handling is not the only use of argumentation. Thus, still in the same setting, a drastic case would be to first use argumentation for another purpose (not dealing with at all with inconsistencies) and then apply a different approach to reason with inconsistency.

We also showed it is possible to instantiate Dung's abstract argumentation theory with classical propositional logic and obtain a meaningful result which is different from the maximal consistent subsets of the knowledge base. Indeed, we defined a whole class of instantiations that return different results. Furthermore, we showed that these instantiations are sound in the sense that they satisfy the postulates from argumentation literature (e.g. consistency, closure).

In the rest of the section, we also showed the link between argumentation semantics and the semantics used in ontology-based data access and linked the ranking-based and extension-based semantics.

3. Reducing the number of extensions

In this chapter we present the works that help to reduce the number of extensions or maxi-consistent sets. This is beneficial in the situations when there are many extensions, which leads to very little or no skeptical conclusions. The first section presents our work on this challenge in the argumentation framework context, the second and the third sections are devoted to the logical context without argumentation. The second section defines a class of inference relations based on a selection of maxi-consistent subsets of an inconsistent knowledge base expressed in propositional logic whereas the third section studies how to reduce the number of repairs in an OBDA setting.

3.1 Supported inference and extension selection

This section presents two approaches for deriving more arguments from an abstract argumentation framework than the ones obtained using sceptical inference, that is often too cautious. The first approach consists in selecting only some of the extensions. We point out several choice criteria to achieve such a selection process. The second approach consists of the definition of a new inference policy, between sceptical and credulous inference.

Observe that when the number of extensions is large, using a sceptical / credulous approach can be sub-optimal. Namely, if there is a lot of extensions, only few (if any) arguments are in all of them. Thus, using sceptical inference gives almost no information. Conversely, the credulous approach may result in too many arguments.

There exist settings for abstract argumentation where preferences, weighted attacks or similar extra information are considered (Cayrol et al., 2010; Dunne et al., 2011; Coste-Marquis et al., 2012; Amgoud and Vesic, 2014). This additional information can be exploited to reduce the number of extensions. In contrast, the problem addressed here is to increase the number of accepted arguments when there is no further data, i.e. other data except the arguments and the attacks between them.

The two main acceptance policies are sceptical and credulous policies. We say that x is *sceptically accepted* under semantics σ (or in short *s-sceptically accepted*) iff $\sigma(\mathscr{F}) \neq \emptyset$ and $x \in \bigcap_{\mathscr{E} \in \sigma(\mathscr{F})} \mathscr{E}$. Argument x is *credulously accepted* under semantics σ iff $x \in \bigcup_{\mathscr{E} \in \sigma(\mathscr{F})} \mathscr{E}$. We denote the set of sceptically accepted arguments by $\mathtt{Sc}_\sigma(\mathscr{F})$ and the set of credulously accepted arguments by $\mathtt{Cr}_\sigma(\mathscr{F})$. We denote by $\mathscr{R}_{\downarrow S}$ the restriction of attack relation $\mathscr{R}$ on set S.

3.1.1 Pairwise comparison of extensions

This subsection studies the way to select the "best" extensions based on the following process:

1. Compare all pairs of extensions based on a given criterion (e.g. the number of arguments in one extension not attacked by the other extension)
2. Choose the "best" extension(s) given the winners of pairwise comparisons

Definition 3.1.1 — Pairwise comparison criteria. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation graph and σ be a semantics. Let $\mathscr{E}, \mathscr{E}' \in \sigma(\mathscr{F})$. Then:

1. $\mathscr{E} \succeq_{nonatt} \mathscr{E}'$ if the number of arguments in $\mathscr{E}$ non attacked by $\mathscr{E}'$ is greater than or equal to the number of arguments in $\mathscr{E}'$ non attacked by arguments of $\mathscr{E}$
2. $\mathscr{E} \succeq_{strdef} \mathscr{E}'$ if the number of arguments in $\mathscr{E}$ strongly defended from $\mathscr{E}'$ by $\mathscr{E}$ is greater than or equal to the number of arguments in $\mathscr{E}'$ strongly defended from $\mathscr{E}$ by $\mathscr{E}'$
3. $\mathscr{E} \succeq_{delarg} \mathscr{E}'$ if the cardinality of any largest subset S of $\mathscr{E}$ such that *if all the attacks from S to $\mathscr{E}'$ are deleted then $\mathscr{E}$ is an extension of* $(\mathscr{E} \cup \mathscr{E}', \mathscr{R}_{\downarrow \mathscr{E} \cup \mathscr{E}'})$ is greater than or equal to the cardinality of any largest subset S' of $\mathscr{E}'$ such that *if all the attacks from S' to $\mathscr{E}$ are deleted then $\mathscr{E}$ is an extension of* $(\mathscr{E} \cup \mathscr{E}', \mathscr{R}_{\downarrow \mathscr{E} \cup \mathscr{E}'})$
4. $\mathscr{E}' \succeq_{delatt} \mathscr{E}'$ if the maximal number of attacks from $\mathscr{E}$ to $\mathscr{E}'$ that can be deleted such that $\mathscr{E}$ is still an extension of $(\mathscr{E} \cup \mathscr{E}', \mathscr{R}_{\downarrow \mathscr{E} \cup \mathscr{E}'})$ is greater than or equal to the maximal number of attacks from $\mathscr{E}'$ to $\mathscr{E}$ that can be deleted such that $\mathscr{E}'$ is still an extension of $(\mathscr{E} \cup \mathscr{E}', \mathscr{R}_{\downarrow \mathscr{E} \cup \mathscr{E}'})$

The two first criteria are based on the number of non attacked or (strongly) defended arguments. The last two ones are based on a notion of robustness from attacks stemming from the other extension. One could also consider other criteria, for example by comparing the total number of attacks from $\mathscr{E}$ to $\mathscr{E}'$ and the total number of attacks from $\mathscr{E}'$ to $\mathscr{E}$. For a criterion γ, we write $\mathscr{E} \succ_\gamma \mathscr{E}'$ iff $\mathscr{E} \succeq_\gamma \mathscr{E}'$ and it is not the case that $\mathscr{E}' \succeq_\gamma \mathscr{E}$. We also write $\mathscr{E} \sim_\gamma \mathscr{E}'$ iff $\mathscr{E} \succeq_\gamma \mathscr{E}'$ and $\mathscr{E}' \succeq_\gamma \mathscr{E}$. This relation between the extensions allows us to select only some of the extensions.

Definition 3.1.2 — Copeland-based extensions. Let $\gamma \in \{nonatt, strdef, delarg, delatt\}$ be one of the criteria from Definition 3.1.1. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation system and σ a semantics. We define the set of Copeland-based extensions (CBE) as follows

$$\mathtt{CBE}_{\sigma,\gamma}(\mathscr{F}) = \underset{\mathscr{E} \in \sigma(\mathscr{F})}{\arg\max} |\{\mathscr{E}' \in \sigma(\mathscr{F}) \mid \mathscr{E} \succeq_\gamma \mathscr{E}'\}| - |\{\mathscr{E}'' \in \sigma(\mathscr{F}) \mid \mathscr{E}'' \succeq_\gamma \mathscr{E}\}|$$

We call this selection "Copeland-based" since it is inspired by the Copeland's rule from voting theory (Moulin, 1988). Of course, one can envisage other ways to select the extensions given criterion γ, for instance all voting methods based on the majority graph (Brams and Fishburn, 2002).

Clearly, selecting some extensions is a way to increase the number of sceptically accepted arguments (and to decrease the number of credulously accepted arguments):

Observation 12 For every $\gamma \in \{nonatt, strdef, delarg, delatt\}$, for every semantics σ, for every AS $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, for every $x \in \mathscr{A}$:

- $\mathtt{CBE}_{\sigma,\gamma}(\mathscr{F}) \subseteq \sigma(\mathscr{F})$
- if x is σ-sceptically accepted then x is $\mathtt{CBE}_{\sigma,\gamma}$-sceptically accepted
- if x is $\mathtt{CBE}_{\sigma,\gamma}$-credulously accepted then it is σ-credulously accepted.

We can show that the semantics defined in this section satisfy the same properties as the underlying semantics they are built from, with the notable exception of directionality.

Proposition 3.1.1 Let x be any property among I-maximality, admissibility, strong admissibility, reinstatement, weak reinstatement, CF-reinstatement. If the semantics σ satisfies property x, then the semantics $\mathtt{CBE}_{\sigma,\gamma}$ satisfies property x.

Note that the relations among different semantics do not carry over in case of CBE approach. For instance, it is not guaranteed that each CBE-stable extension is also a CBE-preferred extension.

3.1.2 Comparing extensions by global evaluation

In the previous subsection we considered different criteria for *pairwise comparison* of extensions. In this subsection we define the score of an argument as the number of extensions it appears in. One may justify this choice of score as some kind of generalization of the principles behind sceptical acceptance. For sceptical acceptance a "good" argument is an argument that appears in all extensions. But, if no such argument exists, it could make sense to consider that arguments that appear in many (or almost all) extensions are also good, and typically better than the ones that appears in less extensions. Note that one can use other scores in the construction and obtain similar results.

Definition 3.1.3 — Scores and support vectors. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation system, σ a semantics, and a an argument. We define $\mathtt{ne}$ as the number of extensions a appears in. Formally, $\mathtt{ne}_\sigma(a, \mathscr{F}) = |\{\mathscr{E} \in \sigma(\mathscr{F}) \mid a \in \mathscr{E}\}|$. For an extension $\mathscr{E} \in \sigma(\mathscr{F})$, with $\mathscr{E} = \{a_1, \dots, a_n\}$ we define its support as $\mathtt{vsupp}_\sigma(\mathscr{E}, \mathscr{F}) = (\mathtt{ne}_\sigma(a_1, \mathscr{F}), \dots, \mathtt{ne}_\sigma(a_n, \mathscr{F}))$.

When $\mathscr{F}$ and σ are clear from the context, we write $\mathtt{ne}(a)$ and $\mathtt{vsupp}(\mathscr{E})$ instead of $\mathtt{ne}_\sigma(a, \mathscr{F})$ and $\mathtt{vsupp}_\sigma(\mathscr{E}, \mathscr{F})$.

The next definition introduces some popular choices of aggregation functions.

Definition 3.1.4 — Aggregation functions. Let $v = (v_1, \dots, v_n)$ be a vector of natural numbers. We denote by $sum(v)$ the sum of all elements of v, by $max(v)$ the maximal element of v, by $min(v)$ the minimal element of v, by $leximax(v)$ the re-arranged version of v where $v_1, \dots, v_n$ are put in decreasing order, by $leximin(v)$ the re-arranged version of v where $v_1, \dots, v_n$ are put in increasing order. Let $v = (v_1, \dots, v_n)$ and $v' = (v'_1, \dots, v'_n)$ be two vectors of natural numbers. We say that $v <_{lex} v'$ iff $\exists j \in 1, \dots, n (\forall i \in 1, \dots, j-1, v_i = v'_i)$ and $v_j < v'_j$. We say that $v <_{leximin} v'$ iff $leximin(v) <_{lex} leximin(v')$ and $v <_{leximax} v'$ iff $leximax(v) <_{lex} leximax(v')$.

We can now define order-based extensions as those that maximise the aggregated scores of their arguments. Namely, the idea is to calculate the popularity of an extension by taking into account the popularity of the arguments it contains.

Definition 3.1.5 — Order-based extensions. Let $\mathscr{F} = (\mathscr{A}, \mathscr{R})$ be an argumentation system, σ a semantics, and γ be an aggregation function. We define

$$\mathtt{OBE}_{\sigma,\gamma}(\mathscr{F}) = \arg\max_{\mathscr{E} \in \sigma(\mathscr{F})} \gamma(\mathtt{vsupp}_\sigma(\mathscr{E}, \mathscr{F}))$$

As with the Copeland-based extensions, with order-based extensions we also increase the number of sceptically accepted arguments and decrease the number of credulously accepted arguments.

Observation 13 For every $\gamma \in \{sum, max, min, leximin, leximax\}$, for every semantics σ, for every AS $\mathscr{F} = (\mathscr{A}, \mathscr{R})$, for every $x \in \mathscr{A}$:

- $\mathtt{OBE}_{\sigma,\gamma}(\mathscr{F}) \subseteq \sigma(\mathscr{F})$
- if x is σ-sceptically accepted then x is $\mathtt{OBE}_{\sigma,\gamma}$-sceptically accepted
- if x is $\mathtt{OBE}_{\sigma,\gamma}$-credulously accepted then it is σ-credulously accepted.

Like in the previous subsection, we can show that some properties of semantics are preserved after extension selection.

Proposition 3.1.2 Let x be any property among I-maximality, Admissibility, Strong Admissibility, Reinstatement, Weak Reinstatement, CF-Reinstatement (Baroni and Giacomin, 2007).

If the semantics σ satisfies property x, then the semantics $\mathtt{OBE}_{\sigma,\gamma}$ satisfies property x.

Directionality is again not satisfied.

We skip the study about the links between the criteria. Roughly speaking, it shows that the criteria are incomparable except for the fact that leximin (resp. leximax) refines min (resp. max). The interested reader is referred to the original publication (Konieczny et al., 2015).

3.1.3 Support-based acceptance policy

This subsection presents a completely different approach for selecting arguments. We focus on arguments that appear in the greatest number of extensions to construct what we call "candidate sets". Then, an argument is called *supportedly accepted* if it is in all the candidate sets.

Roughly speaking, a candidate set is a set constructed by adding arguments to the empty set, starting from the arguments having the biggest $\mathtt{ne}$ score and continuing as long as the resulting set stays conflict-free.

Definition 3.1.6 — Candidate sets. Let $\succeq_{\mathtt{ne}}$ be defined by $x \succeq_{\mathtt{ne}} y$ if and only if $\mathtt{ne}_\sigma(x,\mathscr{F}) \geq \mathtt{ne}_\sigma(y,\mathscr{F})$. Let $\mathscr{F} = (\mathscr{A},\mathscr{R})$ be an AS and let σ be a semantics. Let $|\mathscr{A}| = m$. We say that a permutation θ of $\{1,\ldots,m\}$ is compatible with $\succeq_{\mathtt{ne}}$ if and only if $a_{\theta(1)} \succeq_{\mathtt{ne}} \cdots \succeq_{\mathtt{ne}} a_{\theta(m)}$. A set $\mathscr{E} \subseteq \mathscr{A}$ is a candidate set of $\mathscr{F}$ under semantics σ iff there exists a permutation θ of $\{1,\ldots,m\}$ such that $>_\theta$ is compatible with $\succeq_{\mathtt{ne}}$ and $\mathscr{E}$ is obtained by the following greedy procedure:

$S := \emptyset$;
for $j = 1,\ldots,m$ do
  if $(\mathtt{ne}_\sigma(a_{\theta(j)},\mathscr{F}) \geq 1)$ and $(S \cup \{a_{\theta(j)}\}$ is conflict-free$)$
  then $S := S \cup \{a_{\theta(j)}\}$
end for;
$\mathscr{E} := S$.

We denote the set of candidate sets of $\mathscr{F}$ under σ by $\mathtt{CS}_\sigma(\mathscr{F})$.

Note that, in general, neither each candidate set is an extension nor each extension is a candidate set. Observe also that the construction of candidate sets is reminiscent to the one of preferred subbases from a stratified belief base with respect to the inclusion-based ordering (Benferhat et al., 1993a); here the belief base consists of all the arguments and the stratification is based on the $\mathtt{ne}_\sigma(.,\mathscr{F})$ score.

We now define an argument as being supportedly accepted if it is in all candidate sets.

Definition 3.1.7 — Supported acceptance. Let $\mathscr{F} = (\mathscr{A},\mathscr{R})$ be an AS, σ be a semantics and let $x \in \mathscr{A}$. We say that x is supportedly accepted under semantics σ iff $x \in \bigcap_{\mathscr{E}\in\mathtt{CS}_\sigma(\mathscr{F})}$. We denote the set of supportedly accepted arguments $\mathtt{Sp}_\sigma(\mathscr{F})$.

We can show that supported inference is "between" sceptical and credulous inference.

Proposition 3.1.3 For every AS $\mathscr{F} = (\mathscr{A},\mathscr{R})$, for every semantics σ returning conflict-free extensions:

$$\mathtt{Sc}_\sigma(\mathscr{F}) \subseteq \mathtt{Sp}_\sigma(\mathscr{F}) \subseteq \mathtt{Cr}_\sigma(\mathscr{F})$$

Note that the condition that σ returns conflict-free extensions is necessary to ensure the link between sceptical and supported acceptance. However, this is not an issue, since all the well-known semantics return conflict-free sets.

A major drawback of credulous inference is that the set of inferred arguments is not always conflict-free. This is problematic since all these arguments cannot be accepted together in such a case. Sceptical inference does not suffer from this problem since the set of inferred arguments is ensured to be conflict-free. Supported inference satisfies this important property:

Observation 14 For any $\mathscr{F}$, the set of supportedly accepted arguments is conflict-free.

Note that this set is not necessarily admissible. This should not be shocking since the same observation can be made for the set of sceptically accepted arguments.

We drop the study of the links between supported inference and other approaches presented in the previous subsections. The interested reader is again referred to the original publication (Konieczny et al., 2015).

In this section, we saw how to reduce the number of extensions of an argumentation system. The next section studies the ways to reduce the number of maxi-consistent subsets of a knowledge base and analyses the consequence relations obtained from them.

3.2 Rational inference relations from maxi-consistent subsets selection

When reasoning from an inconsistent knowledge base, a natural approach is to take advantage of the maximal consistent subsets of the base. In this section, we study inference relations based on selection of some maximal consistent subsets of an inconsistent knowledge base, leading thus to inference relations with a stronger inferential power. We study the principles that must be satisfied by the selection process to ensure that it leads to an inference relation which is rational. Our formal setting is classical propositional logic. Thus we consider a language defined from a finite set of propositional variables and the usual connectives. A belief base is a finite set of formulas.

3.2.1 Inference from selected maximal consistent subsets

Standard notions when facing inconsistent belief bases are minimal inconsistent subsets, that encode the sources of conflicts in the base; and maximal consistent subsets, which can be considered as the potential repairs of the inconsistent belief base. We already introduced the notation $\mathtt{MC}(\mathtt{K})$, let us now define the set of minimal inconsistent subsets of a knowledge base.

Definition 3.2.1 — $\mathtt{mus}$. $\mathtt{mus}$ is a mapping defined as follows: for every belief base $\mathtt{K}$, $\mathtt{mus}(\mathtt{K})$ is the set of all minimal (for set inclusion) inconsistent subsets of $\mathtt{K}$:

- $\mathtt{K}' \subseteq \mathtt{K}$
- $\mathtt{K}'$ is not consistent
- If $\mathtt{K}'' \subset \mathtt{K}'$, then $\mathtt{K}''$ is consistent

We will first define mappings that attach a score to each formula α of $\mathtt{K}$ and then aggregate those scores (in the following, we will simply sum up the scores; as will be discussed later, other aggregation functions could be considered alternatively).

Definition 3.2.2 — scoring function. A *scoring function* s associates with a belief base $\mathtt{K}$ and a formula $\alpha \in \mathtt{K}$ a non-negative real number $s(\mathtt{K},\alpha)$ which is equal to 0 if and only if α is a trivial formula (i.e., such that $\alpha \equiv \top$ or $\alpha \equiv \bot$).

We now present three examples of scoring functions. The first one is based on the number of maximal consistent sets a formula belongs to.

Definition 3.2.3 — $\#\mathtt{MC}$. Let $\mathtt{K}$ be a belief base and $\alpha \in \mathtt{K}$. We define:

$$\#\mathtt{MC}(\mathtt{K},\alpha) = \begin{cases} 0 & \text{if } \alpha \text{ is trivial} \\ |\{\mathtt{K}_i \in \mathtt{MC}(\mathtt{K}) \mid \alpha \in \mathtt{K}_i\}| & \text{otherwise} \end{cases}$$

Another interesting example of a scoring function is based on the number of minimal inconsistent sets a formula belongs to. The scale must be reversed here, since this number must be minimized if one wants to give some preference to the less conflicting formulae. The addition of 1 in the second part of the definition is introduced in order to make sure that only trivial formulas can get the score 0.

Definition 3.2.4 — $\#\mathtt{mus}$. Let $\mathtt{K}$ be a belief base and $\alpha \in \mathtt{K}$. We define:

$$\#\mathtt{mus}(\mathtt{K},\alpha) = \begin{cases} 0 \text{ if } \alpha \text{ is trivial, otherwise} \\ 1 + |\mathtt{mus}(\mathtt{K})| - \\ \qquad |\{\mathtt{K}_i \subseteq \mathtt{K} \mid \mathtt{K}_i \in \mathtt{mus}(\mathtt{K}), \alpha \in \mathtt{K}_i\}| \end{cases}$$

We could also use an inconsistency measure in order to attach scores to formulae. Let us consider the measure *MIV*, introduced by Hunter and Konieczny (2010), which is based both on the number of minimal inconsistent subsets containing a formula and on their cardinalities. The idea is that belonging to a large inconsistent set puts less blame on a formula than belonging to a small set.

Definition 3.2.5 — $\mathtt{MIV}$. Let $\mathtt{K}$ be a belief base and $\alpha \in \mathtt{K}$. We define:

$$\mathtt{MIV}_{\mathtt{K}}(\alpha) = \sum_{M \in \mathtt{mus}(\mathtt{K}), \alpha \in M} \frac{1}{|M|}$$

We can use $\mathtt{MIV}$ to define a scoring function $\mathtt{miv}$ as follows:

Definition 3.2.6 — $\mathtt{miv}$. Let $\mathtt{K}$ be a belief base and $\alpha \in \mathtt{K}$. We define

$$\mathtt{maxmiv}(\mathtt{K}) = max_{\alpha \in \mathtt{K}} \mathtt{MIV}_{\mathtt{K}}(\alpha)$$

and

$$\mathtt{miv}(\mathtt{K},\alpha) = \begin{cases} 0 & \text{if } \alpha \text{ trivial} \\ 1 + \mathtt{maxmiv}(\mathtt{K}) - \mathtt{MIV}_{\mathtt{K}}(\alpha) & \text{otherwise} \end{cases}$$

On this ground, the score of any subset of $\mathtt{K}$ can be computed by aggregating (e.g., using sum) the scores of its elements:

Definition 3.2.7 — $\mathtt{score}^s_{\mathtt{K},sum}$. Let s be a scoring function. Let $\mathtt{K}$ be a belief base, and let $\mathtt{K}_i \subseteq \mathtt{K}$. We note $\mathtt{score}^s_{\mathtt{K},sum}(\mathtt{K}_i) = \sum_{\alpha \in \mathtt{K}_i} s(\mathtt{K},\alpha)$.

In particular, we have:

$$\mathtt{score}^{\#mc}_{\mathtt{K},sum}(\mathtt{K}_i) = \sum_{\alpha \in \mathtt{K}_i} \#\mathtt{MC}(\mathtt{K},\alpha).$$

$$\mathtt{score}^{\#mus}_{\mathtt{K},sum}(\mathtt{K}_i) = \sum_{\alpha \in \mathtt{K}_i} \#\mathtt{mus}(\mathtt{K},\alpha).$$

$$\mathtt{score}^{\mathtt{miv}}_{\mathtt{K},sum}(\mathtt{K}_i) = \sum_{\alpha \in \mathtt{K}_i} \mathtt{miv}(\mathtt{K},\alpha).$$

Let us now show how to infer conclusions from a belief base. We first need the following notation that will prove convenient: $\mathtt{MC}(\mathtt{K},\alpha) = \{\mathtt{K}_i \subseteq \mathtt{K} \mid \mathtt{K}_i \cup \{\alpha\} \in \mathtt{MC}(\mathtt{K} \cup \{\alpha\})\}$.

Definition 3.2.8 — Inference from subsets with best scores. Let $\mathtt{K}$ be a belief base and α and β be two formulas. Let s be a scoring function. We define $\mathtt{MC}_{\mathtt{score}^s_{\mathtt{K},sum}}(\mathtt{K},\alpha) = \{\mathtt{K}_i \in \mathtt{MC}(\mathtt{K},\alpha)$ and there exists no $\mathtt{K}'_i \in \mathtt{MC}(\mathtt{K},\alpha)$ such that $\mathtt{score}^s_{\mathtt{K},sum}(\mathtt{K}'_i) > \mathtt{score}^s_{\mathtt{K},sum}(\mathtt{K}_i)\}$. We say that $\alpha \mid\!\sim^s_{\mathtt{K},sum} \beta$ if and only if either α is inconsistent, or for every $\mathtt{K}_i \in \mathtt{MC}_{\mathtt{score}^s_{\mathtt{K},sum}}(\mathtt{K},\alpha)$ we have $\mathtt{K}_i \cup \{\alpha\} \models \beta$.

3.2.2 Logical properties

We now formally evaluate the introduced methods w.r.t. their logical rationality. There has been much work on the issue of determining the minimum logical properties that any nonmonotonic inference relation should satisfy (Gabbay, 1985; Makinson, 1994; Kraus et al., 1990; Lehmann and Magidor, 1992). There is now a wide consensus on the fact that the minimal set of expected properties is the one of *preferential inference relations* (Lehmann and Magidor, 1992), also called *system P*, and that an interesting subclass is the one of *rational inference relations* (Kraus et al., 1990), also called *system R*. Preferential inference relations are characterized by the following postulates:

$$\alpha \mathrel{|\!\sim} \alpha \quad (Ref)$$

$$\frac{\models \alpha \leftrightarrow \beta,\ \ \alpha \mathrel{|\!\sim} \gamma}{\beta \mathrel{|\!\sim} \gamma} \quad (LLE)$$

$$\frac{\models \alpha \rightarrow \beta,\ \ \gamma \mathrel{|\!\sim} \alpha}{\gamma \mathrel{|\!\sim} \beta} \quad (RW)$$

$$\frac{\alpha \wedge \beta \mathrel{|\!\sim} \gamma,\ \alpha \mathrel{|\!\sim} \beta}{\alpha \mathrel{|\!\sim} \gamma} \quad (Cut)$$

$$\frac{\alpha \mathrel{|\!\sim} \gamma,\ \beta \mathrel{|\!\sim} \gamma}{\alpha \vee \beta \mathrel{|\!\sim} \gamma} \quad (Or)$$

$$\frac{\alpha \mathrel{|\!\sim} \beta,\ \alpha \mathrel{|\!\sim} \gamma}{\alpha \wedge \beta \mathrel{|\!\sim} \gamma} \quad (CM)$$

A *rational inference relation* is a preferential relation that also satisfies the RM (*rational monotony*) postulate:

$$\frac{\alpha \mathrel{|\!\not\sim} \neg\beta \quad \alpha \mathrel{|\!\sim} \gamma}{\alpha \wedge \beta \mathrel{|\!\sim} \gamma} \quad (RM)$$

We can show that the three relations we defined above satisfy all those postulates. Indeed, a more general result holds: we can use any scoring function and any aggregation function that satisfies the four properties specified in Definition 3.2.10.

Definition 3.2.9 — Aggregation function. $\oplus$ is an *aggregation function* if for every positive integer n, for every non-negative real number $x_1, \ldots, x_n$, $\oplus(x_1, \ldots, x_n)$ is a non-negative real number.

Definition 3.2.10 — Properties of aggregation function. An aggregation function $\oplus$ satisfies:

- **Composition** if $\oplus(x_1, \ldots, x_n) \leq \oplus(y_1, \ldots, y_n)$ implies $\oplus(x_1, \ldots, x_n, z) \leq \oplus(y_1, \ldots, y_n, z)$
- **Decomposition** if $\oplus(x_1, \ldots, x_n, z) \leq \oplus(y_1, \ldots, y_n, z)$ implies $\oplus(x_1, \ldots, x_n) \leq \oplus(y_1, \ldots, y_n)$
- **Symmetry** if for every permutation θ, $\oplus(x_1, \ldots, x_n) = \oplus(\theta(x_1, \ldots, x_n))$
- **Monotonicity** if for every $z > 0$ we have $\oplus(x_1, \ldots, x_n, z) > \oplus(x_1, \ldots, x_n)$

Composition, Decomposition and Symmetry were introduced by Konieczny et al. (2004). We add here a new property, Monotonicity, and slightly change Composition and Decomposition for dealing with tuples of different sizes.

Observe that *sum* satisfies the above conditions. Another well-known aggregation function satisfying the previous four properties is lexicographic aggregation (Konieczny et al., 2004; Dubois et al., 1996; Moulin, 1988).

One can then define the score of any subset of the input belief base is the aggregation of the corresponding scores.

Definition 3.2.11 — $\mathtt{score}^s_{\mathtt{K},\oplus}$. Let s be a scoring function and $\oplus$ an aggregation function. Let $\mathtt{K}$ be a belief base and $\mathtt{K}_i \subseteq \mathtt{K}$ with $\mathtt{K}_i = \{\alpha_1, \ldots, \alpha_n\}$.We define

$$\mathtt{score}^s_{\mathtt{K},\oplus}(\mathtt{K}_i) = \oplus_{\alpha \in \mathtt{K}_i} s(\mathtt{K}, \alpha).$$

Based on those scores, one can compare the subsets of the belief base.

Definition 3.2.12 — $\succeq_{\mathtt{K},\oplus}$. Let s be a scoring function and $\oplus$ an aggregation function. Let $\mathtt{K}$ be a belief base, $\mathtt{K}_i, \mathtt{K}_j \subseteq \mathtt{K}$. We state that $\mathtt{K}_i \succeq^s_{\mathtt{K},\oplus} \mathtt{K}_j$ if and only if $\mathtt{score}^s_{\mathtt{K},\oplus}(\mathtt{K}_i) \geq \mathtt{score}^s_{\mathtt{K},\oplus}(\mathtt{K}_j)$.

We can show that using a scoring function and an aggregation operator satisfying Composition, Decomposition, Symmetry and Monotonicity ensures to get a rational inference relation. However, even a more general result holds. We are not obliged to use a scoring function and compare the scores. It is sufficient to use what we call a *monotonic selection relation* to select some of the maximal consistent subsets of the belief base in order to guarantee that the corresponding inference relation is rational. The notion of monotonic selection relation is introduced in the next definition. It says that adding a non trivial formula in a set makes it better.

Definition 3.2.13 — Monotonic selection relation. Given a belief base $\mathtt{K}$, let $\succeq_{\mathtt{K}} \subseteq 2^{\mathtt{K}} \times 2^{\mathtt{K}}$ be a reflexive, transitive and total relation over the powerset of $\mathtt{K}$. $\succeq_{\mathtt{K}}$ is said to be a *monotonic selection relation* if for every consistent set $\mathtt{K}_i \subseteq \mathtt{K}$, for every non-trivial formula $\alpha \in \mathtt{K} \setminus \mathtt{K}_i$, $\mathtt{K}_i \cup \{\alpha\} \succ_{\mathtt{K}} \mathtt{K}_i$. We use the standard notation $\mathtt{K}_i \succ_{\mathtt{K}} \mathtt{K}_j$ for $\mathtt{K}_i \succeq_{\mathtt{K}} \mathtt{K}_j$ and not($\mathtt{K}_j \succeq_{\mathtt{K}} \mathtt{K}_i$).

For instance, consider the relation $\succeq_{card}$ that compares the subsets of the knowledge base $\mathtt{K}$ based on their cardinality:

Definition 3.2.14 — $\succeq_{card}$. For every two subsets $\mathtt{K}_i$, $\mathtt{K}_j$ of $\mathtt{K}$,
$\mathtt{K}_i \succeq_{card} \mathtt{K}_j$ if and only if $|\mathtt{K}_i| \geq |\mathtt{K}_j|$.

Clearly enough, $\succeq_{card}$ satisfies the conditions from Definition 3.2.13, thus it is a monotonic selection relation. On this ground, one can define a selection mechanism which consists in keeping only the best sets with respect to a monotonic selection relation:

Definition 3.2.15 — $\mathtt{MC}_{\succeq_{\mathtt{K}}}$. Given a belief base $\mathtt{K}$, a formula α, and a monotonic selection selection relation $\succeq_{\mathtt{K}}$, we define

$$\mathtt{MC}_{\succeq_{\mathtt{K}}}(K, \alpha) = \{\mathtt{K}_i \in \mathtt{MC}(\mathtt{K}, \alpha) \mid \text{ there exists no } \mathtt{K}'_i \in \mathtt{MC}(\mathtt{K}, \alpha) \text{ such that } \mathtt{K}'_i \succ_{\mathtt{K}} \mathtt{K}_i\}$$

We now show that all the scoring functions studied in this section induce monotonic selection relations. Formally, we identify sufficient conditions under which a relation comparing the maximal consistent subsets based on a scoring function is a monotonic selection relation.

Proposition 3.2.1 Let $\mathtt{K}$ be a belief base, $\oplus$ be an aggregation function satisfying Composition, Decomposition, Symmetry and Monotonicity and let s be any scoring function. Then $\succeq^s_{\mathtt{K},\oplus}$ is a monotonic selection relation.

Let us now generalize Definition 3.2.8:

Definition 3.2.16 — Inference from best subsets wrt $\succeq_{\mathtt{K}}$. Given a belief base $\mathtt{K}$, two formulae α and β, and a monotonic selection relation $\succeq_{\mathtt{K}}$, we state that $\alpha \mid\!\sim^{\mathtt{MC}_{\succeq_{\mathtt{K}}}}_{\mathtt{K}} \beta$ if and only if either α is inconsistent, or for every $\mathtt{K}_i \in \mathtt{MC}_{\succeq_{\mathtt{K}}}(K, \alpha)$ we have $\mathtt{K}_i \cup \{\alpha\} \models \beta$.

It can be shown that each monotonic selection relation induces a rational inference relation:

Proposition 3.2.2 If $\succeq_K$ is a monotonic selection relation, then $\mid\!\sim_K^{MC_{\succeq_K}}$ is rational.

As a corollary of the previous result, we conclude the following.

Corollary 3.2.3 Let K be a belief base, $\oplus$ an aggregation operator satisfying Composition, Decomposition, Symmetry and Monotonocity and let s be any scoring function. Then $\mid\!\sim_K^{MC_{\succeq^s_{K,\oplus}}}$ is rational.

We have seen that for each belief base K and each monotonic selection relation $\succeq_K$, $\mid\!\sim_K^{MC_{\succeq_K}}$ is rational. The converse is also true, namely, that for each rational relation $\mid\!\sim$, there exists a belief base K and a monotonic selection relation $\succeq_K$ such that $\mid\!\sim_K^{MC_{\succeq_K}} = \mid\!\sim$. More formally:

Proposition 3.2.4 For every rational relation $\mid\!\sim$ defined on a logical language built over a finite set of propositional variables, there exists a belief base K and a monotonic selection relation $\succeq_K$ such that $\mid\!\sim_K^{MC_{\succeq_K}} = \mid\!\sim$.

Putting the last two results together, we obtain the following proposition.

Proposition 3.2.5 A relation $\mid\!\sim$ is rational if and only if there exists a belief base K and a monotonic selection relation $\succeq_K$ such that $\mid\!\sim_K^{MC_{\succeq_K}} = \mid\!\sim$.

Note: This work was presented in two conferences (Konieczny et al., 2018, 2019). It is only after that, during the 2019 BRAON workshop[1] that it was pointed to us by Hans Rott that our Proposition 3.2.5 is a consequence of a result proved in the nineties by del Val (1997) regarding preferential relations. Note that the proofs are different: our proofs of satisfaction of properties are direct, in contrast to the proof by del Val.

3.3 Applying inconsistency measures for repair semantics in OBDA

In this section, we briefly mention a similar study[2] as in the previous one, but for a more practical point of view. We report parts of our recent work (Yun et al., 2018e) in which we place ourselves in the Ontology Based Data Access setting, define a framework that allows to select repairs, show how desirable properties of its components guarantee the desirable postulates of the whole framework, propose an efficient algorithm for computing the output of the framework and apply it in choosing the best packaging for strawberries. For brevity reasons, we informally present a subset of those results and refer the reader to the published version for more details.

The framework is composed of three layers. First, an inconsistency value is used in order to attach a number to each fact of the knowledge base. (This is exactly what we called a scoring function in the previous section.) Second, as in the previous section, we use an aggregation function to calculate the score of each repair. Third, we use one of the existing inconsistency tolerant inference relations, e.g. IAR, ICR.

We do not recall the definitions and intuitions related to inconsistency values, since we already mentioned them in Subsection 2.3.2 and Section 3.2. Similarly, we do not discuss aggregation functions in more detail here.

We developed an efficient algorithm for calculating the output of our framework and applied it in order to choose the best packaging for the strawberries. An online poll consisting of 66 questions was submitted to an audience of 21 professionals from the food industry. We distinguished four kinds of professionals: the wholesalers, the floorwalkers, the quality managers and the warehouse managers. The questions were aimed at collecting the individual vision of each person about the

[1] http://www4.uma.pt/braon2019/index.html

[2] in the two previous sections we minimized the number of extensions or the number of maxi-consistent sets; in this section, we minimize the number of repairs

characteristics of four packagings: the wooden packaging (WP), the plastic packaging with a plastic film (PPF), the plastic packaging with a rigid lid (PRL) and the opened plastic packaging without lid (OPL).

The answers of this poll were formalised into a set of 50 facts and 160 rules. In our application scenario, the inconsistency of the knowledge base comes from the fusion of the divergent visions of the several professionals about the four aforementioned packagings. These visions were explicitly expressed using the rules. A group of packaging experts constructed another set of 18 rules constituting expert knowledge. For instance, the rule $\forall x(keepHumidity(x) \rightarrow badFridgeConservation(x))$ states that if x is a packaging that keeps humidity then x is a bad packaging for fridge conservation. Lastly, a set of 34 negative constraints representing conflicting atoms, such as $\forall x(notBadEffectOnFruits(x) \wedge badEffectOnFruits(x) \rightarrow \perp)$, and incompatibilities between packagings, such as $\forall x,y,z,t(OPL(x,y) \wedge PRL(z,t) \rightarrow \perp)$, was added. The formalisation yielded a set of 33 repairs where each repair corresponds to the vision of a collection of individuals about a single packaging. For instance, the repair $OPL_2 = \{OPL(po, floorwalker0)$, $OPL(po, wharehouse\_manager0)\}$ corresponds to the vision of $floorwalker$ 0 and $wharehouse$ $manager$ 0 about the OPL. Amongst the 33 repairs, 16 concerned the WP, 6 concerned the PRL, 9 concerned the PPF and 2 concerned the OPL. The different number of repairs is explained by the diverse quantity of disagreements amongst individuals. For instance, only two wholesalers disagreed about the characteristics of the OPL whereas eight wholesalers disagreed about the characteristics of WP. In our model, the size of the repair corresponds to the number of individuals that agree on all the characteristics of a specific packaging. Note that the repairs are not ranked solely based on their cardinality.

Somehow surprisingly, the ranking on repairs was extremely clear as it showed that $WP > PRL \sim PPF > OPL$ (see Table 3.1). Indeed, all the repairs about the WP were ranked above the other repairs. The repairs about the PRL were ranked roughly equally with the repairs about PPF and the repairs about OPL were last. The ranking was evaluated by a group of packaging experts, who confirmed that the ranking on packagings was intuitive with respect to the data of the KB. Indeed, the experts acknowledged the fact that the WP was ranked first because its characteristics were less contested by the professionals.

The complete KB in DLGP format as well as a JAVA implementation of the tool for computing the output of our framework is accessible at `https://gite.lirmm.fr/yun/IJCAI2018`.

3.4 Summary

This section presented our work on reducing the number of extensions, maxi-consistent sets or repairs. Note that the work in argumentation is orthogonal to some other works (Coste-Marquis et al., 2005) that aimed at defining more prudent inference relations for Dung's argumentation frameworks (i.e., the objective is to derive less arguments). Contrariwise to our work, instead of selecting some extensions or defining a new inference policy, their approach consists in strengthening the usual conflict-freeness property to *indirect* conflict-freeness. Thus a prudent extension cannot contain two arguments when there exists an indirect attack among the first one and the second one. When the credulous policy and the preferred semantics (or the stable semantics) are considered, the set of derivable arguments from prudent extensions is included in the set of arguments derivable from the standard extensions.

Our work in reducing the number of maxi-consistent subsets is related to a number of approaches where inference is defined from maximal subsets of defaults (Reiter, 1980; Poole, 1988; Makinson, 2005). As already mentioned, the work of del Val (1997) implies our Proposition 3.2.5. Also, some propositions reported by Gärdenfors and Makinson (1994) are quite close to our results. Indeed, one can find several characterization results for nonmonotonic inference operators based on selection functions over maximal consistent subsets in their paper. The work by Gärdenfors

WP_1
$\vdots$
WP_{16}
PPF_1, PRL_1
PPF_2, PRL_2
PPF_3
PPF_4
PPF_5, PRL_3, PRL_4
PPF_6
PRL_5
PRL_6
PPF_7
$\vdots$
PPF_9
OPL_1
OPL_2

From “worst” to “best” repairs

Table 3.1: Ranking on repairs. For simplicity, repairs are denoted by the packaging they are referencing

and Makinson (1994) nevertheless departs from our own one from several aspects. Thus, while we focus on finite classical propositional logic, Gärdenfors and Makinson considered more general logical systems. Their representation theorem for rational inference also requires the base to be consistent and deductively closed, while in our approach, the base is finite, so it is not deductively closed, and it can be inconsistent.

In this section we also briefly presented our more practical work on applying inconsistency measures for repair semantics in OBDA, where we showed some concrete ways to select repairs, studied efficient algorithms and applied our framework in selecting the best packaging for strawberries.

4. Judgment aggregation

In this chapter we report our work in judgment aggregation (Lang et al., 2014, 2016, 2017). Judgment aggregation studies the problems related to aggregating a finite set of yes-no individual judgments, cast on a collection of logically interrelated issues. Such a finite set of issues forms the *agenda*. It can be seen as a generalisation of preference aggregation (Dietrich and List, 2007a).

Until a few years ago, the judgment aggregation literature had focused considerably more on studying impossibility theorems, than on developing and investigating specific aggregation rules. This field development approach departs from the, admittedly much older, field of voting theory. Nevertheless, several recent and independent papers have started to explore the concrete judgment aggregation rules, beyond the well known premise-based and conclusion-based rules (Dietrich and Mongin, 2010; Slavkovik and van der Torre, 2009). While the premise- and conclusion-based rules can only be applied if there exists a prior labelling of the agenda issues as premises and conclusions, the following rules are defined for every agenda: quota-based rules (Dietrich and List, 2007b), distance-based rules (Pigozzi, 2006; Miller and Osherson, 2009; Endriss et al., 2012; Duddy and Piggins, 2012), generalizations of Condorcet-consistent voting rules (Nehring et al., 2014; Lang et al., 2011), and rules based on the maximisation of some scoring function (Lang et al., 2011; Dietrich, 2014). Some of these rules obviously generalize well-known voting rules.

Our contribution is threefold. First, as there is so far no *compendium* of judgment aggregation rules, we give one: we list most of the rules that have been proposed recently, in a structured way. This part of the chapter does not give novel results, but serves as a partial survey. Second, we compare in a systematic way these rules in terms of inclusion relationships. Third, we consider a few key properties and identify those of the considered rules that satisfy them.

We follow earlier work in judgment aggregation (List and Puppe, 2004) in using a constraint-based version of judgment aggregation to represent properties like transitivity of preferences. As it is common in voting theory, we consider irresolute rules (also called correspondences) rather than functions, that is, a rule outputs a non-empty set of collective judgments.

The outline of the chapter is as follows. The general definitions are given in Section 4.1. In Section 4.2 we review the rules we study in the chapter. Majority preservation is a key property of rules, as it generalizes Condorcet-consistency. We focus on majority-preservation in Section 4.3

and show which of the rules defined in Section 4.2 satisfy it. In Section 4.4 we address inclusion and non-inclusion relationships between our rules. In Section 4.5 we study the rules from the point of view of important properties.

4.1 Formal setting

We use the standard propositional language $\mathscr{L}$ together with a standard notion of logical consistency. We denote atomic propositions by p,q etc. and formulas by α,β etc.

An *agenda*[1] $\mathscr{A}$ is a finite set of propositions of the form $\{\varphi_1,\neg\varphi_1,\ldots,\varphi_m,\neg\varphi_m\}$, where for all i, $\varphi_i \in \mathscr{L}$ and φ_i is neither a tautology nor a contradiction, and is a non-negated formula, (i.e. it is not of the form $\neg\alpha$). We refer to a pair $(\varphi,\neg\varphi)$ as an *issue*. The *pre-agenda* $[\mathscr{A}]$ associated with $\mathscr{A}$ is $[\mathscr{A}] = \{\varphi_1,\ldots,\varphi_m\}$. We slightly abuse notation and write φ_i instead of $\neg\neg\varphi_i$ for $\varphi_i \in [\mathscr{A}]$.

An agenda is *endowed with a notion of consistency* which preserves logical consistency. Formally, $\mathscr{A}$ comes with a set of *($\mathscr{A}$-)consistent* judgment sets; an ($\mathscr{A}$-)consistent judgment set is logically consistent, but the converse does not necessarily hold. Without loss of generality, the agenda's consistency notion is defined as *logical consistency given some fixed formula*: a set of formulas S is consistent if $S\cup\{\gamma\}$ is logically consistent, where γ is some exogenously fixed non-contradictory formula, which we call the *integrity constraint*. When γ is not specified, by default it is equal to $\top$, in which case the notion of consistency associated with the agenda coincides with standard logical consistency.

A *judgment* on $\varphi \in [\mathscr{A}]$ is either φ or $\neg\varphi$. A *judgment set* J for $\mathscr{A}$ is a subset of $\mathscr{A}$. J is *complete* if and only if for each $\varphi \in [\mathscr{A}]$, either $\varphi \in J$ or $\neg\varphi \in J$. A judgment set for $\mathscr{A}$ is *rational* if it is complete and consistent. Let $\mathscr{J}_{\mathscr{A}}$ be the set of all rational judgment sets for $\mathscr{A}$. For every consistent $S \subseteq \mathscr{A}$, the set of rational extensions of S, i.e. $\{J \mid J \in \mathscr{J}_{\mathscr{A}} \text{ and } S \subseteq J\}$, is denoted as $ext(S)$.

A $\mathscr{J}_{\mathscr{A}}$*-profile*, or simply a *profile*, is a finite sequence of rational individual judgment sets, i.e. $P = \langle J_1,\ldots,J_n\rangle$ for some n, where J_i is the judgment set of voter i. We slightly abuse notation and write $J \in P$ when $J = J_i$ for some i, and we write $|P|$ to denote the number of judgment sets in P. We sometimes denote P as (J_i,J_{-i}), where $J_{-i} = \langle J_j, 1\le j\le n, j\neq i\rangle$.

Given two rational judgment sets J and J' we define the *Hamming distance* d_H: $d_H(J,J')$ as the number of issues on which J and J' disagree. We also define the Hamming distance between two profiles $P = \langle J_1,\ldots,J_n\rangle$ and $P' = \langle J'_1,\ldots,J'_n\rangle$ as $D_H(P,P') = \sum_{i=1}^n d_H(J_i,J'_i)$, and between a judgment set and a profile as $d_H(J,P) = \sum_{i=1}^n d_H(J,J_i)$.

We define $N(P,\varphi)$ as the number of all voters in P whose judgment set contains φ, i.e.

$$N(P,\varphi) = |\{i \mid \varphi \in J_i, J_i \in P\}|$$

The *majoritarian judgment set* associated with profile $P = \langle J_1,\ldots,J_n\rangle$ contains all the elements of the agenda that are supported by a strict majority of judgment sets in P, i.e. $m(P) = \left\{\varphi \in \mathscr{A} \mid N(P,\varphi) > \frac{n}{2}\right\}$. A profile P is *majority-consistent* when $m(P)$ is a consistent subset of $\mathscr{A}$.

An (irresolute) *judgment aggregation rule* F maps every profile P, defined on every agenda $\mathscr{A}$, to a nonempty set of rational judgment sets in $\mathscr{J}_{\mathscr{A}}$. When for all profiles P, $F(P)$ is a singleton, then F is said to be *resolute*. Like in voting theory, resolute rules can be defined from irresolute ones by coupling them with a tie-breaking mechanism.

We also studied the link between judgment aggregation rules presented in this chapter and voting rules. For brevity reasons, we do not include that discussion in the present habilitation. The interested reader is referred to the corresponding publication (Lang et al., 2017).

[1] In this chapter, we use the notation $\mathscr{A}$ for an agenda. We used the same notation for the set of arguments. However, there is no danger of confusion, since in this chapter we do not study argumentation.

4.2 Judgment aggregation rules

We now define five (overlapping) families of judgment aggregation rules.

4.2.1 Rules based on the majoritarian judgment set

A judgment aggregation rule F is *based on the majoritarian judgment set* when for every two $\mathcal{J}_{\mathcal{A}}$-profiles P and P' such that $m(P) = m(P')$, we have $F(P) = F(P')$. These rules can be viewed as the judgment aggregation counterparts of voting rules based on the pairwise majority graph.

Recall that we denote the set of maximal for set inclusion consistent subsets of S by $\mathtt{MC}(S)$. Let us denote the set of maximal for cardinality consistent subsets of S by $\mathtt{MCC}(S)$. We will now use notations MC and MCC to denote two judgment aggregation rules we define in the next definition. We prefer to use this notation as it is already present in the literature (Lang et al., 2017). We strongly believe that there is no danger of confusion.

Definition 4.2.1 — Maximal Condorcet and maxcard Condorcet rules. For every $\mathcal{J}_{\mathcal{A}}$-profile P, the maximum Condorcet rule (MC) and the maxcard Condorcet rule (MCC) are defined as follows:

$$\textsc{mc}(P) = \{ext(S) \mid S \in \mathtt{MC}(m(P))\}, \tag{4.1}$$

$$\textsc{mcc}(P) = \{ext(S) \mid S \in \mathtt{MCC}(m(P))\}. \tag{4.2}$$

Clearly, $\textsc{mcc}(P) \subseteq \textsc{mc}(P)$. The output of the rule MC is called *Condorcet admissible set* by Nehring et al. (2014). The rule MCC is called *Slater* rule (Nehring et al., 2014), and $\textsc{Endpoint}_{d_H}$ (Miller and Osherson, 2009).

4.2.2 Rules based on the weighted majoritarian set

The *weighted majoritarian set* associated with a profile P is the function $N(P,.)$. A judgment aggregation rule F is *based on the weighted majoritarian set* when for every two $\mathcal{J}_{\mathcal{A}}$-profiles P and P', if for every $\varphi \in \mathcal{A}$ we have $N(P,\varphi) = N(P',\varphi)$, then $F(P) = F(P')$. These rules can be viewed as the judgment aggregation counterparts of voting rules that are based on the weighted pairwise majority graph Since $m(P)$ can be recovered from $N(P,.)$, every rule based on the majoritarian judgment set is also based on the weighted majoritarian set.

Definition 4.2.2 — Median rule. For every $\mathcal{J}_{\mathcal{A}}$-profile P, the median rule (MED) is defined as follows:

$$\textsc{med}(P) = \arg\max_{J \in \mathcal{J}_{\mathcal{A}}} \sum_{\varphi \in J} N(P,\varphi). \tag{4.3}$$

This rule appears in many places under different names: $\textsc{Prototype}$ (Miller and Osherson, 2009), *median rule* (Nehring et al., 2014), *maximum weighted agenda rule* (Lang et al., 2011), *simple scoring rule* (Dietrich, 2014) and *distance-based procedure* (Endriss et al., 2012). Variants of this rule have been defined by Konieczny and Perez (2002) and Pigozzi (2006). For completeness we give here the equivalent distance-based formulation of MED, although we consider more generally the family of distance-based rules in Section 4.2.4. For every $\mathcal{J}_{\mathcal{A}}$-profile P, the distance-based rule $F^{d_H,\Sigma}$ is defined as follows:

$$F^{d_H,\Sigma}(P) = \arg\min_{J \in \mathcal{J}_{\mathcal{A}}} \sum_{J_i \in P} d_H(J_i, J). \tag{4.4}$$

It is not difficult to establish that $F^{d_H,\Sigma}$ coincides with MED (Lang et al., 2011; Dietrich, 2014). The following rule generalizes the *ranked pairs* voting rule (Tideman, 1987). It proceeds by considering the elements φ of the agenda in non-increasing order of $N(P,\varphi)$ and fixing each agenda issue value to the majoritarian value if it does not lead to an inconsistency.

Definition 4.2.3 — Ranked agenda rule. Let $\mathscr{A} = \{\psi_1,\dots,\psi_{2m}\}$. For every $\mathscr{J}_{\mathscr{A}}$-profile P, RA consists of those judgment sets $J \in \mathscr{J}_{\mathscr{A}}$ for which there exists a permutation $(\varphi_1,\varphi_2,\dots,\varphi_{2m})$ of the propositions in $\mathscr{A}$ such that $N(P,\varphi_1) \geq N(P,\varphi_2) \geq \cdots \geq N(P,\varphi_{2m})$ and J is obtained by the following algorithmic procedure:

$S := \emptyset$
for $k = 1,\dots,2m$ do
  if $S \cup \{\varphi_k\}$ is consistent **then** $S \leftarrow S \cup \{\varphi_k\}$
  end if
end for
$J := S$

In plain words, RA assigns iteratively a truth value to each proposition of the agenda, whenever it does not produce an inconsistency with propositions already assigned, following an order compatible with $N(P,.)$. An equivalent non-procedural definition is the following: for every profile P, define $>^{RA}_P$ by: $J >^{RA}_P J'$ if there is an $\alpha \in \mathbb{N}$ such that

1. for all $\psi \in \mathscr{A}$, $N(P,\psi) > \alpha$ implies [$\psi \in J$ if and only if $\psi \in J'$], and
2. $J \cap \{\varphi \mid N(P,\varphi) = \alpha\} \supset J' \cap \{\varphi \mid N(P,\varphi) = \alpha\}$.

Then $\text{RA}(P) = \{J \in \mathscr{J}_{\mathscr{A}} \mid J \text{ undominated in } >^{RA}_P\}$.[2]

We skip the LEXIMAX rule (Everaere et al., 2014), which is a refinement of RA. The interested reader is referred to the literature (Lang et al., 2017).

4.2.3 Rules based on elementary changes in profiles

The next family of rules we consider contains rules that are based on minimal set of changes on a profile needed to render the profile majority-consistent. This family of judgment rules can be viewed as the judgment aggregation counterpart of voting rules that are rationalisable by some distance with respect to the Condorcet consensus class (Elkind et al., 2009).

The first rule we consider is called the *Young* rule for judgment aggregation, by analogy with the Young voting rule, which outputs the candidate c minimising the number of voters to remove from the profile so that c becomes a weak Condorcet winner Young (1977). The judgment aggregation generalization consists of removing a minimal number of voters so that the profile becomes majority-consistent, or equivalently, to look for majority-consistent subprofiles of maximum cardinality.

Definition 4.2.4 — Young rule. For every $\mathscr{J}_{\mathscr{A}}$-profile P,

$$\text{Y}(P) = \{ext(m(Q)) \mid Q \in \underset{Q\subseteq P,\ m(Q) \text{ is consistent}}{\arg\max} |Q|\}. \tag{4.5}$$

The next rule we define looks for a minimal number of individual *judgment reversals* in the profile so that P becomes majority-consistent, where a judgment reversal is a change of truth value of one agenda element in one individual judgment set. This rule has been proposed first by Miller and Osherson (2009) under the name FULL_d.

Definition 4.2.5 — Minimal profile change rule. For $P \in \mathscr{J}^n_{\mathscr{A}}$, the MPC rule is defined as:

[2] The proof—almost straightforward—can be found in Lang (2015).

$$\text{MPC}(P) = \{ext(m(Q)) \mid Q \in \underset{Q \in \mathscr{J}_{\mathscr{A}}^n,\ m(Q) \text{ is } \mathscr{A}\text{-consistent}}{\arg\min} D_H(P,Q)\}.$$

4.2.4 Rules based on (pseudo-)distances

For a given constrained agenda, a pseudo-distance d on $\mathscr{J}_{\mathscr{A}}$ is a function that maps pairs of judgment sets to non-negative real numbers, and that satisfies, for all $J, J' \in \mathscr{J}_{\mathscr{A}}$, $d(J,J') = d(J',J)$, and $d(J,J') = 0$ if and only if $J = J'$.

Two pseudo-distances we will use are the Hamming distance d_H, defined in Section 4.1, and the *geodesic distance*[3] on $\mathscr{J}_{\mathscr{A}}$, defined by Duddy and Piggins (2012) as follows. Given three distinct rational judgment sets J, J', J'', we say that J *is between* J' and J'' if $J' \cap J'' \subset J$. Let $G_{\mathscr{A}}$ be the graph whose set of vertices is the set of rational judgment sets $\mathscr{J}_{\mathscr{A}}$ and that contains an edge between J' and J'' if and only if there exists no $J \in \mathscr{J}_{\mathscr{A}}$, $J' \neq J \neq J''$, between J' and J''. Finally, $d_G(J',J'')$ is defined as the length of the shortest path between J' and J'' in $G_{\mathscr{A}}$.

Definition 4.2.6 Let d be a pseudo-distance on $\mathscr{J}_{\mathscr{A}}$ and $\star$ a commutative, associative and non-decreasing function on $\mathbb{R}^+$. The distance-based judgment aggregation rule $F^{d,\star}$ associated with d and $\star$ is defined as

$$F^{d,\star}(P) = \underset{J \in \mathscr{J}_{\mathscr{A}}}{\arg\min} \star (d(J_1,J),\ldots,d(J_n,J)) \tag{4.6}$$

In addition to $F^{d_H,\Sigma}$ we focus on two specific distance-based judgment aggregation rules: $F^{d_G,\Sigma}$ (Duddy and Piggins, 2012), and $F^{d_H,\text{MAX}}$ (Konieczny and Perez, 2002; Lang et al., 2011). From now on, we will use the word "distance" instead of "pseudo-distance" although our rules can be defined more generally for pseudo-distances.

4.2.5 Scoring rules

Dietrich (2014) defines a general class of *scoring rules* for judgment aggregation. Given a function $s : \mathscr{J}_{\mathscr{A}} \times \mathscr{A} \to \mathbb{R}^+$, the rule F_s is defined as

$$F_s(P) = \underset{J \in \mathscr{J}_{\mathscr{A}}}{\arg\max} \sum_{\varphi \in J} \sum_{J_i \in P} s(J_i, \varphi). \tag{4.7}$$

The MED rule (4.3) is a scoring rule (and also a distance-based rule). The reversal score function rev Dietrich (2014) is defined as:

$$\text{rev}(J,\varphi) = min_{J' \in \mathscr{J}_{\mathscr{A}},\ \varphi \notin J'} d_H(J,J'). \tag{4.8}$$

The main motivation for introducing this rule is that it generalizes the Borda voting rule.

4.3 Majority-preservation

Intuitively, a judgment aggregation rule F is majority-preserving if and only if F returns only the extensions of the majoritarian judgment set whenever it is consistent. In case of ties, a majoritarian set can have more than one extension. For example, when we have agenda $\mathscr{A} = \{p, \neg p, q, \neg q\}$ and individual judgments $J_1 = \{p,q\}$ and $J_2 = \{p, \neg q\}$, then $m(\langle J_1, J_2 \rangle) = \{p\}$, which can be extended into two complete collective judgment sets, namely $\{p, \neg q\}$ and $\{p,q\}$.

[3] Our name; no name was given of this distance by Duddy and Piggins (2012).

Definition 4.3.1 A judgment aggregation rule F is *majority-preserving* if and only if for every agenda $\mathscr{A}$ and for every majority-consistent $\mathscr{J}_{\mathscr{A}}$-profile P we have $F(P) = ext(m(P))$. A rule F is *weakly majority-preserving* if and only if for every agenda $\mathscr{A}$ and for every majority-consistent $\mathscr{J}_{\mathscr{A}}$-profile P we have $F(P) \supseteq ext(m(P))$.

Proposition 4.3.1 MC, MCC, MED, RA, LEXIMAX, Y and MPC are majority-preserving. $F^{d_G,\Sigma}$, F_{rev} and $F^{d_H,\text{MAX}}$ are not even weakly majority-preserving.

4.4 Inclusion relationships between the rules

We now establish the following (non)inclusion relationships between the rules.

Definition 4.4.1 Given two judgment aggregation rules F_1 and F_2, we denote:

- $F_1 \subseteq F_2$ when $F_1(P) \subseteq F_2(P)$ holds for every agenda $\mathscr{A}$ and every $\mathscr{J}_{\mathscr{A}}$-profile P.
- $F_1 \subset F_2$ when $F_1 \subseteq F_2$ and $F_1 \neq F_2$.
- F_1 $\texttt{inc}$ F_2 when neither $F_1 \subseteq F_2$ nor $F_2 \subseteq F_1$.

do

Proposition 4.4.1 The inclusion and incomparability relations among the majority-preserving rules, and among the non majority-preserving rules, are represented on Tables 4.1 and 4.2; a $\supset$ sign for row F_1 and column F_2 means that $F_1 \supset F_2$, and an $\texttt{inc}$ sign, that F_1 $\texttt{inc}$ F_2.

	MCC	MED	RA	Y	MPC
MC	$\supset$	$\supset$	$\supset$	inc	inc
MCC		inc	inc	inc	inc
MED			inc	inc	inc
RA				inc	inc
Y					inc

Table 4.1: (Non)inclusion relationships between the majority-preserving rules.

	$F^{d_G,\Sigma}$	F_{rev}
$F^{d_H,\text{MAX}}$	inc	inc
$F^{d_G,\Sigma}$		inc

Table 4.2: (Non)inclusion relationships between the other rules.

Note that for every $F_1 \in \{F^{d_G,\Sigma}, F_{\text{rev}}, F^{d_H,\text{MAX}}\}$ and for every $F_2 \in \{\text{MC}, \text{MCC}, \text{MED}, \text{RA}, \text{Y}, \text{MPC}\}$, it holds that F_1 $\texttt{inc}$ F_2.

4.5 Properties

In preference aggregation, there are three classes of properties (Zwicker, 2014): those that are satisfied by most common rules (such as neutrality or anonymity); those that are very hard to satisfy, and whose satisfaction, under mild additional condition, implies impossibility results; and finally, those that are satisfied by a significant number of rules and violated by another significant number of rules. Similarly, in judgment aggregation, weak properties such as anonymity are satisfied by all our rules, while strong properties such as independence are violated by all our rules. We have already studied an "intermediate" property: *majority-preservation*. Here we consider five more: *unanimity*, *monotonicity*, *homogeneity*, *reinforcement* and *separability*.

4.5.1 Unanimity

Unanimity has been defined for *resolute rules* by Dietrich and List (2008): R is said to satisfy unanimity when for every $\mathcal{J}_{\mathcal{A}}$-profile $P = \langle J_1, \ldots, J_n \rangle$ and every $\varphi \in \mathcal{A}$, if $\varphi \in J_i$ for all $i \leq n$, then $\varphi \in R(P)$. We generalise unanimity to irresolute rules, which gives us a weak and a strong version of unanimity.

Definition 4.5.1 — Weak and strong unanimity. Given $\varphi \in \mathcal{A}$, the $\mathcal{J}_{\mathcal{A}}$-profile P is said to be φ-*unanimous* when $\varphi \in J_i$ for every $J_i \in P$.

- F satisfies *weak unanimity* when for every φ-unanimous profile P, there is a $J \in F(P)$ such that $\varphi \in J$.
- F satisfies *strong unanimity* when for every φ-unanimous profile P, for all $J \in F(P)$ we have $\varphi \in J$.

4.5.2 Monotonicity

In voting, monotonicity states that when the position of the winning alternative improves in some vote *ceteris paribus*, then this alternative remains the winner. We define below a generalisation of this property for (irresolute) judgment aggregation rules. It is a generalization of the monotonicity property defined by Dietrich and List (2007c) for resolute rules.

Definition 4.5.2 — Monotonicity.
Let $P, P' \in \mathcal{J}^n_{\mathcal{A}}$ be two profiles, and $\varphi \in \mathcal{A}$. P' is a φ-improvement of P when

- $P = (J_i, J_{-i})$
- $P' = (J'_i, J_{-i})$
- $\neg\varphi \in J_i$
- $\varphi \in J'_i$
- for all $\psi \in \mathcal{A}$, $\psi \notin \{\varphi, \neg\varphi\}$, $\psi \in J_i$ if and only if $\psi \in J'_i$.

Note that the definition implies that J'_i is consistent, otherwise P' would not be a well-defined profile. A judgment aggregation rule F is monotonic, when for every $P \in \mathcal{J}^n_{\mathcal{A}}$ and its φ-improvement $P' \in \mathcal{J}^n_{\mathcal{A}}$, for any $\varphi \in \mathcal{A}$, it holds that: if $\varphi \in J$ for every $J \in F(P)$, then $\varphi \in J'$ for every $J' \in F(P')$.

Note that not every profile has a φ-improvement for a given $\varphi \in \mathcal{A}$.

4.5.3 Reinforcement

A social preference function F satisfies *reinforcement* if whenever two profiles over disjoint electorates have some output rankings in common, then the profiles obtained by merging the two electorates leads to elect those rankings that are obtained for both profiles. This easily generalizes to judgment aggregation rules as follows.

Definition 4.5.3 For every two profiles $P = \langle J_1, \ldots, J_n \rangle$ and $Q = \langle J_{n+1}, \ldots, J_q \rangle$, we denote $P + Q = \langle J_1, \ldots, J_q \rangle$. We say that a judgment aggregation rule F satisfies *reinforcement* when for every agenda $\mathcal{A}$, and every two profiles P and Q over disjoint electorates, if $F(P) \cap F(Q) \neq \emptyset$ then $F(P + Q) = F(P) \cap F(Q)$.

4.5.4 Homogeneity

Homogeneity says that multiplying the same profile several times does not change the result of the aggregation. Let us write kP for $\underbrace{P + \cdots + P}_{k\ times}$, where $+$ has been defined in Subsection 4.5.3.

Definition 4.5.4 A judgment aggregation rule F satisfies homogeneity when for every $\mathcal{J}_{\mathcal{A}}$-profile P and positive integer k, it holds that $F(kP) = F(P)$.

4.5.5 Agenda separability

The *restriction* of $P = \langle J_1, \ldots, J_n\rangle$ over a sub-agenda $\mathscr{A}_1$ of $\mathscr{A}$ is defined as $P_{\downarrow \mathscr{A}_1} = \langle J_1 \cap \mathscr{A}_1, \ldots, J_n \cap \mathscr{A}_1\rangle$. A judgment aggregation rule F satisfies *independence of irrelevant alternatives* (IIA) if for every two profiles $P, P' \in \mathscr{J}^n_{\mathscr{A}}$, and every $\varphi \in \mathscr{A}$, if $P_{\downarrow\{\varphi,\neg\varphi\}} = P'_{\downarrow\{\varphi,\neg\varphi\}}$, then $\varphi \in F(P)$ iff $\varphi \in F(P')$.

IIA is a very strong requirement, since together with three seemingly innocuous properties, namely universal domain (F is defined for every profile), unanimity principle, and collective rationality (F outputs complete and consistent judgment sets), it implies dictatorship (Dietrich and List, 2007a). In this subsection, we define a weakening of IIA, which we call agenda separability.

Following the idea that only judgments on logically related issues should influence the collective judgment on each issue, we define agenda separability as the property requiring that when two agendas can be split into sub-agendas that are independent from each other, the output judgment sets can be obtained by first applying the rule on each sub-agenda separately and then taking the pairwise unions of judgment sets from the two resulting sets.

A partition $\{\mathscr{A}_1, \mathscr{A}_2\}$ of $\mathscr{A}$ is an *independent partition of* $\mathscr{A}$ if for every $J^1 \in \mathscr{J}_{\mathscr{A}_1}$ and $J^2 \in \mathscr{J}_{\mathscr{A}_2}$, $J^1 \cup J^2$ is γ-consistent.[4]

Definition 4.5.5 — Agenda separability. We say that rule F satisfies *agenda separability* if for every agenda $\mathscr{A}$, every independent partition $\{\mathscr{A}_1, \mathscr{A}_2\}$ of $\mathscr{A}$, and all profiles $P \in \mathscr{J}^n_{\mathscr{A}}$, we have

$$F(P) = \{J^1 \cup J^2 \mid J^1 \in F(P_{\downarrow\mathscr{A}_1}) \text{ and } J^2 \in F(P_{\downarrow\mathscr{A}_2})\}.$$

If F is a resolute rule, then the last line of the definition simplifies into $F(P) = F(P_{\downarrow\mathscr{A}_1}) \cup F(P_{\downarrow\mathscr{A}_2})$.

Also, by associativity of $\cup$, this notion generalises to agendas that can be partitioned into a collection $\{\mathscr{A}_1, \ldots, \mathscr{A}_k\}$ such that for every $J_1 \in \mathscr{J}_{\mathscr{A}_1}, \ldots, J_k \in \mathscr{J}_{\mathscr{A}_k}$, $J_1 \cup \ldots \cup J_k$ is consistent. In that case,

$$F(P) = \left\{ \bigcup_{i=1}^{k} J^i \;\middle|\; J^1 \in F(P_{\downarrow\mathscr{A}_1}), \ldots, J^k \in F(P_{\downarrow\mathscr{A}_k}) \right\}.$$

Independence of irrelevant alternatives is defined for resolute rules only. We show that agenda separability restricted to resolute rules is a weakening of IIA.

Property	MC	MCC	RA	MED	MPC	Y	$F^{d_G,\Sigma}$	$F^{d_H,\mathrm{MAX}}$	F_{rev}
Majority preservation	yes	yes	yes	yes	yes	yes	no	no	no
Weak unanimity	yes	no	yes	no	no	yes	no	no	no
Strong unanimity	no	no	yes	no	no	yes	no	no	no
Monotonicity	yes	yes	yes	yes	no	yes	yes	yes	no
Homogeneity	yes	yes	yes	yes	no	no	yes	yes	yes
Reinforcement	no	no	no	yes	no	no	yes	no	yes
Agenda separability	yes	yes	yes	yes	yes	no	yes	no	yes

Table 4.3: Summary of rules and properties they (do not) satisfy.

Proposition 4.5.1 Any resolute judgment aggregation rule that satisfies IIA is agenda separable.

[4] A stronger notion of independence, which makes sense only when $\gamma = \top$, is *syntactical agenda independence*: a partition $\{\mathscr{A}_1, \mathscr{A}_2\}$ of $\mathscr{A}$ is *syntactically independent* if the sets of propositional variables appearing in $\mathscr{A}_1$ and $\mathscr{A}_2$ are disjoint. Clearly, syntactical agenda independence implies agenda independence. Note that the implication is strict: as a counter-example, consider $\mathscr{A} = \{x, \neg x, x \leftrightarrow y, \neg(x \leftrightarrow y)\} = \mathscr{A}_1 \cup \mathscr{A}_2$, $\gamma = \top$, $\mathscr{A}_1 = \{x, \neg x\}$ and $\mathscr{A}_2 = \{x \leftrightarrow y, \neg(x \leftrightarrow y)\}$.

We omit overlapping agenda separability, which is a stricter property. The interested reader is referred to the original publication (Lang et al., 2016).

4.6 Summary

We have listed a number of existing judgment aggregation rules, and for a number of important properties we have identified those rules that satisfy it.

We can note that RA and MED perform particularly well, in the sense that each of them satisfies a maximal for set inclusion set of properties.

5. Future work and open questions

This chapter concludes the habilitation by summarising our contributions, presenting two applications that we plan to work on in the near future and presenting some open problems.

5.1 Summary

We presented the contributions in different areas but they are all centred around reasoning under inconsistency. The methodological approach is based on defining principles and studying / evaluating systems based on the principles they satisfy. We presented our contributions in the area of argumentation, namely in defining the principles for both extension-based and ranking-based semantics. We also studied the link between the principles.

Another part of our study was to investigate the links between argumentation-based and non argumentation-based approaches. Namely, we studied the link between the reasoning from maxi-consistent subsets and the reasoning based on argumentation and the links between argumentation and OBDA semantics. We also considered the links between extension-based and ranking-based semantics.

We then studied how to decrease the number of extensions or repairs in order to allow for more conclusions while staying consistent. We presented a new type of inference, called supported inference, as well as other ways to decrease the number of argumentation extensions. We also studied the corresponding problem when reasoning from an inconsistent knowledge base. Finally, we considered the question how to reduce the number of repairs in OBDA.

The last part of our contribution was the study of different rules for judgment aggregation. We proposed new principles and studied whether existing rules satisfy those principles.

5.2 Future work

We now consider perspectives for future research. For brevity and coherence reasons, we do not mention all the possibilities in all the areas of research cited in this habilitation. Instead, we concentrate on two applications, which call for both theoretical and practical developments: e-democracy and health.

5.2.1 Applications in e-democracy

E-democracy is a form of government that allows everybody to participate in the development of laws. It has numerous benefits since it strengthens the integration of citizens in the political debate. Several on-line platforms exist; most of them propose to represent a debate in the form of a graph, which allows humans to better grasp the arguments and their relations. However, once the arguments are entered in the system, little or no automatic treatment is done by such platforms. Given the development of online consultations, it is clear that in the near future we can expect thousands of arguments on some hot topics, which will make the manual analysis difficult and time-consuming. Our idea is to use computational argumentation theory in order to detect the most important arguments, estimate the acceptability degrees of arguments and predict the decision that will be taken.

There are numerous platforms for online argument-based discussions, like idebate's debatebase (https://idebate.org/debatabase) and Debategraph (http://debategraph.org), which was used by the White House and CNN. Particularly, considering e-democracy, the first law in France that was preceded by an online discussion was la loi numérique in 2015 (https://www.republique-numerique.fr). The consulting contains more than 8.000 contributions and more than 140.000 votes. All those platforms have one thing in common: there is no automatic reasoning, i.e. the data is not exploited in any way. Our goal is to allow for automatic reasoning and to exploit the data present in those platforms.

There are several things that can be done automatically. First, we can identify key arguments. This can be done by, for example, looking at the topology of the graph, the number of attackers and supporters, the links between the arguments and propositions (options) etc. Second, we can automatically determine whether a given proposition is going to be accepted or rejected even without looking at the votes of the users. For example, if the topic argument is attacked by three arguments, if one of them is not attacked, and another is defended against all of its attackers, we can conclude that it will be rejected. System can automatically do many other things, even without relying on NLP techniques. For instance, we can automatically identify two unconnected parts of the graph and suggest that there are two different lines of attack of an argument / idea.

This automatically generated list of the most influential arguments can help a new user joining the discussion to more easily grasp its current state (e.g. key arguments, main lines of critique). It is also very useful to policy makers who want to quickly understand the public opinion, especially in the case when the number of arguments is elevated.

There are many challenges in order to apply argumentation in e-democracy. New theoretical developments are necessary. For instance, there is still no semantics that can deal with arguments, attacks, supports, positive and negative votes on arguments. Some semantics can treat attacks and supports but not votes. Virtually all bipolar semantics are defined on acyclic graphs only. We plan to develop semantics that can deal with all the necessary components of an argumentation system for e-democracy.

We envisage to conduct a theoretical evaluation of the developed framework. We will define the principles that need to be satisfied. For example, take an argument X whose attacker Y receives a negative vote; in this case, the score of argument X might increase, but should not decrease. The goal of the principle-based approach is to offer explainability to the user so that the system is not seen as a black box. The rationality postulates from the literature allow evaluating argumentation frameworks without votes. We will define rationality postulates for frameworks with votes and prove that they are satisfied by our framework. We also plan to implement the platform and test it.

5.2.2 Applications in health

There are numerous examples of applying argumentation in health care (Longo, 2016; Atkinson et al., 2017; Cyras et al., 2018; Kökciyan et al., 2018). In what follows, we discuss some applications of argumentation and argument-based negotiation in this domain that we envisage to conduct in the near future.

We consider using argumentation as a way to communicate between patients, their families, as well as different practitioners. The idea is that all the practitioners do not gather all the time and there might be inconsistencies between the treatments and advice given by, say an ophthalmologist (telling a patient to stop taking a certain medication three days before an eye surgery) and a cardiologist (recommending the patient not to stop taking that medication). Some data may be in raw format (e.g. patient's heart rate throughout the day and night) and sharing it with all the actors (doctors, families, institutions) is not a solution due to its format and to the privacy issues.

Our goal is to encapsulate some information in argument format and share it by the means of an electronic medical record. For example, we might send an argument to some actors telling them that around 2 pm, there was a considerable increase in patient's heart rate without revealing all the data to everybody. Other arguments may be shared with other actors as well as some links (attacks and supports) between them.

Our hypothesis is that most of the actors have a lot of common goals, namely the health of the patient, so we strongly believe that exchanging arguments can resolve most of the disagreements that arise between them (like in the previous example).

We identified several scientific research tasks that must be done. First, the existing framework has to be applied in the medical domain. In order to translate the arguments from English to the formal language used by the system and vice versa, we need to develop a way to link argumentation theory with the corresponding ontology. Second, the role and different types of supports between arguments has to be taken into account. In this context, there might be a lot of deductive supports between arguments, but some other types of interactions may appear. Third, it might be the case that all the arguments are pairwise acceptable, but that some groups of three or more arguments are contradictory. To model this real-life situation, we need to use argumentation semantics that support n-ary attacks, also known as sets of attacking arguments.

5.3 Open problems

We now comment on some open problems and challenges in argumentation.

5.3.1 Links between approaches

In many areas, all the researchers use the same format of the input. They propose different algorithms, rules, etc. and study their properties. They compare their contributions more or less easily thanks to the common framework. In argumentation, different approaches have different formats and different ways to model the same situation (e.g. the same dialogue between three journalists). This might be an expected consequence of a vast domain like argumentation, which is multidisciplinary in its essence, and has links with linguistics, philosophy, psychology, law, etc. However, this opens numerous questions, like: given many different approaches (Simari and Loui, 1992; Besnard and Hunter, 2001; Amgoud and Cayrol, 2002b; Modgil and Prakken, 2013; Toni, 2014; Cyras and Toni, 2016; Baroni et al., 2018b), what is the link between them? What is the link between ranking-based approaches and probabilistic approaches (Dung and Thang, 2010; Li et al., 2011; Thimm, 2012; Hunter, 2013; Hunter and Thimm, 2017; Polberg and Hunter, 2018; Hunter et al., 2020; Hunter, 2020)? What is the link between ADF (Brewka et al., 2017, 2018) and bipolar argumentation frameworks? What are the links between the intuitions underneath different types of relations in bipolar argumentation (Boella et al., 2010; Nouioua and Risch, 2011; Rago et al., 2016; Polberg and Hunter, 2018; Doder et al., 2020)? Let us note that some of those links have been studied by certain researchers, but many of them have only been tackled in an informal way, through comments or examples.

5.3.2 Benchmarks

Computational argumentation is positioned as part of artificial intelligence and its practical significance is often underlined. We hereby claim that more effort is needed in order to produce real-life benchmarks. There are several types of examples in argumentation. First, we have toy examples introduced by the researchers with a need of illustrating their systems. Second, we have formalizations of real dialogues, e.g. the work by Cabrio and Villata (2014); however, such benchmarks are very rare and often small in size. Third, there are automatically extracted graphs of arguments, but given the precision of state of the art algorithms, they are not reliable since they are error prone. The fourth group is the group of benchmarks generated from more or less random data. Those can be useful to compare existing solvers, but do not provide any insights at how real argumentation graphs look like.

5.3.3 Modelling arguments and attacks

One of the most important applications of argumentation theory is to model human reasoning. A given text in natural language, say English, can be translated to an argumentation graph. The question whether an argument attacks another one is still open for many situations. Let us illustrate this on an example from by Cramer and Guillaume (2018). Let argument A be: *Louis applied the brake. Therefore, the car slowed down.* Let argument B be: *Louis applied the accelerator instead of the brake. Therefore, Louis did not apply the brake.* Pretty much everyone agrees that B attacks A. But whether A attacks B is an object of heated discussions. The answer depends on the philosophical school, mathematical modelization, feeling or previous knowledge of the annotator, and on other factors. Different formalisms allow for different types of attacks and there are several differences, e.g. can a sub-argument (or a hypothesis) of an argument rebut another argument? We think that the question of comparing different types of attacks (and related preferences) did not receive enough attention. For example, is the more specific argument (e.g. penguins do not fly) always stronger than a general argument (e.g. birds fly)? Does this mean that there is a unidirectional attack; or is there a symmetric attack accompanied by a strict preference? The similar question can be asked by replacing the concept of *specificity* with *recency*. Understanding real arguments and attacks between them can underline important modelling questions that would stay unexplored otherwise.

5.3.4 How intuitive are our models?

Let us finish this habilitation with, in our opinion, the most important open question. It is often stated that argumentation is an intuitive way to represent data since this format is easier for humans to grasp than a formalization based on numbers or logical formulas. We think that this claim must be considered and examined in much more detail, both theoretically and practically, namely through experiments with humans. The result obtained by examining this question should guide us in defining and developing the future systems.

6. Selected publications

Here we provide a list of selected papers closely related to the topics presented in this habilitation.

Argumentation

- Leendert van der Torre, Srdjan Vesic. The Principle-Based Approach to Abstract Argumentation Semantics. Book chapter. In Handbook of Formal Argumentation. Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, Leendert van der Torre (Eds.), College Publications, 2018.
  `http://www.cril.univ-artois.fr/~vesic/2018_HOFA.pdf`
- Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, Srdjan Vesic. Acceptability semantics for weighted argumentation frameworks. In 26th International Joint Conference on Artificial Intelligence, IJCAI 2017.
  `http://www.cril.univ-artois.fr/~vesic/2017_IJCAI_WAF.pdf`
- Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, Srdjan Vesic. Ranking Arguments With Compensation-Based Semantics. In 15th International Conference on Principles of Knowledge Representation and Reasoning, KR 2016.
  `http://www.cril.univ-artois.fr/~vesic/2016_KR.pdf`
- Srdjan Vesic. Identifying the Class of Maxi-Consistent Operators in Argumentation. Journal of Artificial Intelligence Research, 2013.
  `http://www.cril.univ-artois.fr/~vesic/2013_JAIR.pdf`

Reducing the number of extensions

- Sébastien Konieczny, Pierre Marquis, Srdjan Vesic. On Supported Inference and Extension Selection in Abstract Argumentation Frameworks. In 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2015.
  `http://www.cril.univ-artois.fr/~vesic/2015_ECSQARU.pdf`

- Sébastien Konieczny, Pierre Marquis, Srdjan Vesic. Rational Inference Relations from Maximal Consistent Subsets Selection. In 28th International Joint Conference on Artificial Intelligence, IJCAI 2019.
  http://www.cril.univ-artois.fr/~vesic/2019_IJCAI.pdf
- Bruno Yun, Srdjan Vesic, Madalina Croitoru, Pierre Bisquert. Inconsistency Measures for Repair Semantics in OBDA. In 27th International Joint Conference on Artificial Intelligence, IJCAI 2018, (contr).
  http://www.cril.univ-artois.fr/~vesic/2020_IJCAI.pdf

JUDGMENT AGGREGATION

- Jérôme Lang, Marija Slavkovik, Srdjan Vesic. Agenda Separability in Judgment Aggregation. In 13th AAAI Conference on Artificial Intelligence, AAAI 2016.
  http://www.cril.univ-artois.fr/~vesic/2016_AAAI.pdf
- Jérôme Lang, Gabriella Pigozzi, Marija Slavkovik, Leendert van der Torre, Srdjan Vesic. A partial taxonomy of judgment aggregation rules and their properties. Social Choice and Welfare, 2017.
  http://www.cril.univ-artois.fr/~vesic/2017_SCW.pdf

7. Bibliography and index

Bibliography

Carlos Alchourrón, Peter Gardenfors, and David Makinson. On the logic of theory change : Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

Leila Amgoud and Jonathan Ben-Naim. Ranking-based semantics for argumentation frameworks. In *International Conference on Scalable Uncertainty Management, SUM 2013*, pages 134–147, 2013.

Leila Amgoud and Jonathan Ben-Naim. Axiomatic foundations of acceptability semantics. In *International Conference on Principles of Knowledge Representation and Reasoning, KR 2016*, pages 2–11, 2016.

Leila Amgoud and Philippe Besnard. Bridging the gap between abstract argumentation systems and logic. In *International Conference on Scalable Uncertainty Management SUM 2009*, pages 12–27, 2009.

Leila Amgoud and Philippe Besnard. A formal analysis of logic-based argumentation systems. In *International Conference on Scalable Uncertainty Management, SUM 2010*, pages 42–55, 2010.

Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002a.

Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002b.

Leila Amgoud and Srdjan Vesic. Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning*, 55:585–606, 2014.

Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, and Srdjan Vesic. Ranking arguments with compensation-based semantics. In *International Conference on Principles of Knowledge Representation and Reasoning, KR 2016*, pages 12–21, 2016.

Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, and Srdjan Vesic. Acceptability semantics for weighted argumentation frameworks. In *International Joint Conference on Artificial Intelligence, IJCAI 2017*, pages 56–62, 2017a.

Leila Amgoud, Jonathan Ben-Naim, and Srdjan Vesic. Measuring the intensity of attacks in argumentation graphs with shapley value. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI 2017*, pages 63–69. AAAI Press, 2017b.

Leila Amgoud, Elise Bonzon, Marco Correia, Jorge Cruz, Jérôme Delobelle, Sébastien Konieczny, João Leite, Alexis Martin, Nicolas Maudet, and Srdjan Vesic. A note on the uniqueness of models in social abstract argumentation. *CoRR*, abs/1705.03381, 2017c. URL `https://arxiv.org/abs/1705.03381`.

Ofer Arieli. Conflict-free and conflict-tolerant semantics for constrained argumentation frameworks. *Journal of Applied Logic*, 13(4):582–604, 2015.

Kenneth J. Arrow. *Social Choice and Individual Values*. Yale University Press, 1951.

Kenneth J. Arrow, Amartya Sen, and Kotaro Suzumura, editors. *Handbook of Social Choice and Welfare*. North-Holland, 2002.

Katie Atkinson, Pietro Baroni, Massimiliano Giacomin, Anthony Hunter, Henry Prakken, Chris Reed, Guillermo Ricardo Simari, Matthias Thimm, and Serena Villata. Towards artificial argumentation. *AI Mag.*, 38(3):25–36, 2017.

Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artif. Intell.*, 175(9-10):1620–1654, 2011.

Jean-François Baget, Salem Benferhat, Zied Bouraoui, Madalina Croitoru, Marie-Laure Mugnier, Odile Papini, Swan Rocher, and Karim Tabia. Inconsistency-Tolerant Query Answering: Rationality Properties and Computational Complexity Analysis. In *JELIA 2016*, pages 64–80, 2016.

Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007.

Pietro Baroni, Massimiliano Giacomin, and Giovanni Guida. Towards a formalization of skepticism in extension-based argumentation semantics. In *Proceedings of the 4th Workshop on Computational Models of Natural Argument (CMNA'04)*, pages 47–52, 2004.

Pietro Baroni, Massimiliano Giacomin, and Giovanni Guida. SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168:162–210, 2005.

Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *Knowledge Eng. Review*, 26(4):365–410, 2011a.

Pietro Baroni, Paul E. Dunne, and Massimiliano Giacomin. On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175(3-4): 791–813, 2011b.

Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49, 2015.

Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors. *Handbook of Formal Argumentation*. College Publications, 2018a.

Pietro Baroni, Massimiliano Giacomin, and Beishui Liao. A general semi-structured formalism for computational argumentation: Definition, properties, and examples of application. *Artif. Intell.*, 257:158–207, 2018b.

Pietro Baroni, Antonio Rago, and Francesca Toni. From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. *International Journal of Approximate Reasoning*, 105:252–286, 2019.

Trevor Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

Salem Benferhat, Claudette Cayrol, Didier Dubois, Jérôme Lang, and Henri Prade. Inconsistency management and prioritized syntax-based entailment. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence, IJCAI 1993*, pages 640–647, 1993a.

Salem Benferhat, Didier Dubois, and Henri Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Conference on Uncertainty in Artificial Intelligence UAI'93*, pages 411–419, 1993b.

Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2):203–235, 2001.

Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, 2008.

Meghyn Bienvenu. On the complexity of consistent query answering in the presence of simple ontologies. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

Guido Boella, Dov M. Gabbay, Leendert W. N. van der Torre, and Serena Villata. Support in abstract argumentation. In *Computational Models of Argument: Proceedings of COMMA 2010*, pages 111–122, 2010.

Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Conference on Artificial Intelligence AAAI 2016*, pages 914–920, 2016a.

Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 914–920, 2016b.

Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. Argumentation ranking semantics based on propagation. In *Computational Models of Argument - Proceedings of COMMA 2016*, pages 139–150, 2016c.

Steven J. Brams and Peter C. Fishburn. Voting procedures. In Amartya K. Sen Kenneth J. Arrow and Kotaro Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 1 of *Handbook of Social Choice and Welfare*, pages 173 – 236. Elsevier, 2002.

Gerhard Brewka, Stefan Ellmauthaler, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Abstract dialectical frameworks. an overview. *FLAP*, 4(8), 2017.

Gerhard Brewka, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Weighted abstract dialectical frameworks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 1779–1786, 2018.

Elena Cabrio and Serena Villata. Node: A benchmark of natural language arguments. In *Computational Models of Argument - Proceedings of COMMA 2014*, pages 449–450, 2014.

Martin Caminada. Semi-stable semantics. In *International Conference on Computational Models of Argument COMMA'06*, pages 121–130, 2006a.

Martin Caminada. On the issue of reinstatement in argumentation. In *European Conference on Logics in Artificial Intelligence JELIA'06*, pages 111–123, 2006b.

Martin Caminada. Comparing two unique extension semantics for formal argumentation: Ideal and eager. In *The 19th Belgian-Dutch Conference on Artificial Intelligence, BNAIC 2007*, pages 81–87, 2007.

Martin Caminada. Rationality postulates: applying argumentation theory for non-monotonic reasoning. In P. Baroni, D. Gabbay, M. Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation*, pages 771–759. College Publications, 2018.

Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence Journal*, 171 (5-6):286–310, 2007.

Martin Caminada and Srdjan Vesic. On extended conflict-freeness in argumentation. In *Proceedings of the 24th Benelux Conference on Artificial Intelligence BNAIC 2012*, 2012. (contr).

Martin Caminada, Walter Carnielli, and Paul Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22(5):1207–1254, 2012.

Claudette Cayrol, Caroline Devred, and Marie-Christine Lagasquie-Schiex. Acceptability semantics accounting for strength of attacks in argumentation. In *European Conference on Artificial Intelligence ECAI 2010*, pages 995–996, 2010.

Sylvie Coste-Marquis, Caroline Devred, and Pierre Marquis. Prudent semantics for argumentation frameworks. In *Proceedings of the 17th International Conference on Tools with Artificial Intelligence (ICTAI'05)*, pages 568–572. IEEE, 2005.

Sylvie Coste-Marquis, Sébastien Konieczny, Pierre Marquis, and Mohand Akli Ouali. Selecting extensions in weighted argumentation frameworks. In *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA'12)*, pages 342–349, 2012.

Marcos Cramer and Mathieu Guillaume. Directionality of attacks in natural language argumentation. In *Proceedings of the fourth Workshop on Bridging the Gap between Human and Automated Reasoning*, pages 40–46, 2018.

Madalina Croitoru and Srdjan Vesic. What can argumentation do for inconsistent ontology query answering? In V. S. Subrahmanian W. Liu and J. Wijsen, editors, *Proceedings of the 7th International Conference on Scalable Uncertainty Management, (SUM'13)*, pages 15–29. Springer, 2013.

Kristijonas Cyras and Francesca Toni. ABA+: assumption-based argumentation with preferences. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016*, pages 553–556, 2016.

Kristijonas Cyras, Brendan Delaney, Denys Prociuk, Francesca Toni, Martin Chapman, Jesús Domínguez, and Vasa Curcin. Argumentation for explainable reasoning with conflicting medical recommendations. In *Proceedings of the Joint Proceedings of Reasoning with Ambiguous* *and Conflicting Evidence and Recommendations in Medicine (MedRACER 2018) and the 3rd International Workshop on Ontology Modularity, Contextuality, and Evolution (WOMoCoE 2018)*, pages 14–22, 2018.

Célia da Costa Pereira, Andrea Tettamanzi, and Serena Villata. Changing one's mind: Erase or rewind? In *International Joint Conference on Artificial Intelligence IJCAI'11*, pages 164–171, 2011.

Alvaro del Val. Nonmonotonic reasoning and belief revision: Syntactic, semantic, foundational, and coherence approaches. *J. Appl. Non Class. Logics*, 7(2):213–240, 1997.

Jérôme Delobelle, Sébastien Konieczny, and Srdjan Vesic. On the aggregation of argumentation frameworks: operators and postulates. *J. Log. Comput.*, 28(7):1671–1699, 2018.

Franz Dietrich. Scoring rules for judgment aggregation. *Social Choice and Welfare*, 42(4):873–911, 2014.

Franz Dietrich and Christian List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007a.

Franz Dietrich and Christian List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 4(19):391 – 424, 2007b.

Franz Dietrich and Christian List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23:269–300, 11 2007c. ISSN 1474-0028.

Franz Dietrich and Christian List. A liberal paradox for judgment aggregaton. *Social Choice and Welfare*, 31(1):59–78, June 2008.

Franz Dietrich and Philippe Mongin. The premiss-based approach to judgment aggregation. *Journal of Economic Theory*, 145(2):562–582, 2010.

Dragan Doder, Srdjan Vesic, and Madalina Croitoru. Ranking semantics for argumentation systems with necessities. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1912–1918, 2020. (contr).

Didier Dubois, Hélène Fargier, and Henri Prade. Refinements of the maximin approach to decision-making in fuzzy environment. *Fuzzy Sets and Systems*, 81:103–122, 1996.

Conal Duddy and Ashley Piggins. A measure of distance between judgment sets. *Social Choice and Welfare*, 39(4):855–867, 2012.

Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Non-Monotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77:321–357, 1995.

Phan Minh Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, 231:107–150, 2016.

Phan Minh Dung and Phan Minh Thang. Towards (probabilistic) argumentation for jury-based dispute resolution. In *International Computational Models of Argument (COMMA'10)*, pages 171–182, 2010.

Phan Minh Dung, P. Mancarella, and F. Toni. Computing ideal skeptical argumentation. *Artificial Intelligence*, 171:642–674, 2007.

Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, 2011.

Paul E. Dunne, Wolfgang Dvorák, Thomas Linsbichler, and Stefan Woltran. Characteristics of multiple viewpoints in abstract argumentation. *Artificial Intelligence*, 228:153–178, 2015.

Edith Elkind, Piotr Faliszewski, and Arkadii M. Slinko. On distance rationalizability of some voting rules. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 108–117, 2009.

Ulle Endriss, Umberto Grandi, and Daniele Porello. Complexity of judgment aggregation. *Journal of Artificial Intelligence Research*, 45:481–514, 2012.

Patrice Everaere, Sébastien Konieczny, and Pierre Marquis. Counting votes for aggregating judgments. In *International conference on Autonomous Agents and Multi-Agent Systems*, pages 1177–1184, 2014.

Dov M. Gabbay. Theoretical foundations for nonmonotonic reasoning in experts systems. In K. Apt, editor, *Logic and Models of Concurrent Systems*. Springer Verlag, 1985.

Dov M. Gabbay and Odinaldo Rodrigues. Equilibrium states in numerical argumentation networks. *Logica Universalis*, 9(4):411–473, 2015.

Sarah Alice Gaggl and Stefan Woltran. The CF2 argumentation semantics revisited. *Journal of Logic and Computation*, 23(5):925–949, 2013.

Peter Gardenfors. *Knowledge in Flux – Modeling the dynamics of epistemic states*. Cambridge, MA, MIT Press, 1988.

Nikos Gorogiannis and Anthony Hunter. Instantiating abstract argumentation with classical logic arguments: Postulates and properties. *Artificial Intelligence Journal*, 175:1479–1497, 2011.

Peter Gärdenfors and David Makinson. Nonmonotonic inference based on expectations. *Artificial Intelligence*, 65(2):197 – 245, 1994. ISSN 0004-3702.

Anthony Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.

Anthony Hunter. Generating instantiated argument graphs from probabilistic information. In *ECAI 2020 - 24th European Conference on Artificial Intelligence*, pages 769–776, 2020.

Anthony Hunter and Sébastien Konieczny. On the measure of conflicts: Shapley inconsistency values. *Artificial Intelligence*, 174(14):1007–1026, 2010.

Anthony Hunter and Matthias Thimm. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research*, 59:565–611, 2017.

Anthony Hunter, Sylwia Polberg, and Matthias Thimm. Epistemic graphs for representing and reasoning with positive and negative influences of arguments. *Artif. Intell.*, 281:103236, 2020.

Nadin Kökciyan, Isabel Sassoon, Anthony P. Young, Martin Chapman, Talya Porat, Mark Ashworth, Vasa Curcin, Sanjay Modgil, Simon Parsons, and Elizabeth Sklar. Towards an argumentation system for supporting patients in self-managing their chronic conditions. In *The Workshops of the The Thirty-Second AAAI Conference on Artificial Intelligence*, pages 455–462, 2018.

Sébastien Konieczny and Stéphanie Roussel. A reasoning platform based on the MI shapley inconsistency value. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty ECSQARU*, pages 315–327, 2013.

Sébastien Konieczny, Jérôme Lang, and Pierre Marquis. DA^2 merging operators. *Artif. Intell.*, 157 (1-2):49–79, 2004.

Sébastien Konieczny, Pierre Marquis, and Srdjan Vesic. On supported inference and extension selection in abstract argumentation frameworks. In *Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, (ECSQARU'13)*, pages 49–59, 2015.

Sébastien Konieczny, Pierre Marquis, and Srdjan Vesic. New inference relations from maximal consistent subsets. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018*, pages 649–650, 2018.

Sébastien Konieczny, Pierre Marquis, and Srdjan Vesic. Rational inference relations from maximal consistent subsets selection. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI'19*, pages 1749–1755, 2019.

Sébastien Konieczny and Ramon Pino Perez. Merging information under constraints: a logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.

Sébastien Konieczny and Ramón Pino Pérez. Logic based merging. *Journal of Philosophical Logic*, 40(2):239–270, 2011.

Sarit Kraus, Daniel J. Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.*, 44(1-2):167–207, 1990.

Jérôme Lang. Twenty-five years of preferred subtheories. In Thomas Eiter, Hannes Strass, Miroslaw Truszczynski, and Stefan Woltran, editors, *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation - Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, volume 9060 of *Lecture Notes in Computer Science*, pages 157–172. Springer, 2015.

Jérôme Lang, Gabriella Pigozzi, Marija Slavkovik, and Leendert van der Torre. Judgment aggregation rules based on minimization. In *Theoretical Aspects of Rationality and Knowledge*, pages 238–246, 2011.

Jérôme Lang, Marija Slavkovik, and Srdjan Vesic. A weakening of independence in judgment aggregation: agenda separability. In *21th European Conference on Artificial Intelligence ECAI 2014*, pages 1055–1056, 2014.

Jérôme Lang, Marija Slavkovik, and Srdjan Vesic. Agenda separability in judgment aggregation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence, AAAI 2016*, pages 1016–1022, 2016.

Jérôme Lang, Gabriella Pigozzi, Marija Slavkovik, Leendert van der Torre, and Srdjan Vesic. A partial taxonomy of judgment aggregation rules and their properties. *Social Choice and Welfare*, 48(2):327–356, 2017.

Daniel J. Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.

João Leite and João Martins. Social abstract argumentation. In *International Joint Conference on Artificial Intelligence IJCAI'11*, pages 2287–2292, 2011.

Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. Inconsistency-tolerant semantics for description logics. In *International Conference on Web Reasoning and Rule Systems (RR)*, pages 103–117, 2010.

Hengfei Li, Nir Oren, and Timothy J. Norman. Probabilistic argumentation frameworks. In *International Workshop on Theory and Applications of Formal Argument (TAFA'11)*, pages 1–16, 2011.

Christian List and Clemens Puppe. Aggregating sets of judgments: Two impossibility results compared. *Synthese*, 104(1-2):207–235, 2004.

Luca Longo. Argumentation for knowledge representation, conflict resolution, defeasible inference and its integration with machine learning. In *Machine Learning for Health Informatics - State-of-the-Art and Future Challenges*, pages 183–208. 2016.

David Makinson. General pattern in nonmonotonic reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume III*, pages 35–110, 1994.

David Makinson. *Bridges from Classical to Nonmonotonic Logic*. Texts in Computing. King's College Publication, 2005. ISBN ISBN 1-904987-00-1.

Michael K. Miller and Daniel N. Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601, 2009.

Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artif. Intell.*, 195:361–397, 2013.

Till Mossakowski and Fabian Neuhaus. Modular semantics and characteristics for bipolar weighted argumentation graphs. *CoRR*, abs/1807.06685, 2018. URL `http://arxiv.org/abs/1807.06685`.

Hervé Moulin. *Axioms of cooperative decision making*. Cambridge University Press, 1988.

Klaus Nehring, Marcus Pivato, and Clemens Puppe. The Condorcet set: Majority voting over interconnected decisions. *Journal of Economic Theory*, 151:268–303, 2014.

Farid Nouioua and Vincent Risch. Argumentation frameworks with necessities. In *Scalable Uncertainty Management - 5th International Conference, SUM 2011, Dayton, OH, USA, October 10-13, 2011. Proceedings*, pages 163–176, 2011.

Emilia Oikarinen and Stefan Woltran. Characterizing strong equivalence for argumentation frameworks. In *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning, (KR'10)*, pages 123–133, 2010.

Gabriella Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006.

Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *Journal on Data Semantics*, 10:133–173, 2008.

Sylwia Polberg and Anthony Hunter. Empirical evaluation of abstract argumentation: Supporting the need for bipolar and probabilistic approaches. *International Journal of Approximate Reasoning*, 93:487–543, 2018.

David Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36(1):27 – 47, 1988.

Graham Priest. Paraconsistent logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 6, pages 287–393. Dordrecht: Kluwer Academic Publishers, 2002.

Fuan Pu, Jian Luo, Yulai Zhang, and Guiming Luo. Argument ranking with categoriser function. In *International Knowledge Science, Engineering and Management Conference KSEM'14*, pages 290–301, 2014.

Antonio Rago, Francesca Toni, Marco Aurisicchio, and Pietro Baroni. Discontinuity-free decision support with quantitative argumentation debates. In *International Conference on Principles of Knowledge Representation and Reasoning (KR'16)*, pages 63–73, 2016.

Iyad Rahwan and Guillermo R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009.

Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

Nicholas Rescher and Ruth Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–219, 1970.

Lloyd Shapley. A values for n-person games. H.W. Kuhn and A.W. Tucker, eds., Contributions to the theory of games, vol. II. *Annal of Mathematics Studies*, (28), 1953.

Guillermo Ricardo Simari and Ronald Prescott Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53(2-3):125–157, 1992.

Gabriella Pigozziand Marija Slavkovik and Leendert van der Torre. A complete conclusion-based procedure for judgment aggregation. In *1rst International Conference on Algorithmic Decision Theory*, volume 5783 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2009.

Matthias Thimm. A probabilistic semantics for abstract argumentation. In *European Conference on Artificial Intelligence ECAI 2012*, pages 750–755, 2012.

T. Nicolaus Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, 1987.

Francesca Toni. A tutorial on assumption-based argumentation. *Argument Comput.*, 5(1):89–117, 2014.

Leendert van der Torre and Srdjan Vesic. The principle-based approach to abstract argumentation semantics. In P. Baroni, D. Gabbay, M. Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation*, pages 797–837. College Publications, 2018.

Bart Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC 1996)*, pages 357—-368, 1996.

Srdjan Vesic. Identifying the class of maxi-consistent operators in argumentation. *Journal of Artificial Intelligence Research*, 47:71–93, 2013.

Srdjan Vesic and Leendert van der Torre. Beyond maxi-consistent argumentation operators. In *13th European Conference on Logics in Artificial Intelligence (JELIA'12)*, pages 424–436, 2012. (contr).

Gerard Vreeswijk. Interpolation of benchmark problems in defeasible reasoning. In *Proceedings of the Second World Conference on the Fundamentals of Artificial Intelligence WOCFAI 1995*, pages 453–468, 1995.

Peyton Young. Extending Condorcet's rule. *Journal of Economic Theory*, 16(2):335–353, 1977.

Bruno Yun, Madalina Croitoru, and Srdjan Vesic. How to generate a benchmark of logical argumentation graphs? In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of the 7th International Conference on Computational Models of Argument, (COMMA'18)*, pages 475–476. IOS Press, 2018a. (contr).

Bruno Yun, Madalina Croitoru, Srdjan Vesic, and Piere Bisquert. Graph theoretical properties of logic based argumentation frameworks. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS'18)*, pages 2148–2149, 2018b. (contr).

Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Toward a more efficient generation of structured argumentation graphs. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of the 7th International Conference on Computational Models of Argument, (COMMA'18)*, pages 205–212. IOS Press, 2018c. (contr).

Bruno Yun, Srdjan Vesic, Madalina Croitoru, and Piere Bisquert. Viewpoints using ranking-based argumentation semantics. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of the 7th International Conference on Computational Models of Argument, (COMMA'18)*, pages 381–392. IOS Press, 2018d. (contr).

Bruno Yun, Srdjan Vesic, Madalina Croitoru, and Piere Bisquert. Inconsistency measures for repair semantics in OBDA. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, (IJCAI'18)*, pages 1977–1983. AAAI Press, 2018e. (contr).

Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Sets of attacking arguments for inconsistent datalog knowledge bases. In *Proceedings of the 8th International Conference on Computational Models of Argument, (COMMA'20)*, pages 419–430, 2020a. (contr).

Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Ranking-based semantics for sets of attacking arguments. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 3033–3040, 2020b. (contr).

William Zwicker. Voting: an introduction. In *Handbook of Computational Social Choice*. Oxford University Press, 2014.

Index