Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Informática

Dissertação de Mestrado

Mestrado em Engenharia Informática

Argumentation Systems with Social Voting

João Martins, 28024

2º Semestre de 2009/10
28 de Julho de 2010

Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Informática

Dissertação de Mestrado

Argumentation Systems with Social Voting

João Martins, 28024

Orientador: Prof. Doutor João Leite

Trabalho apresentado no âmbito do Mestrado em Engenharia Informática, como requisito parcial para obtenção do grau de Mestre em Engenharia Informática.

2º Semestre de 2009/10
28 de Julho de 2010

Acknowledgements

I would like to thank first and foremost my (ex-)future wife, Diana, for the constant support and putting up with me. I would also like to thank my colleagues at FCT/UNL for interesting input and debate. Other people that provided assorted help were David Henriques, Prof. Paula Amaral e Prof. Nair Santos. Last, but definitely not least, my advisor, Prof. João Leite for the regular meetings and continuous help, right up until the last minute.

Abstract

One of the main uses of the Internet is global communication. Within communication we find the more specific case of debate. In fact, there are entire online communities that do not consider it as merely a component, but as their very core. Such communities usually incorporate a voting mechanism allowing the user base to decide the winner. The existing mechanisms of online debating systems are crude and unsophisticated. They reduce entire debates to statements of the type "party 1 wins".

We believe this standard for online debating systems can be highly improved in a variety of different ways. We wish that 1) the granularity of opinion sharing and debate outcome is improved to the argument level, 2) the content generated in each debate be structured in such a way as to become reusable, 3) the system can provide an official, formally justifiable outcome to each debate that takes into account the opinion of the user base.

The purpose of the current dissertation is to take an initial step towards building the formal infrastructure for such a system to exist. For this purpose, we begin by extending Dung's [17] well established abstract argumentation frameworks with voting information. We find that existing many-valued approaches [19, 11, 13, 25] do not have the intuitions to match this new social context, and thus define a novel generic semantics for taking into account social data. We also specialise this generic semantics to two concrete cases whose behaviour we study.

We provide an initial characterisation of the behaviour of these new semantics. We start at the generic level by studying sufficient conditions for the existence of a single outcome, and the behaviour of the semantics for particular topologies of a framework. We prove that for certain families of semantics, the addition of votes is superficial, but that this result does not hold in the general case.

We find that there are significant differences in the behaviour of the two concrete semantics we propose. We identify the core divergence by looking at characterisations used elsewhere in the literature and applying them to our semantic operators. We motivate our choice of preferred semantics based on that characterisation. Finally, we prove a relation between our proposal and Dung's classical one, showing that the spirit of argumentation as initially defined is still present in such a different context. Another relation is proven with a much similar approach [13], increasing our belief in the augmented expressiveness.

We hope this dissertation to be a first step towards formal but truly subjective reasoning in Argumentation Theory.

Keywords: Argumentation, popular opinion, subjective reasoning, fuzzy, many-valued

Contents

List of Figures

List of Tables

1. Introduction

This dissertation's work is inserted within a very specific context and project. The project is a long-term endeavour to research all the facets of a self-managing online debating system capable of autonomously providing formal outcomes to debates. We argue that existing systems have many shortcomings, and that there is enough interest within the Internet to warrant a system such as the one we propose.

The vision for this online debating system (ODS) states that it should naturally accommodate users of two archetypal, distinct levels of expertise. On the one hand, experts are interested in debating issues seriously, making available a wealth of knowledge and sources in the process. To these experts, the ODS must provide an interface for proposing arguments and counterarguments. On the other hand, less expert users who not knowledgeable enough engage in the debate itself should still be able to participate by voting. To them, the ODS must provide an interface for voting over each argument, enabling users to share their beliefs about the expert's statements and line of reasoning. We will call these less expert users the *crowd*.

The foundation of our vision - and main point of divergence from other existing systems - is that the system should maintain a *formal* representation of the ongoing debates. In other words, underlying a natural language discussion is a knowledge base (KB) structured by the arguments proposed by experts in the debate. Furthermore, that formal representation also contains accumulated statistical data which is representative of the crowd's opinion of each argument.

The ODS must be capable of formal reasoning over the knowledge base and elements of crowd opinion. Its purpose is to provide the outcome of the debate in the form of individual argument acceptability. Because the acceptability of each argument will be highly dependent on the crowd's votes, we claim that this reasoning is not only formal, but also *subjective*.

We believe this research to be important, relevant and interesting for a variety of reasons. Existing systems dedicated exclusively to serious debate demonstrate this. One such example is as www.debate.org. In this website, two (and only two) antagonistic users engage in a round-based debate. Several characteristics stand out when reading these debates attentively.

- The typically courteous greetings in the first round.
- The clear definition of the vocabulary used[1].
- The nature and source of the references cited.
- The type of statements made, *e.g.* stating the difference between correlation and causation[2], or rebuttals and undercuts.

This indicates the degree of seriousness with which debates are carried out on this website. The greetings are cordial and demonstrate willingness. Vocabulary is agreed upon initially to

[1]http://www.debate.org/debates/The-modern-death-penalty-is-beneficial-to-the-economy-of-the-United-States./1/

[2]http://www.debate.org/debates/Tax-cuts-are-based-on-a-politically-unsound-philosophy-which-has-never-worked-in-history./1/

avoid misinterpretations and provide a common basis for arguments. The references range from official government documents to scientific and/or speciality articles. Finally, it is easy to glean a deep understanding of the philosophical nuances of rhetoric.

Furthermore, the same website allows users to vote on which party performed better in several criteria. Each criterion is worth a certain number of points. At the end of the final voting round the points are summed up and one of the two users is declared as the winner. This demonstrates an active interest in three main items: firstly, in very serious online debate; secondly, in opinion-sharing in the form of votes; thirdly, in determining the winner of a debate based on popular, or crowd opinion.

We believe that the envisioned ODS can improve this basic formula in several ways.

- The vast majority of studies in Argumentation Theory have a granularity at the argument-level. In other words, instead of two sides where one is a winner and the other a looser, each argument can be considered acceptable or defeated individually, independently of the side it was proposed by. This allows the outcome of a debate to be more than a win-lose or tie situation as in www.debate.org. Instead, each side can have both accepted and defeated arguments. Indeed, it should not be mandatory (not even implicitly) that both sides are antagonistic with respect to every single argument. This would invalidate the need for a point-based score system to reduce several criterion to the same "point unit". This is, at best, unintuitive.

- With the same idea of granularity in mind, perhaps there should be more values than simply True/Accepted and False/Defeated for argument validity. The inclusion of votes presupposes a sort of popular opinion, and popular opinion is seldom universal. One can hardly take any controversial issue and state it as simply true or false, but it is not shocking to state it to have a certain *degree of truth* (e.g., based on a referendum). A realistic argumentation system must incorporate this fuzzy or many-valued approach on argument validity from the very start.

- A discussion need not be made by two single-person opposing parties. In www.debate.org, any third expert who wants to take part in an ongoing discussion can only do so by waiting until it is over and voting. His relevant, expert input is thus discarded. This is clearly a big disadvantage. Once more, taking granularity to the argument level and disregarding "sides" altogether, a debate becomes a much more relaxed structure capable of accommodating larger numbers of experts. This can lead to more engaging, complete, diverse and far-reaching discussions.

- To generate massive (and thus statistically relevant) crowd-based opinion or knowledge, simpler systems like Facebook's "Like" feature, Youtube's "Like" vs "Dislike" feature or media-sharing "tagging" systems have proven more effective. For the less expert users to participate in a website like www.debate.org, they need to examine several criteria separately and rate them individually. We argue that it is better for individual arguments to have an "Agree" and "Disagree" feature. The simplicity of the system allows users to provide their opinion without having to justify it through rating criteria.

Quite apart from the user-side advantages related with personal and debating experience proposed above, there are other advantages in a system with the envisioned formal backbone. As has been said before, the system should be capable of maintaining a formal representation of the structure and knowledge contained in debates. Then, all knowledge that has been thus formalised becomes *reusable*. Surely the same issue has been debated hundreds if not thousands of times across the internet, each time with slightly different sets of arguments being proposed. No single debate can thus be considered complete. An argumentation-based knowledge base enables relation and correlation of information on a previously unknown scale. Suppose a newly started debate in which a party proposes a certain argument a. Suppose some argument b from a completely different debate is found to attack a. In this situation, the contexts of the two previously separate debates are merged together, providing a much more informative and complete view, increasing our confidence in the outcome of the debate.

Our proposal is then to take the arguments proposed by the experts, the notion of attacks between arguments and popular opinion generated by the crowd and define a semantics capable of providing a justifiable outcome to debates.

There is some work in the literature which relates to our endeavour. For example, since we are interested in people's opinions, we may look at the approach proposed in[9]. The relation of our work with the literature will be discussed in greater detail later in this document.

When large-scale debates involve many participants providing their surely contradicting opinions, a fuzzy approach to argument acceptability should better be able to represent this heterogeneity of viewpoints. However, to the best of our knowledge, no fuzzy systems have broached the subject of social opinion and subjectivity with enough depth to be considered for this project, although some come close.

In summary, while there are some studies that deal loosely with issues we must be aware of, no single proposal appears to provide an adequate solution to them all in conjunction.

The expected contributions of this dissertation as are follows:

- A framework capable of representing the data generated by the online system,
- A generic semantics, with the intent of providing the outcome of a debate,
- A study of two concrete semantics from which actual outcomes can be obtained,
- A description of the behaviour of both the generic and concrete semantics,
- Some insight into how this proposal relates to others in the literature.

This dissertation is organised as follows. In Chapter 2 we provide an overview of the state of the art in Argumentation Theory, focusing on issues relevant to our context. In Chapter 3 we discuss the reasons behind decisions that we have made and define the formal framework and respective semantics. Chapter 4 provides a characterisation of the properties and behaviour of our semantics, and Chapter 5 shows how it relates to similar work in the literature. Finally, we conclude and review our contributions in Chapter 6, and subsequently describe the all the open questions and extensions possible for our proposal in Chapter 7.

2. State of the Art

This chapter is intended to provide an overview of the literature related to Argumentation Theory. First and foremost, we start by providing some very broad insights into research on argumentation theory over the past decades. Then, we will focus on individual studies which contain some more particular element that we consider important and relevant for our current proposal.

2.1 Overview of Argumentation

Argumentation is a field of study whose inception has been traced as far back as Aristotle's *Rhethoric* in the 4th century BC. It attempts to understand the principles behind human thought processes and reasoning, and how those principles apply when ideas are shared and uttered in debates and discussions involving multiple parties. Many issues demand study such as what exactly an argument is, how it can be decomposed, how it, or its components, affect the other arguments in the dialogues, what can be considered the outcome of some dialogue, what types of dialogues are there... None of these questions are trivial, and each viewpoint on argumentation will derive many different opinions on how to provide answers to those questions.

As implied by the nature of Aristotle's work on *persuasion*, the most interesting and problematic dialogues are those where the parties disagree on some point, and attempt to convince one another of their personal views on the subject. A very aggressive stance must be taken when attempting to extract meaning from these dialogues since arguments are generally uttered to undermine and defeat some argument of the opponent.

The second half of the 20th century saw renewed interest in the area of argumentation with the philosophers of the time embracing a current of absolutism, whereby formal logics were highly regarded as a reliable way to represent argumentation and other issues through *inference* [33]. Very few, like Toulmin, disagreed, preferring instead a justificatory procedure by arguing that the formal approaches of the time lacked practical value. Moreover, as research into artificial intelligence gained fame and focus, argumentation came under study from a different, but not totally independent perspective.

Douglas Walton provides a very interesting review on current views of argumentation [35]. One of the most important concepts present in his review is that of argument scheme [34]. Argument schemes are semi-formal representations of arguments with major and minor premises, and some sort of conclusion. Associated with each scheme are critical questions, which may undermine the credibility of the inference or the premises. Its main use is in identifying arguments from a natural language text, and possibly in their formalization as formal, structured arguments. Some examples of these schemes are argument from witness testimony, argument from expert opinion, argument from popular opinion, argument from example, argument from analogy, to name but a few. Some critical questions related to the argument scheme of expert opinion are how credible the expert is, how consistent his opinion is with that of other experts, etc.

Enthymemes are another very important concept for our purposes [35]. An enthymeme is an argument that contains a premise or conclusion omission. These typically happen during natural language argumentation, where some premises or conclusions are deemed so obvious as to not require utterance. Clearly, in formal argumentation systems that rely on complete knowledge, enthymemes are a major disruption.

Another attempt at formalization of arguments was proposed by Toulmin [33], as a rebut of his contemporary philosophers' love of classical logics and absolutism. Instead of just premises and a conclusion, he considered an argument to have six components:

- *Claim*: is the equivalent of the conclusion of a formal argument.
- *Datum*: the basis of the claim.
- *Warrant*: explains how we get from the datum to the claim.
- *Qualifier*: characterizes the strength of the step from datum to claim.
- *Backing*: the reasons why the warrant holds.
- *Rebuttal*: situations where the warrant does not hold, or where the conclusion does not hold.

This definition was built to encompass the idea of defeasibility of arguments. At the time, these concepts were not well taken, but they have heavily influenced modern artificial intelligence [33].

Certain studies do not focus on what conclusions can be derived from a debate, but rather how they are generated [18]. Relying on analogies to legal systems, a formalisation of *burdens of proof* is presented. These are important because at any given time, it should be clear who should utter the next argument, either providing defence for his stand or offence for his opponent's. In implementations of debate in multi-agent systems, an agreed upon schematic with regards to burdens of proof is invaluable, lest two agents function differently and a deadlock occur.

Some practical approaches to decision making use argumentation [2]. Decision making in these cases is seen as "reasoning towards action". The study proposes a two step decision making procedure. On a first step, epistemic arguments are evaluated among themselves and are used to attack practical arguments representing courses of action. This eliminates a considerable amount of possible alternatives, leaving the second step to use several types of semantics to order the remaining alternatives by preference. Argumentation is only used in the first step, and this is justified by its explanatory power.

This is one of the strong points of many kinds of argumentation systems: its ability to intuitively justify and explain its choices. Many approaches, especially those based on classical logic, use a *dialogical* or *dialectical* type of proof procedure [35, 18, 28, 1]. These are characterized by argumentation trees, where each node in the tree is an argument. Its children are that argument's attackers, and its parent is the argument it attacks. It is very easy to imagine the three being constructed by one proponent and one opponent, taking turns to attack undefeated arguments.

Finally, a relation between argumentation and game theory has also been proposed in [28]. Game theory is the study of strategical interaction [23], with the objective of predicting the outcome of a "game" where self-interested agents attempt to further their personal goals. Due to their self-interested nature, one cannot expect them to cooperate freely. *Mechanism design* studies how one can design the rules of a game such that the logical course of action for a self-interested agent is desired system-wide. The study in [28] shows that it is possible to design a mechanism for an argumentation system such that each agent is interested in uttering their true desires, and not some fabricated argument in order to gain advantage.

After this generic introduction to the field of Argumentation Theory, we decide to follow up with a set of specific publications in order to provide a more detailed study of proposals which appear particularly relevant in our context.

2.2 Dung's Abstract Argumentation Frameworks

Abstract Argumentation Frameworks (AAF) as defined by Dung in 95 [17] are a graph-based formalism with an argumentation intuition. It is extremely simple and elegant, and yet serves a variety of purposes. It is also very close to human intuition, and the manipulation of arguments in its formal representation does not require any special training or technical knowledge.

The importance of AAFs comes from their ability to 1) represent and solve a variety of complex problems, but most importantly 2) capture semantics of a variety of other formalisms like Default Logic, Inductive Defeasible Logic and Logic Programming to some extent.

We will use the similarity between AAFs and directed graphs throughout this section.

2.2.1 Definition

An Argumentation Framework AF is defined by a pair $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a set of arguments (or nodes), and $\mathcal{R}$ is a binary attack relation on $\mathcal{A}$ (or arcs in the graph).

Properties of sets of arguments are a central point in the definition of the semantics. Let $S \subseteq \mathcal{A}$. S is *conflict-free* if $\nexists_{a,b \in S}, a\mathcal{R}b$, i.e. there is no attack between two arguments belonging to S. A conflict-free set S is called *admissible* iff each argument in S is *acceptable* w.r.t. S. An argument a is acceptable w.r.t. S iff $\forall_{b \in \mathcal{A}} \exists_{c \in S}, b\mathcal{R}a \Rightarrow c\mathcal{R}b$, i.e. all arguments that attack a are in turn attacked by S, *i.e.* S supports a.

There is not much more with regards to definitions. This is a testament to the simplicity of Abstract Argumentation Frameworks.

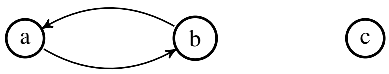


Figure 2.1: Example where set $S = \{c\}$ is a complete but not preferred extension

2.2.2 Semantics

The semantics of AAF are said to be extension-based. Extensions are simply sets of arguments whose validity or acceptability is ensured by the extension itself. Dung's original paper [17] proposed several, and studied properties and relations between them. It is also apparent that their names have not been chosen haphazardly, as one recognises them from several other well-known formalisms.

A *Preferred extension* of an argumentation framework is a maximal (w.r.t. set inclusion) admissible set of AF. Intuitively, it is a maximal set S such that no two arguments in it attack each other, and all arguments which attack $a \in S$ are in turn attacked by S. It always exists, since the empty set {} is admissible. This is an important fact because other semantics may not be defined for certain frameworks, and preferred semantics may be used as a fall back.

A *Stable extension* is a set S of arguments that attack all arguments outside of S, i.e. $\forall_{a \in \mathcal{A} \setminus S} \exists_{b \in S}, b\mathcal{R}a$. It is clear from the definition that any stable extension is also preferred. It is an extremely strong semantics in the sense that it imposes a severe restriction over the extension. As a result, stable extensions may not exist for a variety of frameworks. The simplest way to see this is that the empty set is clearly not stable if $\mathcal{A}$ is non-empty. Despite problems derived from its restrictive nature, it is clearly one of the most credible semantics: the extension can counter-argument or refute anything which it does not claim. It is, in a way, irrefutable in a universal way. Preferred semantics does not have this kind of solid footing, although no argument should be able to attack it without being refuted.

Before *complete extensions* can be defined, it is necessary to introduce the *characteristic function* F_{AF}: $2^{AR} \rightarrow 2^{AR}$. F_{AF}(S) = { a | a is acceptable w.r.t. S }. This function is monotonic (w.r.t. set inclusion), gradually including arguments that have become defensible by S itself. An extension S said to be *complete* iff $S = F_{AF}(S)$. This means that nothing else can be safely justified, if we take S as justified. Intuitively, it is a maximal set w.r.t. argument acceptability, and a fixed point of the F_{AF} function. A preferred extension must be a complete extension since it is maximal by definition, but the reverse is not necessarily true (see Figure 2.1).

The (unique) *grounded extension* of AF, or $GE(AF)$, is the least fixed point of F_{AF}. Clearly, there is a tight relation between grounded extensions and complete extensions due to the use of the characteristic functions. More specifically, the grounded extension is the least complete extension. Intuitively, the grounded extension is the set of all arguments that one can safely assume starting from arguments with are not contested.

Dung also discovered conditions for coincidence between semantics. They rely on properties of the frameworks themselves, not on subsets of arguments. An AF is *well-founded* if there is no infinite chain of attacks. These chains exist when there is a cycle in the argumentation

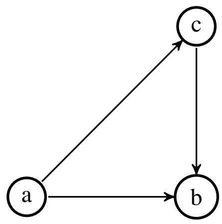


Figure 2.2: Argument a is controversial w.r.t. b

graph, like in Figure 2.1. If AF is well-founded, there is a single extension which is complete, grounded, preferred and stable. Although highly desirable, this property is extremely rare. It is much too common for two arguments to be mutually exclusive. A simple example is a personal debate on whether one should drive or walk to work.

Furthermore, an argumentation framework AF is coherent if each preferred extension of AF is stable. AF is also said to be relatively grounded extension coincides with the intersection of all preferred extensions. These properties are over the semantics themselves, but they can be related with the topology of the argumentation framework itself.

An argument a *indirectly attacks* an argument b if there is an odd-length directed path from a to b. Similarly, a *indirectly defends* b if there is an even-length path from a to b. An argument is *controversial* if it indirectly attacks and indirectly defends another argument, as in Figure 2.2. An AF is *uncontroversial* if none of its arguments are controversial. Finally, AF is *limited controversial* if there is no infinite sequence of arguments such that an argument is controversial w.r.t. its predecessor. Clearly, every uncontroversial framework is also limited controversial.

Dung showed that every limited controversial AF is coherent, and that every uncontroversial AF is both coherent and relatively grounded. Then, each limited controversial AF has a non-empty complete extension and at least one stable extension. Furthermore, if AF is an uncontroversial argumentation framework and $a \notin GE(AF)$ is an argument not attacked by $GE(AF)$, then a belongs to some complete extension, and some complete extension attacks a.

2.2.3 Problem-solving Capabilities

One of the major points of Abstract Argumentation Theory is its ability to capture and solve problems. The importance of the different semantics becomes clear in this context. Dung's original paper [17] formalises both the Stable-Marriage Problem with Gays and N-Person games in AAFs. We will focus on the contributions of his work in N-Person games.

N-Person Game theory proposes to find the best distribution of payoffs to all participants in the game. These payoffs are called *imputations*, and are represented by vectors where each position i represents the payoff for participant i. An imputation *dominates* another if some subset of the dominator's participants are strictly better off than in the dominated imputation. A solution S to an N-Person Game is a set of imputations, obeying the following two rules:

1. No $s \in S$ is dominated by an $s' \in S$.

2. Every $s \notin S$ is dominated by some $s' \in S$.

One may build an argumentation framework where $\mathcal{A}$ is the set of imputations, and the $\mathcal{R}$ is given by dominance between imputations. Then, it should not be hard to see that a solution to the game will be a stable extension to this argumentation framework, and vice-versa. So, it is possible to use an AAF and the stable semantics to solve an N-Person Game. Much more to the point, we know that stable extensions do not always exist. When they do not, there is no solution to the problem. However, this does not mean that the problem is uninteresting. Indeed, it has been shown that there are games without solutions that still model meaningful economic systems [17].

Using preferred semantics ensures that there will be some sort of solution, and that the solution is still free from inner conflict and from outer attack. It will, in fact, provide some kind of partial solution to the problem! This had been an open issue in N-Person Game theory due to the complex and complicated specifics of trying to find solutions for these types of games. Abstract Argumentation Theory's much higher level of abstraction made it possible to find perfectly viable and interesting solutions even when traditional ones are not available.

In the Stable Marriage Problem with Gays, couples are drawn from a single set (the original problem deals with two sets, male and female). This may result in a love-triangle which made the problem instance unsolvable with traditional solvers. Once again, while stable semantics were able to report a complete solution for the problem, preferred semantics will provide a partial solution that may not marry everyone.

2.2.4 Relation with Other Formalisms

Without going into detail, it will suffice to say that Reiter's Default Logic, and Pollock's Inductive Defeasible Logic map into abstract argumentation to some extent. Much like in the two problems we have just seen, stable semantics capture Default Logic's expected semantics, and preferred extensions may be used instead of partial but meaningful partial solutions. The semantics of Inductive Defeasible Logic do not map to the stable semantics, but instead relate to the grounded extension of the corresponding argumentation frameworks.

For our purpose, Abstract Argumentation Theory's relation with Logic Programming is much more relevant, and will thus be presented with more detail. More precisely, we will present Dung's results with regards to logic programming with negation as *possible infinite failure*.

The first result regards stable model semantics for logic programs. Let P be a logic program and G_P the set of all ground clauses of P. Let K be a set of negated literals. An atom k is said to be a defeasible consequence of P, K if there is a sequence of ground atoms $(e_0, ..., e_n)$, with $e_n = k$ such that for each e_i either:

1. $e_i \leftarrow$ belongs to the set of ground clauses of P

2. $e_i \leftarrow a_1, ..., a_t, \neg a_{t+1}, ..., \neg a_{t+r}$ belongs to G_P, a_1 through a_t belong to $\{e_1, ..., e_{i-1}\}$ and $\neg a_{t+1}, ..., \neg a_{t+r} \in K$

We say that K is a support set for k w.r.t. P. One may transform a logic program P into an argumentation framework $AF_P = \langle \mathcal{A}, \mathcal{R} \rangle$ in the following manner:

$\mathcal{A} = \{(K, k) \mid K$ is a support set for k w.r.t. $P\} \cup \{(\{\neg k\}, \neg k) \mid k$ is a ground atom$\}$

$(K, h)\mathcal{R}(K', h')$ iff $h^* \in K'$, where h^* is the complement of h.

Now, let M be a Herbrand interpretation. Recall that M is a stable model of P iff it is the least Herbrand model of P_M. P_M is obtained from P by deleting all clauses with $\neg b$ in their body, where $b \in M$, and deleting negative literals from all other clauses. Let $CM = \{\neg a \mid a$ is a ground atom and $a \notin M\}$.

Dung demonstrated that M is a stable model of P iff there is a stable extension E of AF_P s.t. $M \cup CM = \{k \mid \exists(K, k) \in E\}$.

The second result is about well-founded semantics. Recall that the well-founded model of a logic program P is the least fixed point of the operator

$V_p(I) = \neg.GUS(I) \cup T_P(I)$

where I is a partial interpretation, $T_P(I) = \{b \mid \exists C \in G_P$ s.t. head(C) = b and body(C) is true w.r.t. $I\}$, $GUS(I)$ is the greatest unfounded set of P w.r.t. I, and $\neg.M = \{\neg a \mid a \in M\}$.

Let GE be the grounded extension of AF_P, then the well-founded model of P, $WFM = \{h \mid \exists(K, h) \in GE\}$.

Furthermore, with negation as finite failure, there are correspondences between 1) stable extensions and Clark's completion, and 2) grounded extension and Fitting's model. It is then clear that there is a very important relation between abstract argumentation theory and logic programming.

2.2.5 Further work

Since Dung's seminal paper, further interesting work has been developed, a review of which can be found in [7]. Some inspiration may be had from looking at principles of extension-based semantics, as they define important and desirable properties of semantics. Some of these are concepts that have been defined before for sets of arguments, but generalised to whole semantics. To the best of our knowledge, no such principles exist for labelling-based semantics, which would be more relevant to our work.

Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an arbitrary argumentation framework. Then, let $\mathcal{E}_\mathcal{S}(AF) \subseteq 2^\mathcal{A}$ be the set of all extensions for framework AF under semantics $\mathcal{S}$. Let $\mathcal{D}_{\mathcal{S}}$ be the set of frameworks for which semantics $\mathcal{S}$ is defined, i.e. $\mathcal{D}_{\mathcal{S}} = \{AF \mid \mathcal{E}_{\mathcal{S}}(AF) \neq \emptyset\}$.

- *Conflict-free principle*: A semantics $\mathcal{S}$ satisfies the conflict-free principle iff $\forall AF \in \mathcal{D}_{\mathcal{S}}, \forall E \in \mathcal{E}_{\mathcal{S}}(AF), E$ is conflict-free.
- *Admissibility principle*: A semantics $\mathcal{S}$ satisfies the Admissibility principle iff $\forall AF \in \mathcal{D}_{\mathcal{S}}, \forall E \in \mathcal{E}_{\mathcal{S}}(AF)$,

  $a \in E \Rightarrow (\forall b \in \mathcal{A}, b\mathcal{R}a \Rightarrow E\mathcal{R}b$

- *Reinstatement principle*: A semantics $\mathcal{S}$ satisfies the Reinstatement principle iff $\forall AF \in \mathcal{D}_{\mathcal{S}}, \forall E \in \mathcal{E}_{\mathcal{S}}(AF)$,

  $(\forall b \in \mathcal{A}, b\mathcal{R}a \Rightarrow E\mathcal{R}b \Rightarrow a \in E$

- *Language independence principle*: A semantics $\mathcal{S}$ satisfies the Language-independence principle iff given an isomorphism between two argumentation frameworks, the language independence principle states that their extensions also "inherit" the isomorphism.
- *I-maximality principle*: A set of extensions $\mathcal{E}$ is I-maximal iff $\forall E_1, E_2 \in \mathcal{E}$, if $E_1 \subseteq E_2 \Rightarrow E_1 = E_2$. A semantics $\mathcal{S}$ satisfies the I-maximality principle iff $\forall AF \in \mathcal{D}_{\mathcal{S}}\ \mathcal{E}_{\mathcal{S}}(AF)$ is I-maximal. Intuitively, this principle ensures that the extensions of a given semantics do not contain redundant information by having certain extensions contained within others.
- *Directionality principle*: A semantics $\mathcal{S}$ satisfies the Directionality principle if and only if $\forall AF \in \mathcal{D}_{\mathcal{S}}, \forall S \in \mathcal{US}(AF), \mathcal{AE}_{\mathcal{S}}(AF, S) = \mathcal{E}_{\mathcal{S}}(AF \downarrow_S)$, where $\mathcal{AE}_{\mathcal{S}}(AF, S) \triangleq \{(E \cap S | E \in \mathcal{E}_{\mathcal{S}}(AF)\} \subseteq 2^S$.

  $\mathcal{US}(AF)$ are the unattacked sets of AF, and $AF \downarrow_S = \langle S, \mathcal{R} \cap (S \times S)\rangle$, i.e. the restriction of an argumentation framework to a subset of arguments. This principle ensures that unattacked sets are not affected by the remainder of the framework w.r.t. extensions.

A quick overview of the properties of some famous semantics can be seen in Table 2.1.

More recent semantics have also been developed in order to refine Dung's original work. Particularly, *stage* and *semi-stable* semantics try to maximise *range*. The range of a set S is the union between itself and all the arguments its attacks. This is an interesting way to relate an extension with as many arguments as possible, without going as far as stable semantics. When stable semantics exist, stage and semi-stable semantics coincide with it (since the range of a stable extension is the set of all arguments). However, it is interesting to note that even if stable extensions do not exist, either semantics will still pick a set that maximises the range, i.e., that

	Preferred	Stable	Complete	Grounded
Admissibility	yes	yes	yes	yes
Reinstatement	yes	yes	yes	yes
I-Maximality	yes	yes	no	yes
Directionality	yes	no	yes	yes
	Stage	Semi-stable	Ideal	CF2
Admissibility	no	yes	yes	no
Reinstatement	no	yes	yes	no
I-Maximality	yes	yes	yes	yes
Directionality	no	no	yes	yes

Table 2.1: Properties of some extension-based semantics

reaches the most arguments. Once more, this is a semantics that will allow us to obtain some sort of information that traditional approaches aren't able to harness.

Other semantics include the *ideal* semantics, which define a unique extension as the maximal subset of all preferred extensions, and *CF2* semantics which uses the idea of strongly connected components of a graph to induce extensions.

More can be said on the *justification states* of arguments, namely, whether an argument is

- *Justified*: it belongs to all extensions of a framework
- *Defensible*: it belongs to some extensions of a framework
- *Overruled*: it does not belong to any extension of a framework

Furthermore, Pollock [24] introduced another classification for arguments that made explicit use of the attack relation $\mathcal{R}$. It is possible to "combine" these two classifications resulting in an assortment of combinations, of which only 7 are useful or meaningful [7].

2.2.6 Contributions

The contributions of Dung's original work and its extensions are many-fold. It presents an extremely simple paradigm for reasoning about argumentation. Although many formalisms are shown to be equivalent, it is important to have different perspectives on the same issues. This allows us perceive what may not be readily visible in other representations.

Secondly, despite their simplicity, or perhaps because of it, Abstract Argumentation Frameworks are an extremely elegant formalism capable of capturing several hard computational problems and other known and widely used formalisms, such as non-monotonic logics, logic programming semantics. Furthermore, its simplicity permits its use as a clear, unifying framework, helping us realise and identify previously unseen relations between all these mechanisms.

2.3 Value-based Abstract Argumentation Frameworks

Value-based Abstract Argumentation Frameworks [9] extend Dung's original AF to deal with, not surprisingly, *values*. They are an attempt to "address issues about the rational justification of choices" [10]. Values are concepts and ideals which are desirable by themselves. They are not probabilities, certainties or strengths, but instead something that an argument is said to uphold.

However, different people prefer different values. Therefore, *audiences* become a decisive factor on whether arguments are accepted or not, and thus, the result of a discourse. Different audiences with different opinions, goals and situations will yield different conclusions in a debate.

Examples in [9] deal precisely with situations where both parties in a discourse agree on the premises and the consequences of each action, and yet disagree as to which course to take. Values and audiences, in these cases, may function as tie-breakers, deeming certain courses preferable to others.

Argument schemes are an important method for the identification of arguments and their roles in a debate or discourse. They provide semi-formal schematics that apply to many common types of arguments, from Aristotle's practical syllogism to the sufficient condition scheme for practical reasoning [34]. Both of these schemes define a desired state of affairs, and possible ways of reaching it through a course of actions. The conclusion of the argument is which action to take.

Unfortunately, this approach does not differentiate between alternative actions - how does one choose an action above another, if both actions equally satisfy the goal? If actions are annotated with values, however, it becomes a question of debating which value is preferable, and this leads us once again to concept of audience. One proposal is formalised in the following [6]:

- In the circumstances R, we should perform action A to achieve new circumstances S, which will realise some goal G which will promote some value V.

The generic goal from the previous two schemes is decomposed into three distinct entities: the circumstances resulting from the action, the goal, and the promoted value. This opens up interesting new alternatives in the decision procedure: different actions may result in same state of affairs, different states of affairs ensure the same goal, or different goals are possible for promoting the same value. These were options not representable in previous schemes, yet it is clear that they have an impact on how we reason and what decisions we make.

The general idea for Value-based Argumentation Frameworks (VAF) is that audiences order their preferences over values. For arguments which are undeniable (e.g. it *is* raining), the Truth value is used, and Truth is mandatorily ranked above all other values.

2.3.1 Definition

VAFs are an extension of Dung's AFs. Instead of being a pair, VAFs are 5-tuples $VAF = \langle \mathcal{A}, \mathcal{R}, V, val, P \rangle$, where $\mathcal{A}$ and $\mathcal{R}$ are the same as in a regular AF. V is a (non-empty) set of values, $val : \mathcal{A} \mapsto V$ (a mapping), and P is the set of all possible audiences (total orderings of V).

One may then define an argumentation framework for a specific audience ad, VAF_{ad}. It is the same as the previous VAF, except P is replaced by a transitive, irreflexive and asymmetric relation $Valpref_{ad} \subseteq V \times V$ derived from the chosen ordering $ad \in P$.

The notions of defeat, acceptability, conflict-freedom and preferred extensions can be defined for a particular audience ad. Informally, an argument a *defeats* argument b for an audience ad if a attacks b and $val(b)$ is not preferred to $val(a)$ in audience a. The definitions of acceptability and conflict-freedom are thus based on the notion of defeat, not attack, and are very much equivalent to Dung's.

It is possible to construct an equivalent Abstract Argumentation Framework [17] by simply removing those attacks which do not succeed due to audience value preference. One of the more useful properties of this work is that if VAF contains cycles of arguments in which at least one value differs from the others, the cycle will be broken [8]. Therefore, we know that the preferred extension will coincide with the grounded extension, and there is a polynomial algorithm to compute them.

2.3.2 Example

One example found in [9] deals with a travel arrangement. Suppose Trevor and Katie need to travel to some destination. They may chose either the plane or the train. Both are perfectly valid choices, and both will accomplish the goal equally well. However, travelling by train promotes comfort, while travelling by plane promotes speed. Katie values comfort above speed, while Trevor thinks otherwise. Furthermore, it is boring to go alone.

From these preferences, it is possible to individually assign preferences to Trevor and Katie. Katie might prefer comfort, while Trevor might prefer speed. However, this might not be an issue unless they rank boredom above those other values. In that case, they need to come to terms with one another, and make a decision. Value-based argumentation frameworks provide the infrastructure from which the several possible extensions can be derived.

One interesting issue that this example iconifies is how individual preferences, represented as a directed graph, can be joined together to create a global, not totally ordered set and graph.

2.3.3 Contributions

There are two main contributions in this work. The most obvious is the ability to deal with values, which is an often overlooked concept in formal argumentation systems, but which sees widespread use in real-world argumentation every day. On a more practical note, the addition of values to argumentation frameworks may serve as a tie-breaking mechanism when two parties in a debate agree upon premises and consequences of actions, but whose fundamental beliefs and ideals differ. In this situation, while regular argumentation frameworks may fail obtain a "coherent" set of extensions, value-based argumentation frameworks use values and an audience's preference over these values to prune arguments which do not match their views.

Audiences are the second contribution. Parties in different positions of the political spectrum may be advocate mutually exclusive solutions to the same problem, promoting certain values such as "free market" or "family". In these situations, the electorate will pick their candidate according to the values they believe in. This is a crucial point: the accepted arguments in a discourse will depend largely on the opinion of the observers. This work may then be seen as a first step in making a formal approach at subjective argumentation.

2.4 Generic Gradual Valuations

Generic Gradual Valuations are one of the two proposals to be found in [13]. It is said to be a local, gradual method to assign labels to arguments. It is local because labels are assigned to arguments based solely on their direct attackers. It is a very generic formalisation capable of capturing many other formalisms, by specifying the generic components of the valuation.

We cannot continue without a quick reference to the global approach also found in this article. It is based on the idea that indirect defenders are in fact *supporters*, and thus help strengthen an argument instead of minimising the damage done by an attacker. In this situation, it may happen that an argument attacked by more arguments may be stronger than one without certain attacks. The reason is that certain attackers may actually belong to a support path.

In our view, this idea goes against some solid principles of argumentation and is quite unintuitive. For this reason, we will not focus on it.

2.4.1 Definition

The definition is surprisingly simple. Assume a totally ordered set W with a minimum element V_{min} and a subset V of W containing V_{min} and a top element V_{max}. A Generic Gradual Valuation is simply any v such that

- $v(a) \geq V_{min}$.
- $v(a) = \top_o$ if $\mathcal{R}^{-}(a) = \emptyset$.
- $v(a) = g(h(v(a_1), ..., v(a_n)))$, where $\mathcal{R}^{-}(a) = \{a_1, ..., a_n\}$.

- $h(x) = x$.
- $h() = V_{min}$.

- For any permutation $(x_{i1}, ..., x_{in})$ of $(x_1, ..., x_n)$, $h(x_{i1}, ..., x_{in}) = h(x_1, ..., x_n)$.
- $h(x_1, ..., x_n, x_{n+1}) \geq h(x_1, ..., x_n)$.
- If $x_i \geq x'_i$, then $h(x_1, ..., x_i, ..., x_n) \geq h(x_1, ..., x'_i, ..., x_n)$.

- $g(V_{min}) = V_{max}$.
- $g(V_{max}) < V_{max}$.
- g is non-increasing (if $x \leq y$, then $g(x) \geq g(y)$).

The restrictions above are a set of intuitive desirable properties for any argumentation system. It ensures that whatever the concrete semantics, certain patterns of behaviour will be consistent. The intuition is that h is an aggregator function whose purpose is to compute the strength of a set of attackers. g, on the other hand, is the function that will determine how acceptable an argument is based on its attackers.

2.4.2 Contributions

The contribution, especially as it relates to our endeavour, is very clear. Generic Gradual Valuations are a very versatile and expressive framework for assigning labels to arguments, where the labels indicate a set of possible acceptability or truth values to each argument. Such is the expressive power that it is able to capture the discrete rooted semantics of [19], and on the other the continuous gradual valuations of [11].

These valuations may provide us with a flexible mechanism that can either be used as a base or as inspiration for reasoning in a many-valued setting. The downside of directly reusing this semantics is the inexistence of social support. Social opinion is an essential component of our envisioned system, and it is difficult to include it without some forethought in the semantics themselves.

Another contribution is a of [13] is the identification of a set of intuitively desirable properties for the aggregation of attacks and the reduction of the attacked arguments' acceptability. Any functions or operators whose purpose argument aggregation or argument strength reduction should, in principle, satisfy those properties.

2.5 Bipolar Abstract Argumentation Frameworks

While the relation of argument attack has been studied since the inception of Abstract Argumentation Theory, the support relation between arguments has received much less attention. However, some other relations may be considered [21, 31, 32, 4], and some everyday scenarios appear to suggest that some arguments do not necessarily attack others, but instead serve

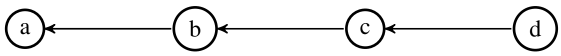


Figure 2.3: A framework for the hiking example

as a support for them. This duality serves to qualify abstract argumentation frameworks that implement them as *bipolar*.

Take the following example, from [15], about a debate on whether to go hiking or not. Four arguments are uttered.

- a: Today we have time, we begin a hike.
- b: The weather is cloudy, clouds are a sign of rain, we had better cancel the hike.
- c: These clouds are early patches of mist, the day will be sunny, without clouds, so the weather will be not cloudy (and we can begin the hike).
- d: These clouds are not early patches of mist, so the weather will be not sunny but cloudy; however, these clouds will not grow, so it will not rain (and we can begin the hike).

Obviously, there are some attacks between these arguments, whose representation can be seen in Figure 2.3. It should be clear that the most important extension in this example is $\{b, d\}$, ruling out hiking (a). Strangely, d supports hiking but is, in practise, the cause for its rejection.

Bipolar argumentation frameworks consider another relation, independent to $\mathcal{R}$, the one of support. Intuitively, the strength of an argument should somehow be an aggregation its attacks and supports, and [15] presents one possible solution.

2.5.1 Definition

Bipolar argumentation frameworks are defined as follows: $BAF = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$. $\mathcal{R}_{att}$ represents attacks as in Dung's work, while $\mathcal{R}_{sup}$ is the *support relation* between arguments.

We say that a *supports* b if there is a path in $\mathcal{R}_{att}$ from a to b. A *supported attack* is a sequence (a, x, b) s.t. a supports x and $x\mathcal{R}_{att}b$. A set S *set-attacks* a if some argument $s \in S$ either directly attacks a or supports an attack towards it. Similarly, S *set-supports* an argument a if some argument $s \in S$ supports a.

Two types of semantics have been studied for these types of argumentation frameworks [15]: extension-based, and labelling-based.

2.5.2 Extension-based semantics

Informally, conflict-freedom of a set S is now given by set-attacks, as opposed to direct attacks. This is said to enforce *internal coherence*. Secondly, a coherent set S (extension) will not set-attack *and* set-support any $a \notin S$ at the same time.

A set S is *+conflict free* iff no element in that set set-attacks another element in S. *Set safety* is based on Dung's controversiality of arguments. Formally, S is *safe* iff $\nexists b \in \mathcal{A}$ s.t. S set-attacks b and either S set-supports b or $b \in$ S. Furthermore, if S is safe, then it is +conflict-free. If S is +conflict-free and closed under $\mathcal{R}_{sup}$, S is safe.

A set S defends a iff every argument that attacks a is direct-attacked (not set-attacked) by an argument in b. This concept is used to define admissibility for a set S:

- *d-admissible*: S is +conflict-free and defends all its elements.
- *s-admissible*: S is safe and defends all its elements.
- *c-admissible*: S is +conflict-free, closed for $\mathcal{R}_{sup}$ and defends all its elements.

Each concept of admissibility in the above list implies the ones above. d-, s- and c-preferred extensions are maximal d-, s-, and c-admissible sets. d-extensions always exist, as the empty set is d-admissible. There are many similarities to Dung's original work [17]. One could further restrict these extensions by making the defends relation based on set-attack instead of direct-attack.

Stable semantics may also be defined analogously for BAFs, using the notion of set-attack instead of direct-attack. Moreover, it has been proven that there if the graph is acyclic there is always a unique stable extension. This extension is only guaranteed to be safe if S is closed for $\mathcal{R}_{sup}$ (and vice-versa).

There are many properties for acyclic BAFs, but given the high implausibility of acyclicity, we will merely refer the interested reader to [15]. Some work in that same reference has been done in converting BAFs into regular abstract argumentation frameworks called *Dung Meta-Frameworks*. The main idea behind this conversion is to aggregate into meta-arguments the connected components (i.e. arguments) under the $\mathcal{R}_{sup}$ relation. Meta-attacks, or *C-attacks*, are inferred from attacks between the sub-arguments. While the two frameworks will be similar, their equivalence is not expected since some specific information is lost during the procedure (e.g. argument-to-argument attacks). Several semantics have proposed for these meta-frameworks which retain most of Dung's properties, but fail to maintain others [15, 14].

2.5.3 Labelling-based semantics

The above extension-based semantics reduced the support relation to another kind of attack, and redefined traditional concepts (e.g. admissibility) on those premises. The labelling-based approach differs radically from this method, and is closely related to that of Section 2.4. Values can be used as labels of continuous domain: they will define the strength of a given argument, depending on the strength and number of attackers and supporters. Notice that values in the VAFs seen previously and in the present BAFs are unrelated.

During the development of these semantics, three principles were kept in mind:

- The value of an argument depends on the values of its direct attackers and of its direct supporters.

- If the quality of the support (resp. attack) increases, then the value of the argument increases (resp. decreases)
- If the quantity of the support (resp. attack) increases, then the value of the argument increases (resp. decreases)

The approach presented (and more fully studied in [3]) is said to be *local* because it depends on *direct* attacks and supports and *gradual* because the domain of values is a continuous interval.

Three functions are used for aggregating the various components necessary to determine the strength of an argument. h_{att} determines the strength of the attackers and h_{sup} does the same for supports. g produces the final strength of the argument based on the two previous functions. These functions must be defined for each situation, but must always obey principles stated above. Let $\{b_1, ..., b_n\}$ be the set of attackers of argument a and $\{c_1, ..., c_m\}$ the set of supporters of argument a. Then,

$$v(a) = g(h_{sup}(v(c_1), ..., v(c_n)), h_{att}(v(b_1), ..., v(b_n)))$$

One instance of these functions, as suggested in [15], is $h_{att}(x_1, ..., x_n) = h_{sup}(x_1, ..., x_n) = max(x_1, ..., x_n)$ and $g(x, y) = \frac{x+y}{2}$.

Notice that since the functions have not been defined, the range of the functions may vary wildly. In the two examples given in [15], the range of h_{att} and h_{sup} is $[-1, 1]$ and $[0, \infty]$. The higher the value of the g function, the more trustworthy an argument is. This approach to bipolarity turns an argumentation framework into a fuzzy system where each argument may be fully accepted, fully refuted, or any value in between.

2.5.4 Contributions

Clearly, these studies are an insightful study into the consideration of a support relation in argumentation. It taps into another commonly used type of argument in real-life which needed a formal representation. The first proposal treats the support relation as a means to define a more general kinds of attack, and enforce stricter rules of coherence upon admissible sets, from which extensions are obtained. In a way, supports are reduced to attacks since supports will not have a great impact on the extensions unless somehow linked to an attack.

The second proposal renounces the boolean requirement of belonging or not to extensions, and provides a fuzzy valuation of arguments. Since support and attack are clearly antagonistic, they must somehow be connected and related, giving rise to the idea that arguments may have strengths, or a continuous measure of credibility. The valuations, denoted as labellings of arguments in this formalisation, become the means by which the strength of an argument is defined. The process through which arguments are evaluated is also extremely customisable, making use of slightly regulated functions. Unlike the previous approach, the support relation in these semantics isn't in any way dependent on the attack relation, being instead on par with it.

2.6 Argument-based Logic Programming

While Dung indicated one way to relate argumentation and logic-programming, other connections may be found. One review that deals mostly with the structured form of arguments from the logic programming perspective can be found in [1].

2.6.1 Defeasible Logic Programming

Defeasible Logic Programming (DeLP) combines techniques from logic programming and defeasible argumentation. The representation is based on logic programming, using facts and rules. Some of these rules are said to be *weak*, and are distinguished syntactically (using a symbol other than $\rightarrow$), and defeasibility stems from the existence of this particular property.

A defeasible rule represents *tentative information*. For example, $lights_on \prec switch_on$ is a rule that states that unless there is information to the contrary, if the switch is on, then the lights are on. If there is no electricity (e.g. $\sim lights_on \prec\sim electricity$), or the bulb is broken, this rule should be overruled. An example of a rule which is *not* tentative but instead definite knowledge is $\sim innocent \prec\sim guilty$.

A program in the DeLP formalism a pair (Π, Δ), where Π is a set of facts and strict rules, and Δ is a set of defeasible rules. Rules are expressed as usual in logic programming, and the formalism includes strong negation $\sim$.

A defeasible derivation of Q from a program (Π, Δ) is denoted by $(\Pi, \Delta) \mid\!\sim Q$ and is a finite sequence of literals $L_1, ..., L_n = Q$ where each L_i is either a fact or the consequent of a rule whose antecedents appear as facts or consequents in previous rules of the derivation.

argument structures are pairs in the form of $\langle \mathcal{A}, H \rangle$, where H is a ground literal and $\mathcal{A} \subseteq \Delta$, s.t.

- $(\Pi, \mathcal{A}) \mid\!\sim H$
- There is no defeasible derivation from $(\Pi, \mathcal{A})$ of contradictory literals.
- There is no proper subset of $\mathcal{A}$ satisfying the two rules above.

We say that $\mathcal{A}$ is an argument supporting the claim H. A query Q succeeds if it is possible to build an argument that supports the query and is undefeated. An argument $\langle \mathcal{C}, P \rangle$ is a *sub-argument* of $\langle \mathcal{A}, H \rangle$ if $\mathcal{C} \subseteq \mathcal{A}$. $\langle \mathcal{B}, S \rangle$ is a *counter-argument* of $\langle \mathcal{A}, H \rangle$ if there exists a sub-argument $\langle \mathcal{C}, P \rangle$ of $\langle \mathcal{A}, H \rangle$ such that P and S disagree (i.e. $\Pi \cup \{P, S\}$ derives complementary literals). P is the counter-argument point.

Now assume that there is an ordering of preference over rules, $\succ$, originating from a syntactic criterion called *generalised specificity* [30]. With this, the notion of defeat can be defined. Let $\langle \mathcal{B}, S \rangle$ be a counter-argument for $\langle \mathcal{A}, H \rangle$ with $\langle \mathcal{C}, P \rangle$ as disagreement sub-argument. Then $\langle \mathcal{B}, S \rangle$ defeats $\langle \mathcal{A}, H \rangle$ if

- $\langle \mathcal{B}, S \rangle \succ \langle \mathcal{C}, P \rangle$ (*proper defeat*)

- $\langle \mathcal{B}, S \rangle \nsucc \langle \mathcal{C}, P \rangle$ and $\langle \mathcal{B}, S \rangle \nprec \langle \mathcal{C}, P \rangle$ (*blocking defeat*)

An *argumentation line* is simply a sequence of arguments $\Lambda = [\langle \mathcal{A}_1, H_1 \rangle, ..., \langle \mathcal{A}_n, H_n \rangle]$ where each element defeats its predecessor. From here it is possible to define the set of supporting arguments and interfering arguments by choosing the odd and even elements respectively. It is clear that we want to avoid infinite argumentation lines generated through repetition of arguments, so we must define *acceptable argumentation lines* as being argumentation lines abiding by these restrictions:

- Λ is finite.
- The sets of supporting and interfering arguments are concordant (their union with Π does not derive complementary literals).
- No argument in the sequence is a disagreement sub-argument of an argument appearing before in Λ.
- A blocking defeater *must* be immediately defeated by a proper defeater, i.e. if $\langle \mathcal{A}_i, H_i \rangle$ is a blocking defeater for $\langle \mathcal{A}_{i-1}, H_{i-1} \rangle$, then $\langle \mathcal{A}_{i+1}, H_{i+1} \rangle$ must be a proper defeater of $\langle \mathcal{A}_i, H_i \rangle$

Because more than one argument may have more than one counter-argument, one must consider not only a single argumentation line, but instead an *argumentation tree* where each branch corresponds to an argumentation line. Each node in the tree is an argument, and the notion of defeat is used to specify which arguments are defeated and which aren't: undefeated if all attacking child nodes are defeated, and defeated if at least one attacking child node is undefeated. More formal details can be found in [1], but the idea is exactly the same as for argumentation lines.

Finally, if the root of a tree is undefeated, then the literal concluded from the argument at the root is *warranted*. Otherwise (i.e. if some node defeats the root), it is *unwarranted*. This is one example of a dialogical or dialectical proof procedure.

On a practical note, there is an interpreter for DeLP, along with several extensions have been developed, such as introducing default negation, explanations to queries, and possibilistic reasoning.

2.6.2 Argument-based Extended Logic Programming with Defeasible Priorities

A interesting approach is taken in [27] with regards to logic programming with both strong and default negation is used. As in the previous study, conflicts are resolved based on a preference relation over rules. However, in this system the preferences (called priorities) will themselves be subject to defeasibility.

Rules are extremely similar to regular logic programming rules, except they have a term associated with them:

$r : L_0 \wedge ... \wedge L_j \wedge \sim L_k \wedge ... \wedge \sim L_m \Rightarrow L_n$

Literals L_i are atoms or strongly negated atoms (e.g. $\neg a$). In the rules, literals may appear negated with default negation $\sim$. Rules have a name r which is a term itself, and which can thus be used in other rules.

An argument is a sequence of rules $[r_0, ..., r_n]$ s.t. all strong literals in the antecedent of a rule are the consequent of a previous rule, and no two distinct rules have the same consequent. Every consequent of an argument A is a *conclusion* of A. A literal L is an *assumption* of A iff $\sim\overline{L}$ appears in a rule in A. An argument is *strict* if it does not contain a defeasible rule.

This introduction of assumptions implies two kinds of conflict between arguments exist: *conclusion-to-assumption* and *conclusion-to-conclusion* attacks. Let S_1 and S_2 be sequences of *strict* rules and the following concatenations $C_1 = A_1 + S_1$, $C_2 = A_2 + S_2$ be arguments. Then, $A1$ attacks $A2$ if C_1 and C_2 conclude complementary literals (conclusion-to-conclusion), or $C1$ contradicts an assumption of $C2$. Priorities among rules are used to compare conflicting arguments.

The notion of defeat is based on *undercutting*, *rebutting* and the priorities of the rules. Rebuts happen in the case of conclusion-to-conclusion conflicts, and which argument wins depends on the strictness and priorities of rules, as well as any undercuts present between the arguments. Undercuts happen in the case of conclusion-to-assumption conflicts. A *strict defeat* implies that the defeated argument does not attack the defeater.

For priorities to be defeasible, a new predicate $\prec$ is added to the language. This predicate will allow priorities to be deduced and defeated "on the fly" by the system. Acceptability, defeat and the characteristic function Γ are then defined w.r.t. what priorities are concluded from a set T of arguments. Arguments are *justified* if they belong to the least fixed point of Γ, and *overruled* otherwise. For more formal details, we refer the reader to [1] or more specifically, [26].

The proof theory for this formalism is given in a dialectical style. A *priority dialogue* is a sequence of arguments alternating between a proponent's and an opponent's such that

- No argument of the proponent repeats itself
- The arguments of the proponent are minimal (and defeat the opponent's previous one)
- The opponent's argument must always defeat the proponent's without considering defeasible priorities (i.e. T is the empty set).

Dialogue trees are trees where all branches are priority dialogues, and they are complete with respect to defeats. An argument is *provably justified* iff there is a dialogue tree with A at its root, and which is won by the proponent (i.e. all branches are won). These proof methods do not vary much between similar systems, as it has been possible to observe.

According to [1], the semantics of this formalism are very akin to Dung's. If static priorities are given, this will correspond to precisely to grounded semantics. When dynamic priorities are considered, different. Some extensions have also been proposed to enhance this formalism.

2.6.3 Argumentation Semantics for Extended Logic Programming

The definition of argument and extended logic program for the current study [29] is exactly the same, and thus will not be repeated. We only exclude the existence of priorities.

As usual, heads of rules in an argument A are called *conclusions* of A, and objective literals subject to default negation in a rule in A are called *assumptions* of A. An argument for conclusion L is minimal if no sub-argument of it also concludes L. The set of minimal arguments associated with P is denoted by $Args_P$.

The notions of undercut, rebut, attack, defeat, strong attack and strong undercut are defined as follows:

- `u`: A_1 *undercuts* A_2 if there is an objective literal L such that L is a conclusion of A_1 and $not\ L$ is an assumption of A_2.
- `r`: A_1 *rebuts* A_2 if there is an objective literal L such that L is a conclusion of A_1 and $\neg L$ is a conclusion of A_2.
- `a`: A_1 *attacks* A_2 if A_1 undercuts or rebuts A_2.
- `d`: A_1 *defeats* A_2 if A_1 undercuts A_2, or A_1 rebuts A_2 and A_2 does not undercut A_1.
- `sa`: A_1 *strongly attacks* A_2 if A_1 attacks A_2, and A_2 does not undercut A_1.
- `su`: A_1 *strongly undercuts* A_2 if A_1 undercuts A_2, and A_2 does not undercut A_1.

Some relations between these different *notions of attack* exist, and they hold for any logic program. They are binary relations, and so their inverse (e.g. $\mathtt{u}^{-1}$) is defined as expected. Here are some of the relations:

- $\mathtt{a} = \mathtt{u} \cup \mathtt{r}$
- $\mathtt{d} = \mathtt{u} \cup (\mathtt{r} - \mathtt{u}^{-1})$
- $\mathtt{sa} = (\mathtt{u} \cup \mathtt{r}) - \mathtt{u}^{-1}$
- $\mathtt{su} = \mathtt{u} - \mathtt{u}^{-1}$

An argument A is acceptable w.r.t. S if all arguments attacking A are attacked by arguments in S. Since arguments can be given by a proponent and by an opponent, we may attribute different attack semantics to each. The opponent may use `u` and the proponent defend himself with another notion of attack, `sa` for instance.

Fixed point semantics can be defined based on a fixed-point operator, $F_{P,x/s}(S) = \{A|A \text{ is } x/y \text{ acceptable w.r.t. } S\}$. Both x and y should be instantiated with one of the above notions of attack.

The authors proved that there are equivalences and subsumptions between some of these fixed-point semantics, creating a *hierarchy of argumentation semantics*. In particular, it is interesting to note that **WFS** = `u/u` and $\mathbf{WFSX}_p$ = `u/a`. Finally, there are results about which semantics satisfy the coherence principle and which generate a conflict-free set of justified arguments. Once again, proof theories are based on dialectical procedures that have been previously seen.

2.6.4 Contributions

Several studies were reported in this section, dealing directly with an argumentation based approach to logic programming. One of the central ideas of the first two was that rules may be prioritised: they do not need to have the same weight in the inference procedure. Much like some arguments may be considered stronger than others, so too can rules. Furthermore, adding defeasibility to rule priorities provides a much more dynamic system with which to work, and somewhat realises the idea that opinions may change over time, preferences themselves being open to challenge.

In the last work, the interesting point is that the developed framework generalises some well-known fixed-point semantics for logic programs. As with any generalisation of this kind, it enables a better viewpoint for comparing between those semantics which already existed, shedding some light and clarity on the issue. It can provide clearer ideas as to why certain semantics have been endorsed while others were found uninteresting.

2.7 Pollock's Defeasible Reasoning with Variable Degrees of Justification II

This work by Pollock is remarkably unlike the system that we envision. It deserves some mention by virtue of being heavily referred to in 4.2.4. In [25], reasoning is made over inference graphs with at least partially structured arguments. The inference graph refers to certain arguments being used to conclude others, thus making undercut and rebut attacks explicit in the graph itself.

There is also an important distinction between epistemic and practical reasoning which deserves further attention. Epistemic reasoning is about what to believe and practical reasoning is about what to do. According to Pollock, "accrual of reasons" exists for practical reasoning, but not for epistemic reasoning. In other words, the more arguments in favour of taking a certain action, the stronger the argument that the action should be taken. In epistemic reasoning, however, the reason to believe a certain statement is only as strong as its stronger supporter. These intuitions about accrual of reasons might very well hold for propositions *against* an argument.

In an online system, it is impossible to define whether debates will be epistemic, practical or both. The notion is still relevant, for example, when considering examples and intuitively validating whether the results appear reasonable or not.

The main contribution of [25] with respect to this dissertation is the definition of a set of properties used to characterise an operator $\blacklozenge$ not unlike g in [13] (refer back to Section 2.4). $a \blacklozenge b$ can be read as a being attacked by b. The properties for $\blacklozenge$ applied to other operators will allow us to gain a deeper understanding of their underlying principles.

2.8 Preference Aggregation

A quick reference to a research area with some similarities is not out of place here. It is already clear that the definition of a semantics will be a balancing act of what properties we want them to satisfy and of how the system will behave. From the literature, we see that Argumentation Theory is not quite as dramatic as Social Choice Theory as regards the satisfiability of properties, but some results in the area of Social Choice Theory may help us understand the process of making a choice between a set of alternative semantics with different behaviours.

Without doubt, the most well known result from Social Choice Theory is Kenneth Arrow's Impossibility Theorem [5]. Very informally, it states that it is impossible for a preference aggregation function to simultaneously satisfy the properties of Unanimity, Independence of Irrelevant Alternatives and Non-Dictatorship. Another way of looking at this result is that if a function has the Unanimity and Independence of Irrelevant Alternatives properties, then it has a dictator, i.e. a single voter whose preferences are always the result of the preference aggregation.

This is one of many similar results that state the inexistence of functions that satisfy certain properties simultaneously. Social Choice Theorists have to decide which functions are best for each application, knowing that there simply isn't a perfect preference aggregation function. These decisions are typically justified by stating why the properties of the chosen function are more desirable than others in a given context.

This method of study is at least partially applicable to the definition of semantics in Argumentation Theory, and this quick reference here should be interpreted as the acceptance that there are no perfect semantics.

3. Framework and Semantics

Now that we have seen some work in the area of Argumentation Theory, in this chapter we will formally define our proposal. We discuss some of the issues surrounding Argumentation Theory in general and apply that discussion to our specific context. This should motivate the driving decisions of our work. Then, we define Social Argumentation Frameworks, an extension of Dung's Abstract Argumentation Frameworks to a social environment. We present completely generic semantics for this framework as the introductory means for obtaining an outcome for a debate. We end the chapter by specifying interesting properties for semantics, and two concrete semantics as a case study along with a light-hearted example.

3.1 Context

In this section we begin by justifying some of the decisions that were made during the development of this work. In Argumentation Theory, as in Social Choice Theory, the justification for the validity or applicability of each element from a set of alternative choices is done by examining it in detail and characterising it according to a set of intuitively desirable properties. Some of these desirable properties are *more* clearly desirable and are universally accepted. On the other hand, certain topics provide many alternatives which are nothing more than differing points of view on the same subject - objectively, they are equally valid. In these situations, the choice may be reduced to a matter of personal preference. As an example that we will explore further on, Pollock [25] defines a series of desirable properties for his $\blacklozenge$ operator and states that he assumes them to be "relatively uncontroversial". Ironically, a particularly important decision of this dissertation is precisely not accepting one of those properties.

The topics that we will address in this section are thus very subjective. We do not claim our approach to be *the* correct approach, but simply that within our context we found them to be the most adequate.

These considerations should also help motivate that Argumentation Theory is not a purely technical pursuit.

3.1.1 Default argument validity

In Chapter 2 we presented Dung's work as a two-valued, graph-based formalism for determining the validity of arguments in a debate. This framework, which has also been adopted in a considerable amount of subsequent work, is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$. It contains the structure of a discussion in the form of a set of abstract arguments and a set of attacks between those arguments.

Within the argumentational context, as we have seen, arguments are assumed to be true by default. There is no reason to doubt a proposition unless there is a logical counter to what has been said or assumed. To further comprehend the reason behind this "postulate" consider logically structured arguments (e.g. [1]): one can intuitively admit that an argument structured as a logical derivation or proof is true unless 1) one of its premises is proven to be false by another argument (undercut) or 2) there is another argument deriving the opposite conclusion (rebut). If neither of these is the case, then one must assume the argument to be true. In a way, this leads to reasoning under uncertainty. Partial knowledge, which exists in any realistic scenario within our context, does not offer the guarantee that every existing counter to an argument can be used in the debate. It is conceivable that an argument which is concluded to be true would in fact be false under complete knowledge (and vice-versa).

Within our project's context there are certain modifications which may interfere with the property of arguments being True by default, in the two-valued sense of Truth. The notion of crowd support seems to indicate that for properly subjective reasoning, there are certain upper bounds to the degree of truth given to an argument. Suppose that many people vote for a certain, unattacked argument, but only 60% agree with it. The remaining 40% of voters would be very misrepresented indeed if a semantics were to simply assign True to such an argument. An even more extreme case is that of an unattacked argument with only only 5% of voters in agreement. Any traditional semantics (i.e., that disregards crowd support) would classify this argument as True since it is unattacked.

This should clarify why arguments being True by default is unacceptable. However, the underlying principle can be adapted to the online debating system context. We will simply say that an argument must be *as true as possible* by default. We should consider the crowd's own support for an argument as an upper bound for the degree of truth given to an argument by the semantics. Any semantics that does not comply with this is clearly violating the crowd's opinion.

This may feel strange to many logicians: things that are commonly accepted as perfectly logical propositions or axioms may not be in accord with the beliefs of the crowd. We must not forget that we are seeking a means of *subjective* reasoning. Our semantics is not expected to evaluate the logical correctness of arguments, but instead their social correctness. It must garner that information from the crowd itself, and only then will social opinion truly influence reasoning.

3.1.2 Extensions vs Labels

Related to the previous topic, we must make a choice whether to opt for an extension-based semantics (arguments are either in our out of the extension) or a labelling-based semantics where arguments are assigned a much more versatile label. It should come as no surprise that we choose the latter. The very existence of an argument's inherent social support which may be neither True nor False indicates that the output of the system must be at least equally expressive. This obviously complicates the semantics but is not regarded as optional, given the nature of our system.

3.1.3 Abstract vs Structured Arguments

In the literature, arguments are deemed to be abstract when no restrictions are made on their nature or structure. The typical way for an argument to be structured (as opposed to abstract) is for it to be based on some sort of logic. In this case, an argument would be a logical derivation/proof from a set of premises to a conclusion. We have already presented the differences between the two approaches in Chapter 2, so we will focus on the implications of choosing one over the other.

As we have introduced it, the ODS must be able to maintain a formal representation of ongoing debates. Abstract arguments allow a lot of flexibility in the identification of what parts of a text in natural language are arguments. This flexibility may mean that certain statements are incorrectly classified as arguments. In this case we will rely on the intelligence of the crowd: an insufficient translation from text to argument will cause the crowd to vote it down, thus reducing its impact in the framework. The same flexibility may also lead to attacks being incorrectly formalised. This is a more serious situation as we assume that attacks are always at “full strength”. We expect it can be relatively easily dealt with by extensions to our proposal. The great advantage of using abstract arguments is that the process of identifying arguments and attacks becomes extremely simple.

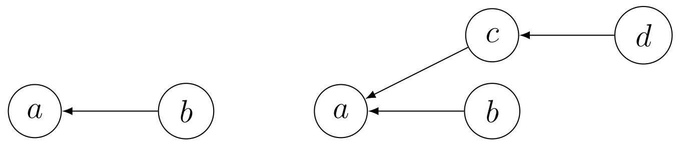
If we consider structured arguments with some underlying logic, then they need to be partitioned into premises and conclusion, both obtained from the text. This process of formalising arguments also requires some technical understanding of the logic being used. On top of this, it is well known that representing textual information or common-sense reasoning in a logic is extremely difficult at best, and virtually impossible for large knowledge bases such as those proposed by debating experts. One instance of this problem is that in common sense reasoning enthymemes are used so often that it is extremely hard for the author of the statement to identify and enumerate them, let alone some third party, be it human or otherwise. Consider the following argument: John is south of a river. He has a boat, so he can go north of the river. It appears to be perfectly obvious, and it is: for humans. For this argument to be logically correct, one would need to state that a river has water, that the boat floats on water, that it can carry John, and that it can be propelled across the water. Without these premises, no logical system is able to derive that John can travel to the opposite side of the river. Similarly, the premises cannot be the target for an attack in a formal context unless they are made explicit. Being able to attack these premises is the crux of the undercutting attacks seen in Sections 2.6.2 and 2.6.3 and as such extremely necessary. In summary, the use of structured arguments makes the translation of natural language text into a debate knowledge base extremely complex and unrealistic.

One could forsake the use of a logic, but still require that arguments have premises and conclusions. In this case, the reasoning steps used in the argument should also be included in the formal representation, as there is no logical guarantee that the conclusion is a consequence of the premises. In other words, the reasoning itself is open to attack. These are perfectly reasonable requirements that lead, however, to the same problem: translation from text to formal knowledge base becomes exceedingly complex. Whatever the case, the translation process is outside the scope of this dissertation and we will simply assume that an excess of complexity in this process significantly hampers its feasibility. Under this assumption and the fact that the whole system hinges on there being a formal knowledge base, we decided for the use abstract arguments.

3.1.4 Bipolarity

Bipolarity, as presented in Chapter 2, is the idea that a single attack relation is not enough to represent real-life debates. Sometimes an argument may not be working against another, but instead serving as a *support*. Support will typically strengthen an argument, but different takes on this notion are possible.

Obviously, incorporating the notion of support is not a trivial matter and therefore one should be absolutely certain that it brings some clear advantage. For example, it exists in [13], although no actual relation is explicitly defined. Instead, a path from an argument to another is said to be an attack branch if its length is odd and a support branch if it is even. By emphasising the role of support branches in their global approach, it happens that argument a is stronger in Figure 3.1b than in Figure 3.1a. This is because a has one support and one attack branch in 3.1b, and one attack branch and no support branch in 3.1a. We argue for the opposite: if one adds an attacker to argument a, then a cannot become stronger under any circumstances. It does not matter whether there is a support branch: another attacker means further evidence that the attacked argument is not true. From this viewpoint, the best case scenario for a would be for the new attacker to be defeated and a maintaining its degree of truth. This illustrates two very different intuitions which do not seem trivially reducible to one another, if at all.



(a) Argument without support

(b) Same argument with support

Figure 3.1: Adding a support branch

As for an explicit support relation, it is quite unclear whether it can be reduced to the notion of attack. One proposal [15] states that it is indeed possible. Here, the support relation is used to define supported attacks. These are then the basis for a redefinition of the acceptability of arguments using the same underlying ideas as in classical work [17]. It is a very limited take on a support relation in the sense that it is reducing it to attacks. It is not immediately obvious that adding a support to an argument increases the probability of it becoming acceptable.

A second, labelling-based proposal in [15] can be seen as a possible answer to these critiques because it considers the two relations as being independent and very much dual. This is enabled by the use of labels to identify the *degree* of acceptability of arguments. To us, whether the

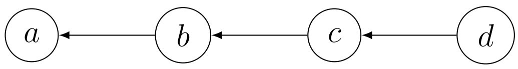


Figure 3.2: A framework for the hiking example

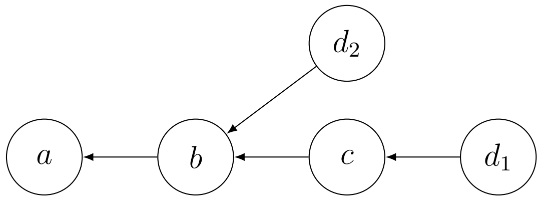


Figure 3.3: Another framework for the hiking example

notion of support brings more expressiveness is still an open question. Take the following motivating example, also based on [15], about a debate on whether to go hiking or not. Four arguments are uttered.

- a: Today we have time, we begin a hike.
- b: The weather is cloudy and clouds are a sign of rain. We should cancel the hike.
- c: These clouds are early patches of mist, the day will be sunny and without clouds, so the weather will be not cloudy (and we can begin the hike).
- d: These clouds are not early patches of mist, so the weather will be not sunny but cloudy; however, these clouds will not grow, so it will not rain (and we can begin the hike).

Obviously, some of these arguments attack one another. This version of the debate can be seen in Figure 3.2. It should be clear that the unique extension, according to Dung, would be $\{b, d\}$, ruling out hiking (argument a). Strangely, d supports hiking but is, in practise, the cause for its rejection.

The conclusion would be that the notion of support is not accurately represented without an explicit support from d to a. However, with a simple reconsideration of what an argument is in the hiking example, one may split argument d into the following:

- d_1: These clouds are not early patches of mist. It will be cloudy, not sunny.
- d_2: These clouds will not grow, so it will not rain (and we can begin the hike).

This division of d into d_1 and d_2 is perfectly viable and can be seen in Figure 3.3. We have simply partitioned the last argument, which we believe was in fact two arguments. The rest remains unchanged. Now notice that d_1 is a rebutting attack on c, stating the opposite information. d_2, however, is an attack on the reasoning behind b. It attacks the assumption that just because there are clouds rain will follow. In this new framework, the unique extension would be $\{a, d_1, d_2\}$. A small alteration to the formal representation of arguments in natural language led to the opposite conclusion: the group should in fact go hiking. Similar revisions led to similar conclusions in most examples presented.

Let us focus on yet another take on support relations. We start with a simple example where a supports b. The idea here is that if a supports b, or in other words if a is evidence for b, then to defeat b one must first defeat a. Until there is no support or evidence for a given proposition, that proposition must be true. A set of supporting arguments can be seen as a conjunction, disjunction or mix of facts (the premises) that lead us to believe the conclusion (the argument which they support). The most obvious and direct solution for removing the support relation is to replace all these arguments with a new, structured argument. Its premises can be the supporting arguments (most logics trivially accommodate conjunctions, disjunctions or a mix as premises). Then, any argument that previously attacked a supporting argument can be considered an undercutting attack while any argument previously attacking the conclusion can be seen as a rebutting attack. Obviously, this is a simplified procedure, but we believe it to be feasible nonetheless. Once again, it is possible to remove the support relation from a framework, without altering its intended semantics.

There may be more intricate and complex cases where these methods cannot be applied. So far, however, no example has convinced us of the *need* for an explicit support relation to represent real-life debates and discussions. In summary, it appears that an explicit support relation does not alter the expressive power of the formal frameworks, although it may simplify the representation of arguments and debates.

In practise, we are not convinced that the motivation for the use of a support relation in argumentation frameworks is strong enough. Because of this and the added complexity of considering a separate support relation, we have simply opted not to include one in our work.

3.1.5 Votes

Since crowd voting is one of the focal points of our system, we must consider 1) what types of votes we allow and 2) how to represent this extra input in our formal framework. We already gave away our choice for 1) and the main reason for it in Chapter 1. As with any statistical system, the more information you have, the greater the accuracy and confidence in the results - and this certainly applies to our context. To subjectively reason about argument validity based on crowd opinion, clearly, the more of that opinion we have as an input, the more credible our results will be. In order to achieve such a high level of participation, we believe that we must make it simple for the crowd to express their opinion. In www.debate.org, a user has to carefully evaluate the entirety of the debate in order to rate each participant over several categories. Each category will attribute points to one of the two participants or to none. This, we believe, has two problems. Firstly, it reduces the whole debate to a score match between two parties. The point value of each category is a very subjective matter in itself, and there is no information about how believable each argument is. It provides a very general overview of each candidate, but very little information about crowd opinion on the *contents* of the debate.

The second problem is that it requires a lot of effort to read several pages of a serious debate - and any rating of candidates is based upon it. We propose a much simpler system: pro/con voting at the argument level. As long as each argument has been properly identified, any user reading only that argument (arguably a small part of a large debate) can easily and quickly state his agreement about that particular piece of the puzzle. An accurate and more detailed overview of the debate is obtained if users read through the whole debate and vote on every proposed argument. The effort necessary to give one's opinions is reduced considerably by allowing users to give only partial information. This ease in opinion sharing should allow our system to gather more crowd opinion, enabling us to provide more credible results.

Regarding the problem of representing the extra input in our framework mentioned in 2), we make a deliberate effort to maintain the existence of votes at every step. To do this, we avoid rewriting the input of pro/con votes as statistical metrics. The purpose of this effort is twofold. Firstly, the context of our project demands it. The envisioned ODS will gather and work with votes, and this framework and semantics are a first foray into its formalisation. Many of the proofs we present also have some intuition based on votes. A typical property will be to analyse the behaviour of the system when a single new vote is added to some argument. Because each vote is an individual event, every framework can be constructed by starting with no votes and adding them one by one. In a way, proofs based on this principle will characterise not just a single event, but the evolution of a framework.

Secondly, by defining a new framework with the updated statistic-based inputs, we would have obfuscated the project's real-world application and context. More to the point however, the semantics would have been defined over our choice of statistical metrics. These would have to be fixed up-front, and we do not wish to do so.

There are two metrics that we believe are of particular interest. By far the most relevant metric is what we call *crowd support*. This measurement should indicate how socially acceptable each argument is from a local perspective. By local we mean that the votes on a should not influence the crowd support of argument b. Furthermore, if a Pro vote is added to an argument, its crowd support should increase, while if a Con vote is added, it should decrease. This metric can be the starting point for attributing argument validity or degree of truth from a set of values and not simple True and False.

Another metric identified as critical is *relevance*. It is not intended to measure the degree of truth of an argument but instead how interesting (relevant!) the crowd thought it was. This raises very interesting questions such as the relative strength of a very credible argument that generated little interest versus that of a very weak argument that generated a lot of interest. We believe that there may be both a global and a local approach to relevance. For the global approach, one can compare the number of votes of a certain argument with the total number of votes in a framework. This will measure the relative interest of an argument with that of all arguments in a framework, resulting in a global, absolute value for how relevant each argument is. The local approach is tied to attacks. Perhaps relevance can be used as a measure of how the defending argument stands up against one its attackers. In this case, it would not matter how an argument compares to all others, but instead how much more or less relevant it is with respect to each of its attackers individually. This difference could be used to lighten or increase the strength of each attack and not of arguments themselves, as suggested by the global approach.

Despite the "relevance of relevance", it is outside the scope of the current dissertation due to time constraints.

3.1.6 The Perfect Argument

The Perfect Argument is a very strong concept when crowd opinion is at stake. A Perfect Argument could be defined as an argument with which no one can ever disagree; a candid universal truth of sorts. Logicians will immediately identify many axioms and valid formulae from their logics, but in practise there are many reasons for such an argument not to exist. Even more so in our context. There will always be a naysayer on the Internet. The fact that one may hide one's true identity simply exacerbates the problem. Someone will deny some fact simply because they can, because they believe it's not true, because they want to annoy others, disrupt a community or a host of other reasons. Even an innocent misclick can lead to a wrongly placed vote and the collapse of the fragile Perfect Argument.

Even were this not the case, the matters open to debate rarely have a logically valid solution. If they did, there would not be so much controversy. Supporting or attacking arguments usually cite references which are themselves debatable subjects: laws, financial reports, books on the topic and others are always open to differing interpretations or points of view. Their credibility can be subject to question. For these reasons, we consider the Perfect Argument on the Internet as absolutely utopian.

We have put so much focus on the practical inexistence of the Perfect Argument because it affects the core of the semantics. Assuming that degrees of truth (term taken from Pollock [25]) exist, then the question is: should an inherently imperfect argument be capable of *completely* defeating another argument? Our answer is no; Pollock's is yes [25]. We investigate the consequences of this matter more formally further on. According to Pollock, as long as an argument a is stronger than another argument b, a should defeat b. Underlying this "relatively uncontroversial" assumption is the difference between practical reasoning and epistemic reasoning.

For reasoning with the opinions of a crowd, we cannot accept that an argument a (with some positive votes) can be utterly defeated by an imperfect argument b. We would be misrepresenting the users who voted for the defeated argument by simply discarding their votes. We accept that a may be attacked into irrelevance, where its weight on the rest of the framework is so small as to become virtually unnoticeable. Indeed, this should happen with very strong attacking arguments, but a should never be defeated totally defeated by arguments other than the Perfect Argument (with which *everyone* agrees (even those that agree with the attacked argument).

One is left to consider the situation where two Perfect Arguments attack one another. The plausibility of such a situation is so small that one may almost ignore, were it not to have a big impact on the semantics themselves, as we will see.

3.1.7 Attack Semantics

When dealing with Pro/Con votes, the informal semantics of the attack relation can become slightly unclear. We wonder whether an attack has the ability to affect *only* the attacked argument or both the attacked and the attacker. Suppose argument b attacks argument a. People have voted for and against both b and a. If the arguments and the attack are properly formulated, one can take a semantic stance that $b \rightarrow \neg a$ in propositional logic. If b is true, then a must be false. Voting-wise, this can be interpreted as a property stating that whoever believes (votes) that b is true must believe that a is false. In this situation we must assume rational voters and a pristine debate, which is not always the case. Even so, with these assumptions, logic-based reasoning appears to capture the correct meaning of attack within the argumentational context. Let us now suppose the contrapositive, $a \rightarrow \neg b$, or if a voter believes a, then he must not believe b. Take the following exchange where b attacks a:

- a: Eight-core computers are the best.
- b: Most applications don't even make use of four cores, so eight-cores can actually be slower.

The implication $b \rightarrow \neg a$ is easy to believe. With more cores, each core typically runs slower (an enthymeme!) and so it can actually perform lower on CPU-bound applications that only make use of a reduced number of cores. In this case, it is clear that eight-core computers are definitely not the best solution. However, if a voter believes that eight-core computers are the best, it is arguable that he must disagree with b. He can agree that it is slower for some applications, but he still believes eight-cores to be better because they outperform quad-cores in his applications of choice.

There are surely many other examples where the contrapositive is perfectly acceptable - they might even outnumber the situations where it does not. For automated systems with rational agents, it might be the best choice. However, in a context where we accept irrational users, it is very important that our semantics does not implicitly assume the crowd's opinions.

This contrapositive issue might already have been discerned from the fact that the embeddings proven by Dung [17] are only with non-monotonic logics such as logic programs. Furthermore, this idea based on classical logic means that a vote in favour of b corresponds to a vote in opposition to a. It is quite possible that the user who voted in favour of b voted against a. In practise then, this user was able to place two votes on a. Because the ODS seeks to accommodate users with different levels of enthusiasm, it is acceptable that the assiduous voter weighs more than the casual browser.

We conjecture that in fact any semantics with statistical metrics on input votes and where the semantics for attack are weakening the attacked argument is implicitly doing such a thing. We wish to abstract ourselves from whether votes are being "passed around" or not. We will not use votes as a "currency" for our semantics nor will we take the classical logic view on the attack relation.

3.1.8 Debate Outcome Uniqueness

As has been said before, the ODS is expected to provide outcomes for ongoing debates. The main purpose of these outcomes is social - the users are interested in knowing which arguments hold formally, and comparing that information with their own beliefs. Furthermore, the objective of the expert users is to attempt to have the majority of their arguments stand up against those of their opponents, so their concern goes without saying.

For these reasons, we believe that the system should only provide a single outcome, not several. Users are engaging in these debates in an attempt to force others to accept their point of view. If the system's answer is that everyone is right because there are different solutions favouring each of the participants, the system's usefulness is next to nil, and it will collapse.

The system must provide a single, unique solution.

3.2 Framework

In this section, we will define the framework underlying the rest of the dissertation.

The framework must be a formal representation of the information the envisioned system is capable of maintaining. In Dung [17], due to the abstract nature of his arguments, a framework is simply $AF = \langle \mathcal{A}, \mathcal{R} \rangle$. This represents the structure of the debate, and we find it perfectly sufficient for that purpose. However, our framework has votes as an additional input, which we include.

Definition 3.1 (Social Argumentation Framework)**.** A *Social Argumentation Framework* is a triple $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, where

- $\mathcal{A}$ is the set of arguments,
- $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation between arguments,
- $V : \mathcal{A} \rightarrow \mathbb{N} \times \mathbb{N}$ is a total function giving the number of *pro* and *con* votes (resp.) for each argument.

Remark 3.1 (Notation)**.** *$\mathcal{V}$ is the set of all possible V functions. Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, $a \in \mathcal{A}$ and $V(a) = (x, y)$. Then, $V^+(a) \triangleq x$ is the number of votes in favour of a and $V^-(a) \triangleq y$ is the number of votes in opposition to a. We use $(a, b) \in \mathcal{R}$ interchangeably with $a\mathcal{R}b$. Let us also denote by $\mathcal{R}^-(a) \triangleq \{a_i : (a_i, a) \in \mathcal{R}\}$ the set of direct attackers of an argument $a \in \mathcal{A}$. Furthermore, we denote the set of arguments from which there is a path to $a \in \mathcal{A}$ as $\mathcal{R}^*(a) = \{b \in \mathcal{A} \mid \exists_{c \in \mathcal{R}^*(a)} (b, c) \in \mathcal{R}\}$, a's ancestors. Finally, let $A \subseteq \mathcal{A}$. Then, $\overline{A} = \mathcal{A} \setminus A$.*

The framework in question still maintains its graph representation, although nodes are now labelled with two natural numbers. From the (direct) attack relation, indirect attack and defence can be derived. Notice that the term defence, as in [13]'s global approach, instead of support. These notions are useful for reasoning about indirectly related arguments.

Definition 3.2 (Indirect attack and defence)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework and $a_1, ..., a_n \in \mathcal{A}$ such that $a_1 \mathcal{R} a_2, a_2 \mathcal{R} a_3 ..., a_{n-1} \mathcal{R} a_n$. If n is even, then we say that a_1 indirectly attacks a_n. If n is odd, then we say that a_1 indirectly defends a_n.

The intuition behind argument defence is simple. Suppose $a, b \in \mathcal{A}$ such that $a\mathcal{R}b$ and $b\mathcal{R}c$. Because the path from a to c is of length 2, we say that a defends c. More specifically, a is protecting c from its attacker, b. a's attack on b reduces b's strength, and consequently the weight of b's attack against c. As a result, c becomes stronger. Moreover, notice that if $a\mathcal{R}b$, then a is also an indirect attacker of b (unlike elsewhere in the literature [13]).

We may extend these notions to sets of arguments.

Definition 3.3 (Set attack and set defence)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework and $A \subseteq \mathcal{A}$. We say that A is *unattacked* if $\forall_{a \in A} \nexists_{b \in \mathcal{A}} (b, a) \in \mathcal{R}$. We say that A is attacked by $b \in \mathcal{A}$ if $\exists_{a \in A}\, b\mathcal{R}a$. We say that A (set-)attacks b if $\exists_{a \in A}\, a\mathcal{R}b$.

The notion of indirect attacks and defence is implicitly defined for sets of arguments much like it was for single arguments.

We follow Dung's naming convention [17] for well-founded frameworks. Our definition is not the the same, but it is equivalent and easier to use in certain proofs.

Definition 3.4 (Well-founded Framework)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework. A sequence $a_1, ..., a_n \in \mathcal{A}$ such that $a_1 \mathcal{R} a_2, ..., a_{n-1} \mathcal{R} a_n, a_n \mathcal{R} a_1$ is called a *cycle*. If there are no such sequences, F is said to be *well-founded*.

Ever since Dung's seminal work cycles have been particular point of interest. For example, if an abstract argumentation framework is well-founded, then all of Dung's semantics coincide [17]. The study of the behaviour of semantics in the presence and absence of cycles is thus a relevant issue. We will attempt to delve deeper and partition arguments into three classes which will be studied separately. For this, we use the notion of cycle dependency.

Definition 3.5 (Cycle dependency)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework. An argument $a \in \mathcal{A}$ is said to be *cycle-dependent* if there is a cycle $a_1, ..., a_n \in \mathcal{A}$ and there is a directed path from $a_i \in \{a_1, ..., a_n\}$ to a and *cycle-independent* otherwise.

One of the main reasons for cycle-dependency is to identify the cycle-*independent* arguments. They cannot be attacked or defended indirectly by a cycle, but can indirectly attack or defend a cycle. As a technical detail, notice that if a belongs to a cycle, it is attacked by some b in that cycle and is therefore implicitly cycle-dependent. To help illustrate, in Figure 3.4, arguments in blue are not cycle-dependent, arguments in red are cycle-dependent and part of a cycle and arguments in yellow are cycle-dependent but not part of a cycle.

Definition 3.6 (Cyclic Form)**.** A social argumentation framework F is said to be in *Cyclic Form* iff $\forall_{a \in \mathcal{A}}$ if $(b, a) \in \mathcal{R}$ then a is cycle-dependent.

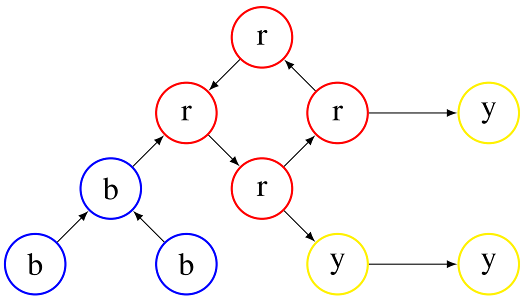


Figure 3.4: Colour coded cycle-dependency

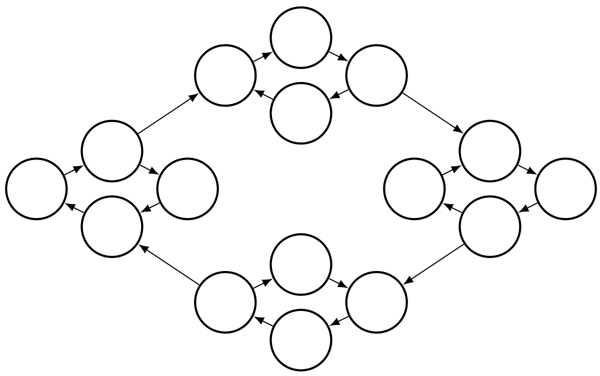


Figure 3.5: A cycle of cycles

This definition is a little more intricate, but very permissive. For example, the argumentation framework in 3.5 is in cyclic form. It is, in a way, a cycle of four cycles of four arguments. In fact any configuration of attacks is still in cyclic form as long as the four cycles exist. The real purpose of the definition is to simply identify frameworks where no tree-like structure (e.g. the blue arguments in Figure 3.4) has a path to a cycle. Isolated arguments do not influence whether a framework is in cyclic form or not. Its relevance will become apparent later on.

Besides the existence of cycles, Dung [17] also provided other conditions for the coincidence and correlation of semantics. These conditions were related to the concept of controversiality. It arises when an argument both indirectly attacks and defends another argument. There is something intuitively undesirable about such arguments, which undermines their validity. We are interested in studying what happens in their absence.

Definition 3.7 (Controversiality)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework. If there are $a, b \in \mathcal{A}$ such that a both indirectly attacks and defends b, then we say that a is *controversial* with respect to b. A framework without controversial arguments is said to be *uncontroversial*.

An example of a framework which is clearly controversial is given in Figure 3.6. In this case, a is controversial with respect to c because it attacks it directly and indirectly defends it through $b_1, ..., b_n$.

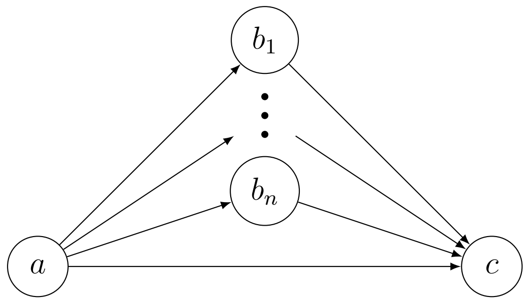


Figure 3.6: Controversial framework

3.3 Semantics

This section is devoted to the definition of our semantics, i.e. the means of obtaining the outcome of a debate where the crowd has voted on arguments. The basis for the semantics is a framework of the type proposed in the previous section. We will begin with some considerations about the input in order to construct our semantics.

Every argument has associated votes. These votes, as we have discussed, derive a value of *crowd support* for the arguments. Therefore, we will need a set from which to draw these values from. Because it is related with the social input given to our framework, we will call this set L_i. We will need to compute the social support values from the given input (i.e. the votes), and for this purpose we will use a function τ.

The semantics' objective is to assign a label to each argument, denoting the degree of acceptability of that argument in the outcome of the debate. We will use another set of values from which to assign these "degrees of acceptability" to arguments. Because these values are the output of the semantics, we will call their set as L_o.

Unattacked arguments do not have any exterior influences. For this reason, their social support, given by τ, should somehow relate to their degree of acceptability quite tightly. For the sake of simplicity, we impose that all input values are also possible output values. Formally, we say that $L_i \subseteq L_o$.

Arguments can be attacked by multiple other arguments. We will also consider that when an argument has more than one attacker, the set of attackers has an aggregated attack strength. Image an argument a which is so poorly constructed that it is attacked by an enormous set of other arguments, $A \subseteq \mathcal{A}$. It is perfectly conceivable that the combined strength of all these attackers can be stronger than the strength of any single attacker (or argument). Neither L_i nor L_o contain values higher than the strength from a single argument, and so we will draw the values for the aggregated attacks from yet another set, L_a.

Once again for simplicity's sake, we impose that aggregated attack strength must exist when there is a single attacking argument. For this reason, the aggregation of a single argument should amount to the argument's strength itself. For this reason, the aggregation value set, L_a, must also be able to represent the values of single arguments. In other terms, $L_o \subseteq L_a$.

We have now motivated an important part of the structure for our semantics. We now provide a quick review. We currenly assume three totally ordered sets $L_i \subseteq L_o \subseteq L_a$. Their meaning is the following:

- L_i: containing the values for the social support of each argument, the social *input*.
- L_a: containing the values for an *aggregation* of several arguments, a *combined attack*.
- L_o: containing the values that our semantics will assign as final argument strength, or its *output*.

Remark 3.2 (Notation for elements of L_i, L_a, L_o)**.** *The top and bottom elements L_i (resp. L_a, L_o), will be denoted by $\top_i$ (resp. $\top_a, \top_o$) and $\bot_i$ (resp. $\bot_a, \bot_o$).*

These sets provide us with the values with which we will work. We now turn our attention to the operators which will manipulate these values in order to obtain the final outcome of the debate. We have already mentioned the idea of an aggregated attack. To implement this idea, we will use an operator $\curlyvee$ which aggregates arguments and outputs a single value corresponding to their combined strength. This aggregated attack strength is a "positive", monotonic value in the sense that adding more arguments makes it increase. We need a $\neg$ operator to turn it into a "negative" force that decreases the attacked argument's strength. The last operator $\curlywedge$ is used to relate an argument's initial crowd support with the negative value of the attack. The relations between these operators and the intuitions presented above should already be somewhat clear.

We will use the notion of semantic framework to centralise all the concepts we have introduced in this section until now. It will contain the ingredients we use to obtain the outcome of the debate.

Definition 3.8 (Semantic framework)**.** A semantic framework is a 7-tuple $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$, where

- L_i : a totally ordered set of social *input* values,
- L_a : a totally ordered set of *combined attack* values,
- L_o : a totally ordered set of *output* values,
- $\tau : \mathcal{A} \times \mathcal{V} \to L_i$,
- $\neg : L_a \to L_a$,
- $\curlywedge : L_i \times L_a \to L_o$,
- $\curlyvee : L_a \times L_a \to L_a$.

We impose no restrictions on these operators except that they be total. This decision is justified by our objective to always provide an answer.

Perhaps some readers familiar with fuzzy logics recognize $\curlywedge$ and $\curlyvee$. This is one of the notations for T-norms and T-conorms respectively. Although the operators are not mandatorily T-norms and T-conorms, one may loan the intuition from these well-known operators for a better understanding of their role.

Remark 3.3 (Notation for τ)**.** *τ depends on a framework F for the information about the votes. We often compare two similar frameworks such as $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ and $F' = \langle \mathcal{A}', \mathcal{R}', V' \rangle$, and because of its extensive use a simplification of the regular notation of τ is desirable: we will denote $\tau(a, V)$ as $\tau(a)$ and $\tau(a, V')$ as $\tau'(a)$. Furthermore, we will also use τ in the following forms $\tau(x, y) \triangleq \tau(V^+(a), V^-(a)) \triangleq \tau(a)$.*

Remark 3.4 (Notation for $\curlyvee$)**.** *To increase clarity, much like $\sum$ for $+$, we use $\bigvee$ for $\curlyvee$. For a sequence of arguments $S = x_1, ..., x_n$ we define $\bigvee_{i=1}^{n} x_i \triangleq x_1 \curlyvee ... \curlyvee x_n$. Furthermore, if $\curlyvee$ is commutative and associative, we may simplify the sequence to a set and use $\bigvee_{x \in S} x \triangleq x_1 \curlyvee ... \curlyvee x_n$ instead.*

The heart of the semantics is in the definition of a model. The model of an argument is an equation that puts all the operators together in order to give that argument a meaningful value.

Definition 3.9 (Model). Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework and $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ be a semantic framework. A model of F under semantics $\mathcal{S}$ is a total mapping $M : \mathcal{A} \rightarrow L_o$ such that $\forall a \in \mathcal{A}$

$$M(a) = \tau(a) \curlywedge \neg \bigvee_{a_i \in \mathcal{R}^-(a)} M(a_i)$$

The definition should clarify how an argument's model is meant to be its crowd support reduced by the aggregation of its attackers. If we assume the normal behaviour for T-norms, then we know that the T-norm conjunction between $\tau(a)$ and its attack should be a reduced value of $\tau(a)$, which is perfectly in accordance with argumentation intuitions: an attacked argument sees its strength reduced.

Remark 3.5 (Notation for model sets)**.** *We denote the set of all models of framework F under semantic framework $\mathcal{S}$ as $\mathcal{M}_{\mathcal{S}}^{F}$. Whenever the framework F or semantic framework $\mathcal{S}$ are unambiguous, they may be ommited. In this case, the set of models is denoted by $\mathcal{M}$. If for $a \in \mathcal{A}$, $\forall_{M,N \in \mathcal{M}}\ M(a) = N(a)$, we say that a has a unique model.*

Definition 3.10 (Model of a set of arguments)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$, $A \subseteq \mathcal{A}$ and $M \in \mathcal{M}$. We say that a total mapping $N : A \rightarrow L_o$ where $\forall_{a \in A} N(a) = M(a)$ is a model of A. Furthermore, if $\forall_{M,N \in \mathcal{M}}\ \forall_{a \in A}\ M(a) = N(a)$, we say that A has a unique model.

It is important to draw attention to the fact that at this stage nothing is known about the number of models of a framework under a certain semantics. We will study this matter further on, as we believe it is of central importance to the use of argumentation in any context. It is interesting that certain other similar approaches do not even broach this subject [13, 12].

3.4 Specific Semantic Frameworks

A generic semantics is now in place. While it is useful in for embedding other systems and providing and umbrella for a wealth of more concrete semantics, we also need concrete, completely defined semantics to provide an actual outcome to debates in the envisioned ODS. For that, we need to specify a semantic framework, and then analyse how it behaves in practise.

Before going to a completely concrete state, one essential initial step is to define some high-level policies for our operators. Despite the flexibility of their definition, there are certain desirable properties that enforce the intuitions we motivated in the beginning of the present chapter. Indeed, as we introduced each operator and their role, some of these behaviour-defining properties might have become clear. We will simply say that a semantic framework is *well-behaved* if it has all of these properties. For example, in argumentation the order of attack is typically irrelevant. For this reason we impose commutativity and associativity to $\curlyvee$ on well-behaved semantic frameworks.

We will use monotonic with the particular meaning of *increasing* monotonic, as opposed to antimonotonic with the meaning of *decreasing* monotonic.

Definition 3.11 (Well-behaved semantics)**.** A semantic framework $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ is *well-behaved* iff

- $\curlywedge$ is monotonic w.r.t. to both arguments,
- $l_i \curlywedge \neg \bot_o = l_i$,
- $\curlyvee$ is monotonic w.r.t. to both arguments,
- $\curlyvee$ is commutative and associative,
- $\curlyvee$ has $\bot_o$ as identity element,
- $\neg$ is antimonotonic,
- τ is monotonic w.r.t. to V^+ and antimonotonic w.r.t. V^-,
- the generalised De Morgan laws apply.

It should be clear that in well-behaved semantic frameworks $\curlyvee_{x \in \{x_1\}} x = x_1$, and that $\curlyvee_{x \in \emptyset} x = \bot_o$. Moreover, an the model of an unattacked argument under a well-behaved semantics is $M(a) = \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^-(a)} M(a_i) = \tau(a) \curlywedge \neg \bot_o = \tau(a) \curlywedge \neg \bot_o = \tau(a)$.

The application of the following definition will not be clear until we start modifying frameworks later on, but it simply states that we are able to derive votes that accurately represent some given social support value.

Definition 3.12 (Reversible semantics)**.** Let $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ be a semantic framework. $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ is *reversible* iff $L_o \in L_i$ and τ is surjective.

We proceed with some considerations for τ. Since we will use only a single τ for this whole dissertation, we define it here once and for all:

$$\tau(a) = \frac{V^+(a)}{V^+(a) + V^-(a)}$$

The function τ should immediately be recognised from the above formula as a measurement for the crowd support of arguments. It is obviously increasing monotonic w.r.t. a positive vote and antimonotonic w.r.t. a negative vote. The result is simply the percentage of votes which are positive for this argument. As can be seen, τ evaluates only the universe of people that voted for each particular argument (hence its being local), and gives us a metric for how much that universe believes in the argument.

With τ is defined, we provide a simple result that has to do with the reversibility of the semantics. The idea is that from a social support value we might want to build a framework that corresponds to that value, and so we need to ensure that values for V^+ and V^- exists.

Proposition 3.1. *Let $x \in \mathbb{Q}$ such that $0 \leq x \leq 1$. Then there is y and z such that $x = \frac{y}{y+z}$ and $\frac{y}{y+z}$ is irreducible.*

Proof. From $x \in \mathbb{Q}$ we know that there is an irreducible fraction $x = \frac{a}{b}$. From $0 \leq x \leq 1$ we know that $a \leq b$. Then, let $y = a$ and $z = b - a \Leftrightarrow z + a = b \Leftrightarrow y + z = b$. □

We now propose two semantic frameworks, the second of which belongs to an appealing family of frameworks. Their names originate from the arithmetic operators they rely on.

3.4.1 Subtraction Semantics

Our first proposal is the subtraction semantics. Clearly, subtraction is the antimonotonic component of this framework. The rest of the operators are expected to be monotonic.

Definition 3.13. (Subtraction semantics) Let

- $L_i = L_o = [0, 1] = \mathbb{R}_1$.
- $L_a = [-\infty, +\infty]$.
- $\neg l = -l$.
- $l_1 \curlywedge l_2 = \max(0, l_1 + l_2)$.
- $l_1 \curlyvee l_2 = l_1 + l_2$.

Then, $\mathcal{S}^{-}$ is the Subtraction Semantic Framework.

The use of bounded subtraction is crucial. Suppose we didn't have bounded subtraction and that the set of attackers of a had an aggregated attack value l_a such that $l_a > \tau(a)$. Then, $M(a) < 0$. If another argument b was only attacked by a, $M(b) = \tau(a) - l_a > \tau(a)$.

Recall that we do not wish to allow an argument's final strength (i.e. its model) to be higher than its crowd support. The use of bounded subtraction ensures that the strength of an argument cannot be increased by its being attacked, a highly desirable property.

Proposition 3.2. *In a Subtraction Semantic Framework, the generalised De Morgan laws apply.*

Proof.

$$\begin{aligned} \neg(l_1 \curlyvee l_2) &= -(l_1 + l_2) \\ &= -l_1 - l_2 \\ &= (-l_1) + (-l_2) \\ &= \neg l_1 \curlywedge \neg l_2 \end{aligned}$$

□

The generalised De Morgan laws are extensively used in certain proofs, along with commutativity. When put together, these two properties enable us to isolate a set of attackers from the model of an argument and apply that partial attack to an argument's crowd support. In the limit scenario, it allows us to consider each attacker separately.

This semantics allows us to transform our framework and semantics into a system of equations of the following form:

$$\begin{cases} M(a_1) = \max\left(0, \tau(a_1) - \sum_{a_i \in \mathcal{R}^{-}(a_1)} M(a_i)\right) \\ \ldots \\ M(a_n) = \max\left(0, \tau(a_n) - \sum_{a_i \in \mathcal{R}^{-}(a_n)} M(a_i)\right) \end{cases}$$

We will wait until Chapter 4 to characterise the behaviour of these semantics further.

3.4.2 T-Norm Semantics

Only slightly more concrete than the generic semantics, but making full use of its notation, we may define T-norm based semantics. There are a few properties which can be proved at this level, being applicable to more than the Product Semantics that follow the T-norm semantics.

Definition 3.14. (T-norm semantics) Let

- $L_i = L_a = L_o = [0, 1]$.
- $\curlywedge$ is a T-norm.
- $\curlyvee$ is the T-conorm associated with the $\curlywedge$ T-norm.

- $\neg l_a = 1 - l_a$.

Then, $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ is a T-norm Semantics.

Proposition 3.3. *In a T-norm based semantic framework, the generalized De Morgan laws apply.*

Proof. Starting from the definition of T-conorm,

$$\begin{aligned} l_1 \curlyvee l_2 &= 1 - (1 - l_1) \curlywedge (1 - l_2) \\ 1 - l_1 \curlyvee l_2 &= (1 - l_1) \curlywedge (1 - l_2) \\ \neg(l_1 \curlyvee l_2) &= \neg l_1 \curlywedge \neg l_2 \end{aligned}$$

□

The Product Semantics is a T-norm Semantics based on the product T-norm. We will follow in more closely instead of other T-norm Semantics because of its interesting properties and simple formulae. It is continuous, archimedean, and does not require us to use $\max$ (as for bounded subtraction). Continuity is interesting because it facilitates the smooth evolution of the outcomes of the framework, e.g. it will not cause a single vote to greatly and single-handedly affect the outcome of a debate.

Definition 3.15. (Product semantics) Let

- $L_i = L_a = L_o = [0, 1] = \mathbb{R}_1$.
- $\neg l = 1 - l$.
- $l_1 \curlywedge l_2 = l_1 \cdot l_2$, the product T-norm.
- $l_1 \curlyvee l_2 = 1 - (1 - l_1) \cdot (1 - l_2)$, the T-conorm derived from the product T-norm.

Then, $\mathcal{S}^{\times} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ is the Product Semantics.

The equation system derivable from the Product Semantics is

$$\begin{cases} M(a_1) = \tau(a_1) \cdot \prod_{a_i \in \mathcal{R}^{\text{-}}(a_1)} (1 - M(a_i)) \\ \dots \\ M(a_n) = \tau(a_n) \cdot \prod_{a_i \in \mathcal{R}^{\text{-}}(a_n)} (1 - M(a_i)) \end{cases}$$

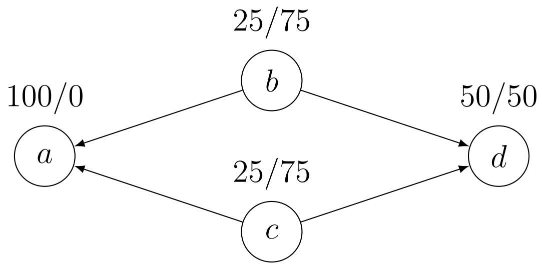


Figure 3.7: The Portuguese Summer, Beaches and Parks

3.4.3 Example

It is very hard to realise the differences of $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$ based on such abstract definitions and motivations. For that reason, we present a very simple illustrative example that illustrates the major differences between the two semantics. We will merely mention the differences here, and discuss them properly later in this document. Imagine the following discussion takes place.

- a: The beach is a good place to be in the Summer.
- d: The park is a good place to be in the Summer.
- b: It is windy in the Summer.
- c: It is cloudy in the Summer.

Now, imagine this discussion is open to voting by a portuguese crowd. *Of course*, the beach being a good place to be is a perfect argument! The park is also considered to be a good idea by the Portuguese, but definitely not as pleasant as the beach. Summers are occasionally windy and cloudy in Portugal, a fact which makes going to the beach or to the park less desirable. It is also a general consensus among the Portuguese that it is typically not very windy nor very cloudy in the Summer.

This light-hearted debate and respective opinion (in the form of votes) could be represented in our system by the framework in Figure 3.7. Since b and c are unattacked, then $M(b) = M(c) = \frac{25}{100} = 0.25$, whatever the semantics. We also know that $\tau(a) = 1$ and $\tau(d) = 0.5$.

Under the Subtraction Semantics, $M(a) = \max(0, 1 - (0.25 + 0.25)) = 0.5$ and $M(b) = \max(0, 0.5 - (0.5 + 0.5)) = 0$. The beach is *clearly* the best place to be at, even formally. No one can dispute that now! In fact, under these semantics we do not even consider the park as an option, which is strange since at least someone thought it was a good place to go.

For the Product Semantics, the model of b and c remains the same, but the model of the other two is altered as follows. $M(a) = 1 \cdot (1 - (1 - (1 - 0.25)(1 - 0.25))) = 0.75 \cdot 0.75 = 0.5625$ and $M(b) = 0.5 \cdot 0.75 \cdot 0.75 = 0.28125$. Notice that this time around, going to the park is still an option, although substantially reduced from its original strength.

4 . Characterisation

Now that the framework and respective semantics are in place, it is our intent to study this proposal at the generic and concrete levels in order to understand its behaviour and see whether it has desirable properties.

In Section 4.1 we will analyse the behaviour of our generic semantics at the model and model set levels according to the concepts of well-foundedness and uncontroversiality. Next, we will look at the impact of the voting extension to our framework and check what sort of expressiveness gains it brings. We do this by checking for the existence of a correspondence between unlabelled and labelled frameworks under certain semantic restrictions. In Section 4.2 we will do a more in-depth study of the concrete semantics, looking at particular cases of cycles and then at operator properties found in the literature.

4.1 Generic semantics

This section deals only with the abstract version of our semantics, although we typically restrict semantic frameworks to the well-behaved case. As said above, we initially look at the behaviour of the system in (a)cyclicality. Next, we define controversiality and study its impact and finally, we expand on the real consequences of adding votes to the framework.

4.1.1 Cycles

Our generic semantics is extremely permissive about the existence of multiple models in the same framework. Within our context, however, it is desirable that the outcome of the ODS be unique. In this section, we propose to investigate the source of the existence of multiple models.

We use a simple idea for this undertaking. First, we identify some framework-level properties that are sufficient for the uniqueness of the model. Then, we attempt to discover components of a larger framework satisfying those properties, and which should, we argue, have a unique model as well. Since those components have a unique model, then whatever the model of the entire framework, their impact (through attacks) on the rest of the framework is static and model-independent. They may thus be treated in a special manner.

The idea is that under a set of simple semantic restrictions it may be possible to propagate the attacks of certain arguments into the rest of the framework, and subsequently remove the attacks. This allows us to consider simplified and reduced but semantically equivalent frameworks. Semantic properties which hold in these simplified frameworks, by virtue of semantic equivalence, also hold in the original framework. This preprocessing process should eventually result in simplified, reduced frameworks where the source of the multiplicity of models can be more easily identified.

We will begin by defining the tools for this method. Since we wish to alter frameworks for precomputing attacks, we need to make sure that every transformation guarantees that the meaning of the framework itself is unaltered.

Definition 4.1 (Semantic equivalence)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ and $F' = \langle \mathcal{A}', \mathcal{R}', V' \rangle$ be social argumentation frameworks. F and F' are *equivalent under semantics* $\mathcal{S}$, denoted by $F \equiv_{\mathcal{S}} F'$, iff $\mathcal{M}_{\mathcal{S}}^{F} = \mathcal{M}_{\mathcal{S}}^{F'}$.

Note that if the set of models is the same for both, then the models themselves have to be the same, immediately implying that $\mathcal{A} = \mathcal{A}'$. Now we define the abstraction which allows us to perform alterations to frameworks.

Definition 4.2 (Social argumentation mapping)**.** Let $\mathcal{F}$ be the set of all social argumentation frameworks. A *social argumentation mapping* is any function $\Phi : \mathcal{F} \to \mathcal{F}$.

Only a subset of social argumentation mappings interest us, however. Namely the ones that preserve semantic equivalence.

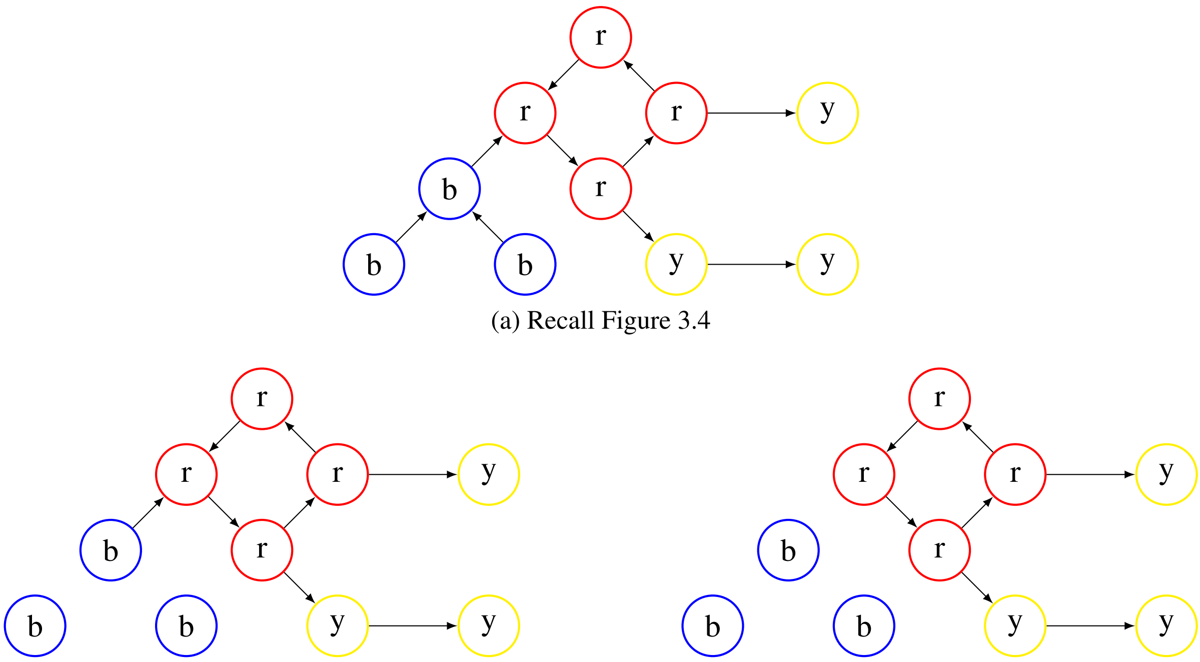
Definition 4.3 ($\mathcal{S}$-preserving social argumentation mapping)**.** A social argumentation mapping Φ is $\mathcal{S}$-preserving iff $\forall_{F \in \mathcal{F}}\, \Phi(F) \equiv_{\mathcal{S}} F$.

We now define a mapping for the purpose of preprocessing attacks. It is important to preprocess only attacks which we know always have the same impact on the rest of the framework. The strength of the attack is derived from that of the attacker himself. As long as the model of an argument a is unique, then all attacks originating from a have a model-independent impact on the rest of the framework.

When an argument a is unattacked under well-behaved semantic frameworks, then we have noted before that $M(a) = \tau(a)$. Since we consider the votes as being static input, then it turns out that an argument being unattacked is one of the simplest sufficient conditions for it having a unique model. Therefore, we can preprocess attacks originating from unattacked arguments using the following mapping.

Definition 4.4 (Hubby mapping)**.** Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework and $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ a reversible, well-behaved semantics. Let Φ_H be the social argumentation mapping such that $\Phi_H(F) = \langle \mathcal{A}', \mathcal{R}', V' \rangle$ where

- $\mathcal{A}' = \mathcal{A}$,
- $\mathcal{R}' = \mathcal{R} \setminus \{(a, b) \in \mathcal{R} \mid \nexists_{c \in \mathcal{A}}\, (c, a) \in \mathcal{R}\}$,
- For V', let $A = \{a \in \mathcal{A} \mid \nexists_{b \in \mathcal{A}}\, (b, a) \in \mathcal{R}\}$ be the set of unattacked arguments, and $B = \{a \in \mathcal{A} \mid \mathcal{R}^{-}(a) \cap A \neq \emptyset\}$ be the set of arguments attacked by some unattacked arguments.
  1. $\forall_{a \in \overline{B}}\, V'(a) = V(a)$.



(a) Recall Figure 3.4

(b) After an application of Φ_H

(c) After a second application of Φ_H

Figure 4.1: Evolution of a framework's topology w.r.t. Φ_H

2. $\forall_{a \in B}\, V'(a) = (x, y)$ is such that $\tau'(x, y) = \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a) \cap A}\; M(a_i)$.

We are aware that there is actually an infinity of possible $V'(a)$ possible, but we can arbitrate it to be any single one of those. They all represent the semantics equally well for the purposes of this dissertation.

An example of a double application of Φ_H can be seen in Figure 4.1. Notice that in the last figure no blue-coloured node attacks any other node. Their attacks have already been propagated into the nodes they attacked, but it had to be done in two steps.

From the definition of Φ_H above, we see that the $\Phi_H(F)$ has the exact same arguments as F and that all attacks originating from unattacked arguments have been removed because of preprocessing. The last step might not be so clear. Firstly, it is worth noting that the change to the voting function itself is a superficial step. What truly interests us at this point is the possibility of updating inherent argument strengths (initially the crowd support) by preprocessing certain attacks we know are invariant. We need only work with values from L_i, L_a and L_o to that purpose, but the extra step - converting back from an updated inherent strength to a vote function - is added for the sake of completeness. We simply impose V' to represent the updated strength value of the preprocessed arguments. This is also the only reason for which we require the semantics to be reversible, as we will see in the proof of Lemma 4.1.

It is crucial that we prove Φ_H to be semantic preserving, so that we may use it without compromising the meaning of the framework. We begin by formalising the result stating that an unattacked argument will always be assigned the same outcome.

Proposition 4.1. *Let* $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ *be a social argumentation framework and* $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ *a well-behaved semantics. Then,* $\forall_{a \in \mathcal{A}}\, \mathcal{R}^{-}(a) = \emptyset \Rightarrow \forall_{M,N \in \mathcal{M}}\, M(a) = N(a)$.

Proof. Let $a \in \mathcal{A}$ such that $\mathcal{R}^{-}(a) = \emptyset$. Then, by definition for any $M \in \mathcal{M}$,

$$\begin{aligned} M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} M(a_i) \\ &= \tau(a) \curlywedge \neg \bot_o \\ &= \tau(a) \end{aligned}$$

The above steps are possible by the definition of well-behaved semantic frameworks. $\tau(a)$ is a constant, and therefore not dependent on M. Therefore, $\forall_{M \in \mathcal{M}}\, M(a) = \tau(a)$, which in turn implies $\forall_{M,N \in \mathcal{M}}\, M(a) = N(a)$. □

We finally reach the first real result of this section, that states the semantic-preserving qualities of Φ_H.

Lemma 4.1 (Hubby Lemma)**.** *Let* $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ *be a reversible, well-behaved semantic framework. Then,* Φ_H *is a* $\mathcal{S}$*-preserving social argumentation mapping.*

Proof. Let $\Phi_H(F) = F' = \langle \mathcal{A}, \mathcal{R}', V' \rangle$, $A = \{a \in \mathcal{A} \mid \nexists_{b \in \mathcal{A}} (b, a) \in \mathcal{R}\}$ be the set of unattacked arguments, and $B = \{a \in \mathcal{A} \mid \mathcal{R}^{-}(a) \cap A \neq \emptyset\}$ be the set of arguments attacked by some unattacked arguments, as before.

By Proposition 4.1, $\forall_{a \in A} \forall_{M,N \in \mathcal{M}^F}\, M(a) = N(a) = \tau(a)$. Then, $\forall_{M \in \mathcal{M}^F} \forall_{b \in B}$,

$$\begin{aligned} M(b) &= \tau(b) \curlywedge \neg \curlyvee_{b_i \in \mathcal{R}^{-}(b)} M(b_i), \text{ which by commutativity} \\ &= \tau(b) \curlywedge \neg \left(\curlyvee_{b_i \in \mathcal{R}^{-}(b) \cap B} M(b_i) \curlyvee \curlyvee_{b_j \in \mathcal{R}^{-}(b) \cap \overline{B}} M(b_j) \right), \text{ by the D.M. laws} \\ &= \tau(b) \curlywedge \neg \Big(\curlyvee_{b_i \in \mathcal{R}^{-}(b) \cap B} M(b_i) \Big) \curlywedge \neg \Big(\curlyvee_{b_j \in \mathcal{R}^{-}(b) \cap \overline{B}} M(b_j) \Big) \\ &= \tau(b) \curlywedge \neg \Big(\curlyvee_{b_i \in \mathcal{R}^{-}(b) \cap B} \tau(b_i) \Big) \curlywedge \neg \Big(\curlyvee_{b_j \in \mathcal{R}^{-}(b) \cap \overline{B}} M(b_j) \Big) \\ &= \tau_b \curlywedge \neg \curlyvee_{b_j \in \mathcal{R}^{-}(b) \cap \overline{B}} M(b_j) \end{aligned}$$

The last step is merely a simplification in the notation. Before that, we isolate the arguments from A attacking b, and precompute their impact on b's inherent strength, $\tau(b)$. The result of the preprocessing of these attacks is stored in an updated inherent strength which we denote by τ_b. In other words, we have simplified models M of arguments $b \in B$ in framework F.

The purpose of the Φ_H is to emulate the above behaviour. The attacks that were precomputed above no longer appear in $\Phi_H(F) = F'$. Therefore, models of F' have the following form by definition:

$$M'(b) = \tau'(b) \curlywedge \neg \curlyvee_{b_j \in \mathcal{R}^{-}(b) \cap \overline{B}} M(b_j)$$

The forms of M and M' are remarkably similar. In fact, we wish to prove their equality, which can be reduced to the equality between τ_b and $\tau(b)$ as follows:

$$M(b) = M'(b)$$
$$\tau_b \curlywedge \neg \curlyvee_{b_j \in \mathcal{R}^-(b) \cap \overline{B}} M(b_j) = \tau'(b) \curlywedge \neg \curlyvee_{b_j \in \mathcal{R}^-(b) \cap \overline{B}} M(b_j)$$
$$\tau_b = \tau'(b)$$

By the reversibility of $\mathcal{S}$ and because $\tau_b \in L_o$, there are $x, y \in \mathbb{N}$ such that $\tau_b = \frac{x}{x+y}$. In other words, we need only to make it so $V'^{+}(b) = x$ and $V'^{-}(b) = y$. □

The most obvious use of this lemma is that any framework that has unattacked arguments may be simplified by an application of Φ_H. Moreover, the proof for Lemma 4.1 not only tells us that Φ_H is a $\mathcal{S}$-preserving social argumentation mapping, but also hints at there being new arguments which might now have a unique model, and which may be preprocessed themselves.

Corollary 4.1. *Let F and $F' = \Phi_H(F)$ be social argumentation frameworks, $\mathcal{S}$ a reversible well-behaved semantics, $A = \{a \in \mathcal{A} \mid \nexists_{b \in \mathcal{A}} (b, a) \in \mathcal{R}\}$ be the set of unattacked arguments and $B = \{b \in \mathcal{A} \mid \mathcal{R}^-(b) \subseteq A\}$ the set of arguments attacked* only *by unattacked arguments. Then, B has a unique model.*

Proof. For all $b \in B$,

$$\begin{aligned} M(b) &= \tau(b) \curlywedge \neg \curlyvee_{b_i \in \mathcal{R}^-(b)} M(b_i) \\ &= \tau(b) \curlywedge \neg \curlyvee_{b_i \in \mathcal{R}^-(b) \cap A} M(b_i) \\ &= \tau(b) \curlywedge \neg \curlyvee_{b_i \in \mathcal{R}^-(b) \cap A} \tau(b_i) \\ &= \tau_b \end{aligned}$$

Therefore, the model of all $b \in B$ is a constant, and B has a unique model. □

The above result suggests an iterative process may be used in order to simplify as much of a framework as possible. This is done under an umbrella of semantic equivalence, so no matter how many times Φ_H is applied, the final framework is still semantically equivalent to the original one. The use of such a procedure in any kind of implementation is clear, as it effectively reduces the size of the problem in an inexpensive way. We capture this iterative process in the following result.

Lemma 4.2. *Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework and $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ a reversible, well-behaved semantics. Then, $\Phi_H(F)$ has a fixpoint.*

Proof. Let $2^{\mathcal{R}}$ be the powerset of all attacks of $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$. Then, $2^{\mathcal{R}}$ is a complete lattice w.r.t. to $\subseteq$. Moreover, for any $F \in \mathcal{F}$ and $\Phi_H(F) = \langle \mathcal{A}, \mathcal{R}', V' \rangle$, $\mathcal{R}' \subseteq \mathcal{R}$. In order words, Φ_H is an order-preserving mapping over $2^{\mathcal{R}}$. In this situation, the Knaster-Tarski Theorem ensures the existence of a fixpoint of $\Phi_H(F)$. □

This result ensures that we can always iterate Φ_H up to a fixpoint, resulting in a preprocessing of all arguments indirectly attacked (and defended) by the unattacked arguments of the initial framework.

Let us now show our first result on an entire framework. We start with a simple result for well-founded frameworks, which by the absence of cycles will behave very nicely.

Theorem 4.1. *Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a well-founded social argumentation framework and $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ a reversible, well-behaved semantics. Then, F has a unique model.*

Proof. Let us prove by induction on the size n of the longest path that there is a semantically equivalent framework without attacks.

Basis: $n = 0$. If the length of the longest path in $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ is 0, then $\mathcal{R} = \emptyset$. Since $F \equiv_{\mathcal{S}} F$ trivially, the basis is verified.

Induction Step $n > 0$. The induction hypothesis: for any framework with a longest path $m < n$ there is a semantically equivalent framework without attacks.

The length of the longest path in F is n. Because the framework is well-founded, the longest directed path cannot be a cycle. Therefore, it has an origin argument a and a destination argument z. If a was attacked by an argument c, then the path of length $n+1$ from c to z would be the longest directed path, which is a contradiction. We conclude that a is unattacked. By the same reasoning, if $a\mathcal{R}b$ and b belongs to the longest directed path, then the path from any argument to b is at most of length 1.

Consider now $\Phi_H(F) = F' = \langle \mathcal{A}, \mathcal{R}', V' \rangle$. Then, $F' \equiv_{\mathcal{S}} F$. b belongs to the set of arguments attacked only by unattacked arguments. So, by the definition of Φ_H, $\nexists_{x \in \mathcal{A}} (x, b) \in \mathcal{R}'$. Since the reasoning holds for all paths of length n, then the length of the longest directed path in F' is $n - 1$. By induction hypothesis, there is $F'' = \langle \mathcal{A}, \mathcal{R}'', V'' \rangle$ such that $\mathcal{R}'' = \emptyset$ and $F \equiv_{\mathcal{S}} F' \equiv_{\mathcal{S}} F''$.

By Proposition 4.1 and because all arguments are unattacked in F'', then all arguments have a unique model in F''. It follows that F'' has a unique model, from which it follows that F' has a unique model, and similarly for F. □

We have thus proven that under well-behaved semantics, any well-founded framework has a unique model. This is a very strong description of the behaviour of acyclic frameworks. In fact, the result holds even for semantics which are not well-behaved. It depends only on the operators of the semantic framework being total, which we impose.

Clearly, these completely preprocessable arguments are not the source of multiple models. We have just seen that as long as there are no cycles, then there is a single model. Our objective is then to simplify and reduce frameworks as much as possible while maintaining semantic equivalence. We hope to identify the core reason that leads to multiple models.

Proposition 4.2 (Preprocessing)**.** *Let F be a social argumentation framework and $\mathcal{S}$ be a reversible, well-behaved semantic framework. Then, there is $F' \equiv_{\mathcal{S}} F$ that is in cyclic form.*

Proof. This proof will be virtually equivalent to that of Theorem 4.1. To reduce verbosity, let $ci(F)$ (resp. $cd(F)$) be the set of cycle-independent (resp. dependent) arguments in F, and let us call the longest path from any $a \in ci(F)$ to any $z \in cd(F)$ the "elected path". Let us prove by induction on the size n of the elected path, that there is a semantically equivalent framework in cyclic form, i.e. such that $\forall_{i \in ci(F)} \forall_{d \in cd(F)} (i, d) \notin \mathcal{R}$.

Basis: $n = 0$. If the length of the elected path is 0, then clearly $\forall_{i \in ci(F)} \forall_{d \in cd(F)} (i, d) \notin \mathcal{R}$. Otherwise, the length of the elected path would be at least 1. Since $F \equiv_{\mathcal{S}} F$ trivially, the basis is verified.

Induction Step $n > 0$. The induction hypothesis: for any framework whose elected path is of length strictly smaller than n, there is a semantically equivalent framework in cyclic form.

The length of the elected path in F is n. Because the framework is well-founded, the elected path cannot be a cycle. Therefore, it has an origin argument $a \in ci(F)$ and a destination argument $z \in cd(F)$. If a was attacked by $c \in ci(F)$, then the path of length $n + 1$ from c to z would be the elected path, which is a contradiction. We conclude that a is unattacked. By the same reasoning, if $a\mathcal{R}b$ and b belongs to the elected path, then the path from any argument $c \in ci(F)$ to b is at most of length 1.

Consider now $\Phi_H(F) = F' = \langle \mathcal{A}, \mathcal{R}', V' \rangle$. Then, $F' \equiv_{\mathcal{S}} F$. b belongs to the set of arguments attacked only by unattacked arguments. So, by the definition of Φ_H, $\nexists_{a \in \mathcal{A}} (a, b) \in \mathcal{R}'$. Since the reasoning holds for all paths of length n, then the length of the longest elected path in F' is $n - 1$. By induction hypothesis, there is $F'' = \langle \mathcal{A}, \mathcal{R}'', V'' \rangle$ such that $\forall_{i \in ci(F)} \forall_{d \in cd(F)} (i, d) \notin \mathcal{R}''$ and $F \equiv_{\mathcal{S}} F' \equiv_{\mathcal{S}} F''$. □

The Preprocessing Proposition (4.2) is the first materialisation of the ability to preprocess a well-founded part of a framework, an ability enabled by Φ_H. It is also an important step in the direction of stating that multiple models, if they arise, do so exclusively because of the existence of cycles. It was mentioned before that Φ_H could be used in an implementation to increase efficiency. This proposition tells us just how far it is possible to go: we must stop at cycle-dependent arguments.

We will also study arguments which do not belong to cycles, but which are nonetheless cycle-dependent. While it is impossible to guarantee that these arguments have a single model, we are able to prove that if they do not it is simply because of the cycles on which they depend. These cycles may have multiple models, a property which crosses over to arguments they attack and defend. We assume a framework which is already in cyclic form - if it was not, the Preprocessing Proposition would allow us to work on the semantically equivalent framework in cyclic form.

We will need to define another mapping for working with cycle-dependent arguments. The objective of this mapping is to sever the links between cycle-dependent arguments and cycles themselves. Let cyclic(F) be the set of arguments belonging to some cycle in F.

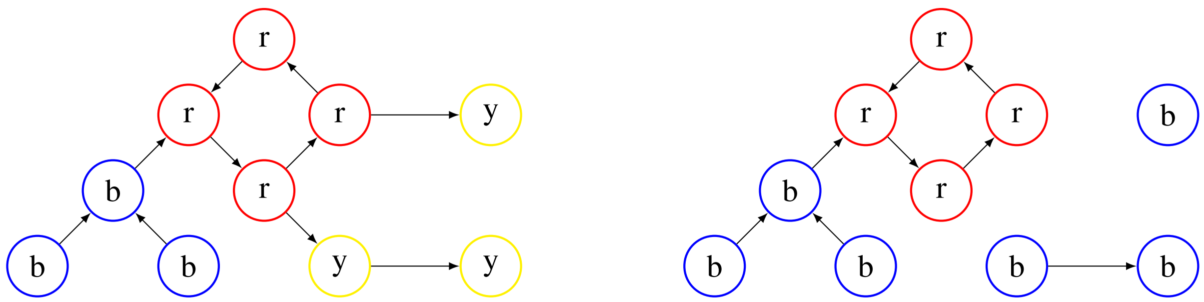
Definition 4.5. Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social argumentation framework, $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ a reversible, well-behaved semantics and $cd(F)$ the set of cycle-dependent arguments of F. Let Φ_W be a social argumentation mapping such that $\Phi_W(F) =$

$\langle \mathcal{A}', \mathcal{R}', V' \rangle$ where

- $\mathcal{A}' = \mathcal{A}$
- $\mathcal{R}' = \mathcal{R} \setminus \{(a,b) \in \mathcal{R} \mid a \in \text{cyclic}(F) \wedge b \in cd(F) \cap \overline{\text{cyclic}(F)}\}$
- For V', let $A = \{a \in \mathcal{A} \mid a \in cd(F) \cap \overline{\text{cyclic}(F)} \wedge \exists_{b \in \text{cyclic}(F)} (b,a) \in \mathcal{R}\}$ be the set of arguments attacked by some argument in a cycle.

  1. $\forall_{a \in \overline{A}} V'(a) = V(a)$.
  2. $\forall_{a \in A} V'(a) = (x,y)$ is obtained from $\tau'(x,y) = \tau(a) \curlywedge \neg \bigcurlyvee_{a_i \in \mathcal{R}^-(a) \cap \text{cyclic}(F)} M(a_i)$.

A single application of Φ_W is shown in Figure 4.2. Unlike Φ_H, Φ_W is meant to be used only once. Notice the new colour coding. There are no more yellow nodes, i.e. there are no more cycle-dependent nodes which do not belong to cycles.



(a) Recall Figure 3.4 (b) Figure 3.4 after an application of Φ_W

Figure 4.2: An example of the use of Φ_W

As mentioned before, we wish to prove that if the cycles have a single model, then so do all cycle-dependent arguments. Suppose a cycle had a several models. Because each model of the cycle interprets the cycle's strength (and thus, attack) differently, each argument attacked by that cycle would also have the same number of models. We thus prove our goal by assuming that cycles have a single model, and then checking that Φ_W is semantic preserving - that is to say, we can preprocess attacks originating from cycles to outside the cycle.

Proposition 4.3. *Let F be a social argumentation framework and $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ be a reversible, well-behaved semantic framework such that cyclic(F) has a unique model. Then, Φ_W is a $\mathcal{S}$-preserving social argumentation mapping.*

Proof. Let $\Phi_W(F) = F' = \langle \mathcal{A}, \mathcal{R}', V' \rangle$, $A = \{a \in \mathcal{A} \mid \exists_{b \in \text{cyclic}(F)} (b,a) \in \mathcal{R}\}$ be the set of arguments attacked by some argument in a cycle. Then, $\forall_{M \in \mathcal{M}^F} \forall_{a \in A}$,

$$
\begin{aligned}
M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^-(a)} M(a_i), \text{ which by commutativity} \\
&= \tau(a) \curlywedge \neg \left(\curlyvee_{a_i \in \mathcal{R}^-(a) \cap \text{cyclic}(F)} M(a_i) \curlyvee \curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j) \right), \text{ by the D.M. laws} \\
&= \tau(a) \curlywedge \neg \left(\curlyvee_{a_i \in \mathcal{R}^-(a) \cap \text{cyclic}(F)} M(a_i) \right) \curlywedge \neg \left(\curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j) \right) \\
&= \tau(a) \curlywedge \neg \left(\curlyvee_{a_i \in \mathcal{R}^-(a) \cap \text{cyclic}(F)} v_{a_i} \right) \curlywedge \neg \left(\curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j) \right) \\
&= \tau_a \curlywedge \neg \curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j)
\end{aligned}
$$

The last two steps are merely a simplification in the notation, made possible by arguments in cycles having a single model (v_{a_i}). Before that, we isolate the arguments belonging to cycles attacking a and precompute their impact on a. The result of the preprocessing of these attacks is denoted by τ_b.

The purpose of the Φ_W is to emulate the above behaviour. The attacks that were precomputed above no longer appear in $\Phi_W(F) = F'$. Therefore, models of F' have the following form by definition:

$$M'(a) = \tau'(a) \curlywedge \neg \curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j)$$

We wish to prove the equality of M and M', which can be reduced to the equality between τ_a and $\tau(a)$ as follows:

$$
\begin{aligned}
M(a) &= M'(a) \\
\tau_a \curlywedge \neg \curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j) &= \tau'(a) \curlywedge \neg \curlyvee_{a_j \in \mathcal{R}^-(a) \cap \overline{\text{cyclic}(F)}} M(a_j) \\
\tau_a &= \tau'(a)
\end{aligned}
$$

By the reversibility of $\mathcal{S}$ and because $\tau_a \in L_o$, there are $x, y \in \mathbb{N}$ such that $\tau_a = \frac{x}{x+y}$. In other words, we need only to make it so $V'^{+}(a) = x$ and $V'^{-}(a) = y$. □

We have just proven the $\mathcal{S}$-preserving properties of Φ_W but only in situations where cycles have unique models. After separating cycles from their dependents, the framework may no longer be in cyclic form, so we may apply the Preprocessing Proposition (4.2) once again to find another equivalent framework in cyclic form. That's the idea of the following proposition. We again assume an argumentation framework in cyclic form because we may obtain it through the preprocessing proposition.

Proposition 4.4 (Postprocessing). *Let F be a social argumentation framework in cyclic form and $\mathcal{S}$ a reversible, well-behaved semantics. If every cycle in F has a unique model under $\mathcal{S}$, then so does F.*

Proof. First, let $\Phi_W(F) = F'$. Then, $F \equiv_{\mathcal{S}} F'$. It is clear that F' is no longer in cyclic form, and that the now cycle-independent arguments are acyclic. That being the case, by the

Preprocessing Proposition (4.2), there is F'', such that $F' \equiv_{\mathcal{S}} F''$ and F'' is in cyclic form. By transitivity, $F \equiv_{\mathcal{S}} F''$. Now, the only cycle-dependent (or indeed attacked) arguments in F'' belong to cycles, as the other attacks were preprocessed away. The remaining arguments are unattacked, so that by Proposition 4.1 they have a unique model.

Every unattacked argument has a unique model, and by hypothesis every argument in a cycle has a unique model. There are no other arguments. Then, F'' has a unique model, and so does F. □

We have proven that if the a cycle does not have multiple models, then neither do any of the cycle-dependent arguments. In conjunction with the previous Preprocessing Proposition,

Theorem 4.2. *Let F be a social argumentation framework and $\mathcal{S}$ a reversible, well-behaved semantics. If every cycle in F has a unique model under $\mathcal{S}$, then so does F.*

Proof. By direct application of Propositions 4.2 and 4.4. □

This theorem is the central result of the current section. We have shown that the existence of many models depends exclusively on the existence of cycles. This is quite important in our context. The greatest practical advantage of this result is that we main restrict our focus to cycles and purely cyclic frameworks. Other arguments will not play a role in deciding whether there are multiple models or not. We will continue down this path in Section 4.2 when we look at specific semantics.

Secondarily, there is also the matter of optimising implementations, which has been referred to several times. Summarising, we have proven that any implementation can remove all cycle-independent arguments and all cycle-dependent arguments not belonging to a cycle and work exclusively with the cycles. Cycles exclusively determine whether there is no, one or multiple models. This by itself may already be an enormous reduction of the problem domain. In this dissertation we do not attempt an implementation, but it is comforting to know some semantic-preserving optimisations exist.

Finally, we suspect this theorem to hold even in the case where the semantic framework is not well-behaved. As long as the functions are total, then there is some kind of guarantee

4.1.2 Controversiality

As is usually the case, very little can be said about completely arbitrary frameworks. Generally speaking, in these cases, particular topological restrictions are important because they allow us to state something about the behaviour of the system in question. Well-founded frameworks are too much to hope for, and we find a weaker restriction in this notion of controversiality, which had already been presented and studied by Dung and has a very strong argumentation intuition.

We believe the examination of the behaviour of a generic semantic framework in an uncontroversial framework is quite relevant. Discussion is carried on by self-interested experts that must rely on the validity and congruence of their own arguments for their personal success. Some of them also have an intimate knowledge of rhetoric, and thus we speculate that controversiality will be much less common that cycles. Results in this section should thus be applicable to many real-life situations.

Controversiality is a straightforward concept from which an immediate topological result can be obtained.

Proposition 4.5. *If F is an uncontroversial social argumentation framework, then it does not contain odd-length cycles.*

Proof. Suppose F has an odd-length cycle $a_1, ..., a_n$. The cycle itself is a path of odd length, n, from a_1 to itself. Now consider the path of two circuits around the cycle, with $2n$ steps. $2n$ is even, and therefore there is a path from a_1 to itself with even length. Then a_1 is controversial with respect to itself, which is a contradiction with F being uncontroversial. There is thus no odd-length cycle in an uncontroversial framework. □

Unfortunately, even-length cycles do not enjoy of the same property. If they did, every uncontroversial framework would necessarily be well-founded, and consequently have a unique model. That would be too simple!

When dealing with cycles in Section 4.1.1, the objective was to figure out conditions for the existence of at most one model. The purpose of this section diverges significantly. The results found here are about the alterations in the models when the tiniest change is introduced in the framework. These changes should all be interpreted at the vote level, when another single vote is added. While τ will be used a lot in this section, it should always be seen as a reflection of the votes.

We will repeatedly analyse the same framework before and after a single (typically positive) vote. We assume a static framework where arguments and attacks never change. To represent the action of adding a vote, it would be possible to define a state transition mapping, similar to the social mappings found in the previous section. We decided against it. It is conceivable, for example, that a single pro or con vote can raise the number of models from one to infinite, or even reduce it to zero. These state transition functions would have to be able to cope with such vast changes in $\mathcal{M}$, and we believe the complexity of the representation would obfuscate the results.

This first proposition only serves the purpose of allowing us to find a path from the argument that was voted on to any argument whose model was altered.

Proposition 4.6. *Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, $F' = \langle \mathcal{A}, \mathcal{R}, V' \rangle$ be social argumentation frameworks, $\mathcal{S}$ a semantic framework and $a \in \mathcal{A}$ such that $\tau'(a) \neq \tau(a)$ and $\forall_{a_i \in \mathcal{A}\setminus\{a\}}\ \tau'(a_i) = \tau(a_i)$. Then $\forall_{M \in \mathcal{M}_\mathcal{S}^F}\ \exists_{M' \in \mathcal{M}_\mathcal{S}^{F'}}\ \forall_{b \in \mathcal{A}\setminus\{a\}}$, if $M(b) \neq M'(b)$ then there is a path from a to b.*

Next, we wish to show that if one increases the social support for a certain argument in an uncontroversial framework (meaning an argument cannot (indirectly) attack itself), then it is impossible for its model to decrease.

Conjecture 4.1. *Let* $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, $F' = \langle \mathcal{A}, \mathcal{R}, V' \rangle$ *be two uncontroversial social argumentation frameworks,* $\mathcal{S}$ *a well-behaved semantic framework and* $a \in \mathcal{A}$ *such that* $\tau'(a) \succeq \tau(a)$ *and* $\forall_{a_i \in \mathcal{A} \setminus \{a\}}\ \tau'(a_i) = \tau(a_i)$. *Then* $\forall_{M \in \mathcal{M}_\mathcal{S}^F}\ \exists_{M' \in \mathcal{M}_\mathcal{S}^{F'}}\ M(a) \preceq M'(a)$.

Remark 4.1 (Duality of results)**.** *Whenever statements of the form* $\tau'(a) \succeq \tau(a)$ *appear, where the intuition is that we are updating frameworks with a new vote, we are assuming that the dual result also applies.*

This conjecture remains unproven. There is much more than meets the eye in this proposition. We will attempt to show why. Recall the envisioned ODS. One option for the structure of the system is dividing it into stages. In a first stage, experts debate by proposing arguments and attacks between those arguments. Subsequently, in another stage, less expert users participate by voting. When that stage commences, intuitively F should be such that $\forall_{a \in \mathcal{A}}\, V^+(a) = V^-(a) = 0$. Every added vote starting from this state can be seen as a discrete event, distinct from all others. In this way, each vote event is a transition from a state of the framework to another, the state being defined by the votes.

If we can find an initial state s for an uncontroversial framework for which a model exists and Conjecture 4.1 holds, then *any* framework state derivable from s also has a model. This is an extremely powerful statement which is hidden behind an apparently obvious and trivial affair.

Delving a bit deeper, for the semantic frameworks we have proposed, when $V^+(a) = V^-(a) = 0$, $\tau(a) = \frac{0}{0}$ is an indeterminate form. Instead of resolving it, consider for argument's sake that the initial state for a framework is $\forall_{a \in \mathcal{A}}\, V^+(a) = 0, V^-(a) = 1$, which is capable of generating virtually the same framework states as the one where $V^+(a) = V^-(a) = 0$ for all $a \in \mathcal{A}$. For most semantic frameworks which derive some sort of social support from the social input we have that $\tau(a) = 0$ is this case. Perfect arguments are inexpressible from this starting state. On the other hand, it holds that $\lim_{V^+(a) \to \infty} \tau(a) = \frac{\infty}{\infty + 1} = 1$. In other words, the crowd has the power to arbitrarily approximate the situation where the argument is perfect.

We may then start with a framework where all arguments are false by default because of the justified assumption that there are no perfect arguments. This framework has a single model, where all arguments are false. Then, from state transitions, we may build any framework except one with perfect arguments. Conjecture 4.1 states that for every model in the framework before a state transition, there is one in the new framework. Then, every framework would have at least one model.

The above reasoning also holds for *any* starting framework. It is possible to start with an arbitrary social argumentation framework and arbitrarily approximate whichever social support values one wishes to obtain. In other words, we do not have to start with an argumentation framework where the arguments are false by default. We simply picked the framework with no perfect arguments closest to one where no votes had been placed.

We have the intuitive reasoning for our belief in this conjecture, which is intimately related with Theorem 4.3 below. Because of the monotonicity of $\curlywedge$, and the increase in $\tau(a)$, the only way for $M(a)$ to decrease is if there is a subset of $\mathcal{R}^-(a)$ of arguments whose model increased.

However, the framework is uncontroversial and thus there is no odd-length cycle which includes a. The behaviour of even-length cycles is typically well-understood. Generally speaking, if one increases the model of an argument c in an even-length cycle, then the model of all arguments in a directed path of odd length from c decreases. The model of all arguments in a directed path of even length from c increases, including c itself (for a more detailed explanation applicable to the Subtraction and Product semantics, see Section 4.2.2). Since the framework is uncontroversial, then c cannot attack itself, and thus its model does not decrease. Furthermore, if it belongs to an even cycle, its model is expected to increase as well.

The next result makes use of the conjecture.

Theorem 4.3 (Controversiality Theorem)**.** *Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, $F' = \langle \mathcal{A}, \mathcal{R}, V' \rangle$ be two uncontroversial social argumentation frameworks, $\mathcal{S}$ a well-behaved semantic framework and $a \in \mathcal{A}$ such that $\tau'(a) \succ \tau(a)$, $\forall_{a_i \in \mathcal{A} \setminus \{a\}}\ \tau'(a_i) = \tau(a_i)$ and Conjecture 4.1 holds. Then $\forall_{M \in \mathcal{M}^F_{\mathcal{S}}}\ \exists_{M' \in \mathcal{M}^{F'}_{\mathcal{S}}}$ if $a_1 \in \mathcal{A} \setminus \{a\}$ is indirectly attacked (resp. defended) by a then $M(a_1) \succ M'(a_1)$ (resp. $M(a_1) \prec M'(a_1)$).*

Proof. By absurdum, let us suppose that there is an argument $a_1 \in \mathcal{A}$ indirectly attacked by a and that $M(a_1) \preceq M'(a_1)$ (the reasoning for indirectly defended arguments is dual). Then,

$$\tau(a_1) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a_1)} M(a_i) \preceq \tau'(a_1) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a_1)} M'(a_i)$$

We know that $\tau(a_1) = \tau'(a_1)$ and that $\curlywedge$ is monotonic w.r.t. to both arguments. Therefore, in order for the above inequation to be verified it follows that

$$\neg \curlyvee_{a_i \in \mathcal{R}^{-}(a_1)} M(a_i) \preceq \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a_1)} M'(a_i)$$

and because $\neg$ is antitonic (if strictly antitonic, the inequation also becomes strict),

$$\curlyvee_{a_i \in \mathcal{R}^{-}(a_1)} M(a_i) \succeq \curlyvee_{a_i \in \mathcal{R}^{-}(a_1)} M'(a_i)$$

Once again, because $\curlyvee$ is monotonic w.r.t. both arguments, there is an argument $a_2 \in \mathcal{R}^{-}(a_1)$ such that

$$M(a_2) \succeq M'(a_2)$$

From a_2 to a_1, there is a path of length 1. By using this same reasoning, we will conclude that there is another argument $a_3 \in \mathcal{R}^{-}(a_2)$ such that

$$M(a_3) \preceq M'(a_3)$$

where there is a path of length 2 from a_3 to a_1.

More generally, for some $a_i \in \mathcal{A}$ in this path under consideration, if $M(a_i) \succeq M'(a_i)$ (resp. $M(a_i) \preceq M'(a_i)$), then that path from a_i to a_1 is of odd (resp. even) length. By Proposition 4.6, this path leads inevitably back to a, and so there is a path from a to a_1. By Proposition 4.1, we know that $M(a) \preceq M'(a)$ and thus that one of the paths from a to a_1 must be even.

By hypothesis, there is also a path from a to a_1 which is odd. a is thus controversial w.r.t a_1, which is a contradiction. Therefore, all arguments indirectly attacked by a in an uncontroversial framework must be such that $M(a_1) \succeq M'(a_1)$. □

This is the main result that characterises the behaviour of frameworks which are not well-founded but still have some topological features. In this case, we use controversiality because with it we can guarantee a certain monotonicity in the update of the framework. This theorem simply describes the changes in the models of arguments reachable from a. The changes are simple, and very intuitive. If the strength of an attacker goes up, the strength of attacked arguments should go down, and similarly for defended arguments.

It is uncontroversiality alone which guarantees this. Consider the framework in Figure 3.6 once again. As n increases, then so does the strength of the combined attack on c. It is quite possible that for a large enough n, the decrease in $M(b_1)$ through to $M(b_n)$ because of an increase in $M(a)$ is large enough to compensate the increase in the strength of the direct attack from a to c. An increase in $M(a)$ might thus be reflected by an increase in $M(c)$, which is directly attacked by a.

The importance of this section is that there is strong evidence to support the belief that 1) there is at least one model for every framework in the generic case, 2) that most real-life frameworks obtained from expert debates are uncontroversial, and 3) that the semantics behave very well in that case.

4.1.3 Reduction to Framework of Perfect Arguments

Our social argumentation frameworks have been extended with a function V containing the votes assigned to each argument. This was done in order to accommodate the context of an online debating system where casual users have the opportunity to participate.

Our semantics is based on this new voting data, following the intuition that for formal but truly subjective reasoning, social support takes a central place. We will demonstrate that our semantics can be emulated by other proposals that do not consider social support, subjective reasoning, or additional data and yet are a special case of our semantics.

The use of "reduction" in the title of this section is not to be taken literally. There is no reduction in the size of the frameworks - in fact, the opposite is true. By reduction we mean instead that our attempts aim to remove V from the argumentation framework while maintaining semantic equivalence. This is roughly the same as saying there is a semantically equivalent abstract argumentation framework, under some restricted semantics. In practise, we will simply try to impose that for all arguments $a \in \mathcal{A}, \tau(a) = 1$, or that $V^+(a) = 1$ and $V^-(a) = 0$.

Not only is the reduction not "for free" in terms of number of arguments, it also implies the complete loss of the social support intuition. By removing votes entirely and simulating them in another manner we have completely lost our context and its intuitions, hampering the subjective reasoning component. Near the end of this section we present the counterpoints to this reduction.

Before the main proof, we begin by defining $\mathcal{S}$-preserving Social Expansion Mappings, which are very similar to $\mathcal{S}$-preserving Social Argumentation Mapping. We name them expansion because the framework resulting from the mapping can have a larger number of arguments than the original one.

Definition 4.6 ($\mathcal{S}$-preserving social expansion mapping)**.** A social argumentation mapping Φ is a $\mathcal{S}$-preserving social expansion mapping iff $\forall_{F \in \mathcal{F}}$, $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$, $\Phi(F) = \langle \mathcal{A}', \mathcal{R}', V' \rangle$ with $\mathcal{A} \subseteq \mathcal{A}'$ and there is a bijection $f : \mathcal{M}^F \rightarrow \mathcal{M}^{\Phi(F)}$ such that $\forall_{a \in \mathcal{A}} \forall_{M \in \mathcal{M}^F} M(a) = f(M)(a)$.

Next, we define the restrictions on the semantic framework that allow the result to hold. As we have said before, we are able to make this reduction only for a specific family of semantic frameworks, and not in the general case. Furthermore, there are extra restrictions quite unlike the ones we have seen previously. We wish to assume every argument is the perfect argument (i.e. true by default), and as such if we allow the perfect argument to completely defeat other arguments, we forsake the capability to represent truth degrees other than True and False. The semantics must, in a way, be resistant to the idea of defeat.

Definition 4.7 (Defeat resistant semantics)**.** Let $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ be a semantic framework, $l_i \in L_i$ and $l_a \in L_a$ such that $l_a \succ \bot_o$. $\mathcal{S}$ is *defeat resistant* iff $l_i \succ l_i \curlywedge \neg \top_o \succ \bot_o$ and $l_i \succ l_i \curlywedge \neg l_a \succ \bot_o$

We now define the mapping which converts a social argumentation framework into an argumentation framework where all arguments are true by default.

Definition 4.8 (Reduction mapping)**.** Let $\mathcal{S}$ be a well-behaved, defeat resistant semantics and the Reduction Mapping $\Phi_R : \mathcal{F} \rightarrow \mathcal{F}$ be defined as

function $\Phi_R(F = \langle \mathcal{A}, \mathcal{R}, V \rangle)$:
$\forall_{a \in \mathcal{A}} \tau_a = \frac{V^+(a)}{V^+(a)+V^-(a)}$ (we store crowd support)
$A = \emptyset$ (a set for new arguments attacking arguments from original $\mathcal{A}$)
for all $a \in \mathcal{A}$ s.t. $\tau_a \prec \top_i \curlywedge \neg \bigcurlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} M(a)$ **do**
    while $\tau_a \prec \top_i \curlywedge \neg \bigcurlyvee_{a_i \in \mathcal{R}'^{-} \cap A(a)} M(a)$ **do**
        $A \leftarrow A \cup \{x\}$
        $\mathcal{A} \leftarrow \mathcal{A} \cup \{x\}$, where $x \notin \mathcal{A}$
        $\mathcal{R} \leftarrow \mathcal{R} \cup (x, a)$
        while $\tau_a \succ \top_i \curlywedge \neg \bigcurlyvee_{a_i \in \mathcal{R}^{-}(a) \cap A} M(a)$ **do**
            $\mathcal{A} \leftarrow \mathcal{A} \cup \{y\}$, where $y \notin \mathcal{A}$
            $\mathcal{R} \leftarrow \mathcal{R} \cup (y, x)$
        end while
    end while
end for
Let V s.t. $\forall_{a \in \mathcal{A}} V^+(a) = 1 \wedge V^-(a) = 0$
return $F' = \langle \mathcal{A}, \mathcal{R}, V \rangle$

We will now briefly explain the behaviour of the algorithm. The objective is to take a given framework, set all votes such that every argument has $\top_i$ as crowd support, and then for every argument whose original crowd support was not $\top_i$, add attacking arguments. Recalling the preprocessing idea, we add attackers in order to update the initial crowd support of $\top_i$ down to (or below) τ_a. In case the attack is excessive and it so happens that the updated crowd support goes below τ_a, the algorithm simply adds attackers to the attacker, thus reducing his impact on a.

This way, each iteration of the algorithm will approximate, by preprocessing, the models of the new framework to the models of the old. If not in a finite number of iterations, at least in infinity it will provide the exact desired value crowd support value.

We only need to prove that the mapping is semantic preserving.

Theorem 4.4 (Reduction Theorem). *Let $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ be a well-behaved, defeat resistant semantic framework. Then, Φ_R is a $\mathcal{S}$-preserving social expansion mapping.*

Proof. Let $F \in \mathcal{F}$ be an arbitrary framework and $\Phi_R(F) = F' = \langle \mathcal{A}', \mathcal{R}', V' \rangle$. From the definition of Φ_R, we know that $\mathcal{A}' = \mathcal{A} \cup A$ and $\mathcal{R}' = \mathcal{R} \cup R$ where A is well-founded. Furthermore, $\forall_{a \in \mathcal{A}'} \, \tau(a) = \top_o$. Since A is acyclic by definition, it has a unique model. Formally, $\forall_{a \in A} \forall_{M', N' \in \mathcal{M}^{F'}} \, M'(a) = N'(a)$.

Given the uniqueness of the model for A, the models for the rest of the framework as as follows. $\forall_{a \in \mathcal{A}} \forall_{M', \in \mathcal{M}^{F'}}$,

$$
\begin{aligned}
M'(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{-}(a)} \, M'(a_i) \\
&= \top_i \curlywedge \neg \Big(\big(\curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} \, M'(a_i) \big) \curlyvee \big(\curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap \overline{A}} \, M'(a_i) \big) \Big) \\
&= \top_i \curlywedge \neg \big(\curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} \, M'(a_i) \big) \curlywedge \neg \big(\curlyvee_{a_i \in \mathcal{R}^{-}(a)} \, M'(a_i) \big) \\
&= \top_i \curlywedge \neg \big(\curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} \, v_{a_i} \big) \curlywedge \neg \big(\curlyvee_{a_i \in \mathcal{R}^{-}(a)} \, M'(a_i) \big) \\
&= \tau_a \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} \, M'(a_i)
\end{aligned}
$$

On the other hand, $\forall_{a \in \mathcal{A}} \forall_{M \in \mathcal{M}^F}$,

$$M(a) = \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} \, M(a_i)$$

Reviewing, we currently have that for $M \in \mathcal{M}^F$ and $M' \in \mathcal{M}^{F'}$

$$
\begin{aligned}
M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} \, M(a_i) \\
M'(a) &= \tau_a \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} \, M'(a_i)
\end{aligned}
$$

It is clear then that if $\tau(a) = \tau_a = \top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} \, M'(a_i)$, the models for both frameworks are the same, and a bijection f exists trivially.

Let us analyse Φ_R. Line 4 repeats the procedure in Lines 4-14 for all arguments needing attackers. The loop in Line 5 may iterate an infinite number of times. By adding a new (perfect) attacking argument to a (Lines 6-8), the increasing monotonicity of $\curlyvee$ and antimonotonicity of $\curlywedge$ ensure that $\top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a) \cap A} M(a)$ decreases. Three situations are now possible,

1. $\tau(a) = \top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a)$, in which case the both conditions on Lines 5 and 9 fail, stopping the algorithm for this argument.

2. $\tau(a) \prec \top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a)$, in which case the attacker is not strong enough, so the condition on Line 9 fails, and another attacker is added by another iteration of the cycle on Line 5.

3. $\tau(a) \succ \top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a)$, in which case the attacker is too strong. Attackers are repeatedly added to the original attacker, until he is weakened enough.

Suppose n is the number of iterations of the loop from lines 5-13. Each iteration reduces the difference between $\tau(a)$ and $\top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a)$. Then, $\lim_{n \to \infty} \top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{\text{-}}(a) \cap A} \tau_a = \tau(a)$.

Then, $\forall_{a \in \mathcal{A}}$,

$$\begin{aligned} M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a_i) \\ M'(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M'(a_i) \\ M'(a) &= \tau_a \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M'(a_i) \\ M'(a) &= \top_i \curlywedge \neg \big(\curlyvee_{a_i \in \mathcal{R}'^{\text{-}}(a) \cap A} M'(a_i) \big) \curlywedge \neg \big(\curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M'(a_i) \big) \\ M'(a) &= \top_i \curlywedge \neg (\curlyvee_{a_i \in \mathcal{R}'^{\text{-}}(a)} M'(a_i)) \end{aligned}$$

The models coincide exactly, and the bijection exists. □

This result demonstrates that it is indeed possible to reduce a social argumentation framework to another where all arguments are labelled with maximum social support. We will call this framework the *representative* framework. Because arguments are "as true as possible by default", the intuitive consequence of the theorem above is that social argumentation frameworks can be equivalently replaced by representative abstract argumentation frameworks (i.e. by removing V altogether).

However, while interesting, this can only be done with the specific defeat resistant semantics. In any other situation, the reduction is impossible without reverting back to True and False arguments which is a great loss of expressiveness. This motivates the belief that the expressive power of our proposal is superior to that of abstract argumentation frameworks, even with multi-valued semantics. We will look at this matter again in later sections.

Furthermore, the conditions on the semantic framework imply that if $\exists_{a \in \mathcal{A}} \tau(a) = \bot_i$, then because $l_i \curlywedge \neg \top_o \succ \bot_o$ the representative framework will have an infinite number of attackers of a. This is another argument in favour of using our social argumentation frameworks.

Corollary 4.2. *Let F be a social argumentation framework and $\mathcal{S}$ be a defeat resistant, well-behaved semantic framework. If $\exists_{a \in \mathcal{A}} \tau(a) = \bot_o$, then $\Phi_R(F)$ is infinite.*

Proof. Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a semantic framework, $\Phi_R(F) = F' = \langle \mathcal{A}', \mathcal{R}', V' \rangle$ such that $\mathcal{A}' = \mathcal{A} \cup A$ and $a \in \mathcal{A}$ such that $\tau(a) = \perp_o$. By absurdum, suppose F' is finite. Then, $\mathcal{R}^{-}(a) = \{a_1, ..., a_n\}$ such that $n < +\infty$. Now, as before,

$$\begin{aligned} M'(a) &= \top_i \curlywedge \neg\big(\curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} M'(a_i)\big) \curlywedge \neg\big(\curlyvee_{a_i \in \mathcal{R}^{-}(a)} M'(a_i)\big) \\ M'(a) &= \tau_a \curlywedge \neg\big(\curlyvee_{a_i \in \mathcal{R}^{-}(a)} M'(a_i)\big) \\ \tau_a &= \top_i \curlywedge \neg\big(\curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} M'(a_i)\big) \end{aligned}$$

Or in other words, because of the defeat resistance of the semantics,

$$\begin{aligned} \tau_a &= \top_i \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A} M'(a_i) \\ &= \top_i \curlywedge \neg M'(a_1) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A \setminus \{a_1\}} M'(a_i) \\ &= l_1 \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A \setminus \{a_1\}} M'(a_i), \text{ with } l_1 \succ \perp_o \\ &= l_2 \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}'^{-}(a) \cap A \setminus \{a_1, a_2\}} M'(a_i), \text{ with } l_2 \succ \perp_o \\ &= ... \\ &= l_{n-1} \curlywedge \neg M'(a_n), \text{ with } l_{n-1} \succ \perp_o \\ \tau_a &= l_n, \text{ with } l_{n-1} \succ \perp_o \end{aligned}$$

Because of the semantic equivalence between F and $\Phi_R(F)$, $\tau_a = \tau(a)$. This is a contradiction because $\tau_a \succ \perp_o$ and by hypothesis $\tau(a) = \perp_o$. Therefore F' cannot be finite. It must contain an infinite number of arguments in order to defeat a totally. □

This corollary holds not only for Φ_R but for any method that attempts this kind of translation. If a semantics is defeat resistant then an infinite number of attackers is necessary to completely defeat any argument. Despite the generality of the result in that sense, the fact that perfect arguments are expected to be non-existent also applies to flawed arguments (with which *everyone* should disagree). A more important question is whether we are able to maintain finiteness of attackers for an arbitrary social support $l_i \in L_i$.

Unfortunately, the proof of Theorem 4.4 does not provide any answer regarding that question. Indeed, we know only that we may arbitrarily approximate the desired value by adding attacking arguments, ensuring that the theorem will hold in infinity. A deeper analysis of semantic frameworks would be necessary in order to state whether there are methods for obtaining a finite representative framework such that $\forall_{a \in \mathcal{A}} \perp_o \prec \tau(a) \prec \top_o$. This analysis is out of the scope and time bounds of this dissertation, and is thus left for future work.

In this section we have seen that with a restricted family of semantics it is possible to reduce our social argumentation frameworks to abstract argumentation frameworks. This shows that under certain conditions, the inclusion of voting can be seen as superfluous and does not bring any extra expressiveness. However, for semantics which are not defeat resistant, social support has no abstract argumentation representation. Neither of our specific semantic framework proposals is defeat resistant, and we do not believe that defeat resistance is the best choice for multi-valued semantics for argumentation. In these circumstances our system is more expressive, as it cannot be embedded or reduced to other proposals without making compromises (such as reducing output to True/False).

A softer consideration is that we lose the social support and voting intuitions with such a reducing approach. Representative frameworks are not particularly informative when it comes to the debates themselves. We become unaware of which arguments were actually proposed in discussion and which are just another layer for ensuring inherent argument strength is present. Furthermore, whenever a new vote is added the representative framework needs to be recomputed, so we can safely say that a representative framework is only a snapshot of a state of the framework.

As of yet, we are unaware of the computational difficulties that may lie in representing an arbitrary social support value by means of new attackers. It may very well be that there are conditions guaranteeing the finiteness of these attackers, and perhaps even efficient implementations. This is a very interesting topic to pursue if we wish to better understand these reductions and how they might be applied.

4.2 Specific semantics

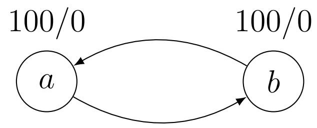
In this section we will initially present two conjectures and motivate them. The conjectures are related to the existence of a single or multiple models for the product semantics.

Next, we will be looking more closely at the two concrete semantics proposed in Section 3.4. We will be looking at several aspects of our concrete semantics on two different levels: the more global model level and the more local semantic framework level. At the model level we find that certain properties are only present for very specific semantics, and it becomes difficult to generalise the conditions of their existence. In Section 4.2.2 we will ascertain the manner in which models behave for cycles. This enables us to better understand the system when the underlying framework is not well-founded nor uncontroversial. In this section we will also attempt to differentiate the Subtraction from the Product semantics, as they appear to have fundamental differences.

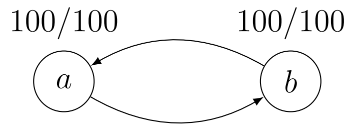
On a more local level, we will look at properties of the actual operators of the semantic frameworks in order to understand the source and core of the divergence between the two as identified in Section 4.2.2. For this purpose, we have identified two authors which provide us with a set of principles and properties for the operators. In Section 4.2.3 we look at the properties presented in [13] and verify whether both our semantics abide by them, and in Section 4.2.4 we will do the same for [25].

4.2.1 Conjectures

In this section we will be providing a very simple example that demonstrates undesirable behaviour in the presence of perfect arguments. Then, we will present two conjectures and motivate them.



(a) Mutually attacking perfect arguments



(b) Same arguments without perfect support

Figure 4.3: Mutually attacking arguments

Consider Figure 4.3a. For both arguments, $\tau(a) = \tau(b) = \frac{100}{100+0} = 1$, which means both of them are perfect. As for Figure 4.3b, $\tau(a) = \tau(b) = \frac{100}{100+100} = 0.5$, meaning they are both imperfect arguments. These are the simplest framework which are not well-founded.

In order not to repeat ourselves, we refer the reader further down to Proposition 4.9. In summary, it enables us to conclude that for the framework of Figure 4.3a, there exists an infinite number of models under $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$. In this particular situation, the equation system associated the two semantics are the same.

We now move on to the framework of Figure 4.3b. It falls once again under the conditions of Proposition 4.9, albeit with the inverse conclusion. $\mathcal{S}^{\times}$ produces a unique model for that framework. The Subtraction Semantics has an infinite number of models according to Proposition 4.8. For the envisioned online system, as discussed before, we are more interested in providing a single outcome than in showing the user an infinite number of solutions. We will be focusing on $\mathcal{S}^{\times}$ for now.

We have two conjectures that we believe to be true, but have not been able to prove formally.

Conjecture 4.2. *Any social argumentation framework has at least one model of F under $\mathcal{S}^{\times}$.*

We have used a simple fixpoint method to look at the behaviour of several types of frameworks containing controversiality and cycles. We initially assign each argument its crowd support, and then computes models based on that assignment. Then, we repeat the procedure with the updated values. We have empirically concluded that this method for the product semantics always approximates a fixpoint with one exception: when crowd support was set at 1 for all the arguments in a cycle. Even then, our calculations have shown there is some model for a variety of topologies. When cycles are involved, we have already seen that multiple models are more likely than none. The existence of at least one fixpoint obtained by iteration is a fairly good indication that there is at least one model.

Equation systems developed from frameworks with several interconnecting cycles have always had a solution, although the complexity of exploring them by hand explodes quite rapidly.

Recent unpublished work in non-monotonic formalisms has evidenced that Brauers' fixpoint theorem might be of use to us. This is left for future work.

We would also refer the reader back to the discussion around Conjecture 4.1, which is another piece of evidence in support of the existence of at least one model.

Conjecture 4.3. *Any social argumentation framework such that* $\forall_{a \in \mathcal{A}} \, \tau(a) < 1$ *has at most one model of* F *under* $\mathcal{S}^{\times}$*.*

Note that $\forall_{a \in \mathcal{A}} \, \tau(a) < 1 \Rightarrow \forall_{M \in F} \, M(a) < 1$. Thus, the problem that we have identified above does not hold in this situation. Together with the belief that perfect arguments are extremely unlikely, the conjecture should not have any problems dealing with the "real-world".

Once again, we have looked at many examples and analysed formulas derived from the equation systems of small but complex frameworks of embedded cycles and never found a hint that there might be a critical situation where the number of models grew.

Perhaps more convincingly, one of the main reasons for the existence multiple solutions of an equation system is when there are more variables than equations. We have exactly the same number of variables and equations - one for each argument - and strongly believe that linear dependence is very difficult (if not impossible) to obtain in the nonlinear system of equations derived from the product semantics.

4.2.2 Cycles

As stated in the introduction of this section, we will now analyse models in isolated cycles. We focus mainly on the issue of there being multiple models, although there are myriad results to be proven in this topic. The initial propositions are based on a framework with the structure (but not V!) found in Figure 4.3. The first result is of minor importance, but can be used to introduce an interesting situation.

Proposition 4.7. *Let* $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ *be a social argumentation framework such that* $a, b \in \mathcal{A}$, $\mathcal{R}^{\text{-}}(a) = \{b\}$, $\mathcal{R}^{\text{-}}(b) = \{a\}$ *and* $\tau(a) \neq \tau(b)$*. Then,* F *has an infinite number of models under* $\mathcal{S}^{\text{-}}$*.*

Proof. By definition, $M(a) = \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a_i)$, which for $\mathcal{S}^{\text{-}}$ is equivalent to $M(a) = \max\left(0, \tau(a) - \sum_{a_i \in \mathcal{R}^{\text{-}}(a)} M(a_i)\right)$. In this particular case, $M(a) = \max(0, \tau(a) - M(b))$. The equation system is thus the following

$$\begin{cases} M(a) = \max(0, \tau(a) - M(b)) \\ M(b) = \max(0, \tau(b) - M(a)) \end{cases}$$

Without loss of generality, assume that $\tau(a) > \tau(b)$. Let $l = \tau(a) - \tau(b) > 0$. Then there is an infinity of models such that $M(a) \in [\tau(a) - l, \tau(a)]$ and $M(b) = 0$ $\square$

So, we have proven that for the simplest, most common of cycles there is an infinite number of models under $\mathcal{S}^{-}$. This result is somewhat inconsequent due to the notion of model maximality. Since it is not a specially important part of this work, we will forgo a formal definition. Essentially, a maximal model is a model which is not dominated by any other model. We say that a model M dominates a model N iff $\forall_{a \in \mathcal{A}} M(a) \succeq N(a)$. With this definition, only one model proposed in the proof of Proposition 4.7 is maximal, namely when $M(a) = \tau(a)$. Justified by the postulate that arguments are always as true as possible, that is the only model that we need to consider, and others become irrelevant.

This reasoning can be found elsewhere, as in Antitonic Logic Programs and the lattice of interpretations [16] as well as in the concept of Pareto Frontier.

The next result targets the case not covered by the previous proposition: $\tau(a) = \tau(b)$.

Proposition 4.8. *Let F be a social argumentation framework and $\mathcal{S}^{-}$ such that $a, b \in \mathcal{A}$, $\mathcal{R}^{-}(a) = \{b\}$, $\mathcal{R}^{-}(b) = \{a\}$ and $\tau(a) = \tau(b)$. Then, F has an infinite number of models.*

Proof. As before, the equation system associated to the above framework is

$$\begin{cases} M(a) = \max(0, \tau(a) - M(b)) \\ M(b) = \max(0, \tau(b) - M(a)) \end{cases} \Leftrightarrow \begin{cases} M(a) = \max(0, \tau(a) - M(b)) \\ M(b) = \max(0, \tau(a) - M(a)) \end{cases}$$
$$\Leftrightarrow \begin{cases} M(a) = \tau(a) - M(b) \\ M(b) = \tau(a) - M(a) \end{cases}$$

The first step above is simply by hypothesis. The second one, however, requires further explanation. By the definition of model, it is clear that 1) no model can be negative and 2) the model's upper bound is the τ of the respective argument. This being the case, we know that the subtractions will never return a number below 0. The two equations are linearly dependent, and therefore the system has an infinite number of models. Models will simply be such that s $M(a) + M(b) = \tau(a)$.

An alternate proof is that since $\tau(a) = \tau(b) > 0$, then there are an infinity of $l_1 > 0$ and $l_2 > 0$ such that $l_1 + l_2 = \tau(a) = \tau(b)$. Therefore,

$$\begin{cases} M(a) = \max(0, \tau(a) - M(b)) \\ M(b) = \max(0, \tau(a) - M(a)) \end{cases} \Leftrightarrow \begin{cases} M(a) = \max(0, l_1 + l_2 - l_2) \\ M(b) = \max(0, l_1 + l_2 - l_1) \end{cases}$$
$$\Leftrightarrow \begin{cases} M(a) = l_1 \\ M(b) = l_2 \end{cases} \qquad \square$$

This result is quite different from Proposition 4.7, if not in its statement. There is no way around the infinite number of models, as before. All possible models belong to the Pareto Frontier, or are incomparable between themselves.

It should now be clear that the Subtraction Semantics suffers from a problematic issue. Whenever two arguments have the same crowd support and are in such a simple cycle, there is no unique model. It is true that the situation described in the previous propositions is somewhat improbable, but it serves as an indicator that more complex frameworks with larger cycles may also suffer from this problem.

As we have discussed before, multiple models are extremely undesirable for the simple reason that our objective is to give *the* solution of the framework. A multitude of solutions becomes incomprehensible to the users of the envisioned system.

We now divert our focus to the Product semantics, in order to identify whether it suffers from the same problem. We start once again with a framework structured like Figure 4.3.

Proposition 4.9. *Let F be a social argumentation framework and $\mathcal{S}^{\times}$ such that $a, b \in \mathcal{A}$, $\mathcal{R}^{-}(a) = \{b\}, \mathcal{R}^{-}(b) = \{a\}$. Then, F has an infinite number of models iff $\tau(a) = \tau(b) = 1$.*

Proof. The equation system derived from $\mathcal{S}^{\times}$ and just a and b is as follows.

$$\begin{cases} M(a) = \tau(a) \cdot (1 - M(b)) \\ M(b) = \tau(b) \cdot (1 - M(a)) \end{cases} \Leftrightarrow \begin{cases} M(a) = 1 - M(b) \\ M(b) = 1 - M(a) \end{cases} \Leftrightarrow \begin{cases} M(a) + M(b) = 1 \\ M(a) + M(b) = 1 \end{cases}$$

This system is clearly linearly dependent, as before. □

So, in the case where $\tau(a) = \tau(b) = 1$, the equation system for $\mathcal{S}^{\times}$ is very much like the one for $\mathcal{S}^{-}$, yielding an infinite number of solutions. The graph of $M(a)$ for this particular case is given by Figure 4.4. The problem is very clearly pin-pointed at $(1, 1)$, where a multitude of solutions can immediately be identified. Interestingly enough, as we approach $(1, 1)$, $M(a)$ approaches 0.5 exactly. This indicates that there might yet be a solution for these special cases, and for more complex ones as well.

Whatever the situation, the number of models continues to be an issue for the Product Semantics. It is not so crucial it is not chronic and present in other situations, as with $\mathcal{S}^{-}$. Let us examine the equation system associated with the previous proof more generally.

$$\begin{aligned} \begin{cases} M(a) = \tau(a) \cdot (1 - M(b)) \\ M(b) = \tau(b) \cdot (1 - M(a)) \end{cases} &\Leftrightarrow \begin{cases} M(a) = \tau(a) - \tau(a)M(b) \\ M(b) = \tau(b) - \tau(b)M(a) \end{cases} \\ &\Leftrightarrow \begin{cases} M(a) = \tau(a) - \tau(a)(\tau(b) - \tau(b)M(a)) \\ M(b) = \tau(b) - \tau(b)M(a) \end{cases} \\ &\Leftrightarrow \begin{cases} M(a) = \tau(a) - \tau(a)\tau(b) + \tau(a)\tau(b)M(a) \\ M(b) = \tau(b) - \tau(b)M(a) \end{cases} \\ &\Leftrightarrow \begin{cases} (1 - \tau(a)\tau(b))M(a) = \tau(a) - \tau(a)\tau(b) \\ M(b) = \tau(b) - \tau(b)M(a) \end{cases} \\ &\Leftrightarrow \begin{cases} M(a) = \dfrac{\tau(a) - \tau(a)\tau(b)}{1 - \tau(a)\tau(b)} \\ M(b) = \tau(b) - \tau(b)M(a) \end{cases} \end{aligned}$$

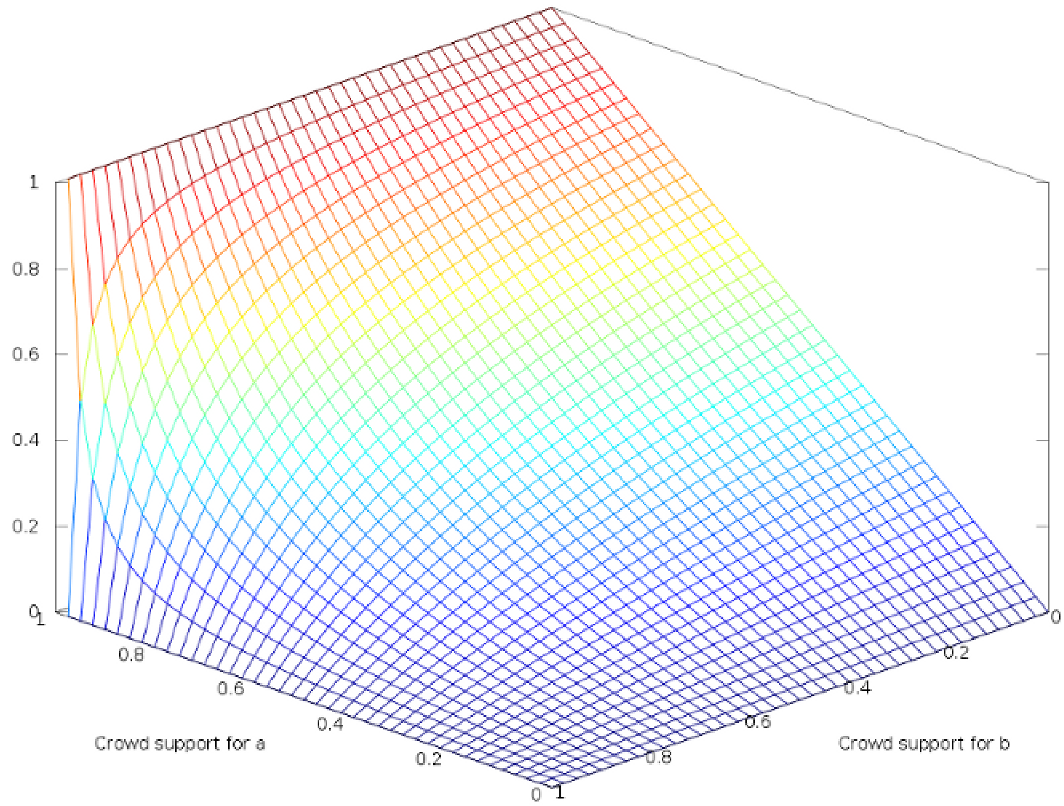


Figure 4.4: $M(a)$ dependent on $\tau(a)$ and $\tau(b)$

So the infinite number of models problem only ever appears when $\tau(a) = \tau(b) = 1$, which is the only case that is in indeterminate form. Any other values will result in a simple fraction with a single solution. This indicates that a larger number of more complicated frameworks will have single models under $\mathcal{S}^{\times}$ than under $\mathcal{S}^{-}$. For reasons previously stated, this is highly desirable.

We may generalise Proposition 4.9 for all even-length cycles. It holds that there is an infinite number of models not only in the simplest, perfect two-argument cycle, but also in any perfect even-length cycle.

Proposition 4.10. *Let F be a social argumentation framework and $\mathcal{S}^{\times}$ such that the sequence $a_1, ..., a_n \in \mathcal{A}$ is a cycle of even length, $\mathcal{R}^{-}(a_i) = \{a_{i-1}\}$ for $2 \leq i \leq n$ and $\mathcal{R}^{-}(a_1) = \{a_n\}$, and $\tau(a_i) = 1, 1 \leq i \leq n$. Then, F has an infinite number of models.*

Proof. The model for a_1 is:

$$\begin{aligned}M(a_1) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} M(a_i)\\ &= \tau(a) \curlywedge \neg M(a_2)\\ &= \tau(a) \cdot (1 - M(a_2))\\ &= 1 - M(a_2)\\ &= 1 - (1 - M(a_3))\\ &= M(a_3)\\ &= ...\\ &= M(a_1)\end{aligned}$$

So a model for F will assign the same value to all odd-numbered arguments in a even cycle. In the same way, it assigns the same value to all even-numbered arguments in that cycle. It is easy to see that the relation between the models of the odd- and even-numbered arguments is the same as that of Proposition 4.9, resulting in an infinite number of models. □

Interestingly, the same does not happen in the same situation for odd-length cycles.

Proposition 4.11. *Let F be a social argumentation framework and $\mathcal{S}^{\times}$ such that the arguments $\mathcal{A} = \{a_1, ..., a_n\}$ are a cycle of odd length, $\mathcal{R}^{-}(a_i) = \{a_{i-1}\}$ for $2 \leq i \leq n$ and $\mathcal{R}^{-}(a_1) = \{a_n\}$, and $\tau(a_i) = 1, 1 \leq i \leq n$. Then, F has a unique model.*

Proof. The model for a_1 is:

$$\begin{aligned}M(a_1) &= \tau(a) \cdot (1 - M(a_2))\\ &= 1 - M(a_2)\\ &= M(a_3)\\ &= ...\\ &= M(a_n), \text{ with } n \text{ odd.}\\ M(a_1) &= 1 - M(a_1)\end{aligned}$$

Now, from $M(a_1) = 1 - M(a_1)$ we get that $2M(a_1) = 1$ and finally that $M(a_1) = \frac{1}{2}$. □

A single model is a promising result - the ideal situation. It must be taken with a grain of salt, however. There is a very tight restriction on this cycle that every single argument must be evaluated the same way and have the same model. One must be wary of such severe bounds on the values of arguments because if they become too tight, we might end up with no model at all. Our research has shown that it does not appear to be the case, but it is a definite subject of interest and further study.

In this section we have provided an initial comparison between $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$. We show several sufficient conditions for the existence of multiple models in both semantics, with the more relaxed and undesirable conditions belonging to $\mathcal{S}^{-}$. The preliminary results presented here are thus evidence that the Subtraction Semantics is less fitting because it is guaranteed to have multiple models in more situations than the Product Semantics.

It would be very interesting to show that in an even cycle, the conditions stated in Proposition 4.10 are not only sufficient conditions but also necessary. Even if this is not the case, we conjecture that the slightest change to the perfect quality of the arguments is enough to break dependences between equations. This might lead to a proposition to the effect that any even-length cycle with an imperfect argument has a unique model.

There is certainly still a lot of work to be done for this section, but time constraints did not permit us to delve deeper. Yet even these propositions are enough to raise some intrigue as to the reason of the divergence between the two semantics. Why is the multiple model situation more symptomatic in $\mathcal{S}^{-}$ than in $\mathcal{S}^{\times}$? Judging from the need for maximal models for $\mathcal{S}^{-}$, there is clearly some inherent difference in the very way they are defined or designed - an underlying principle which is satisfied by one but not the other.

The two next subsections are dedicated to understanding the nature of this principle and of the general guidelines specifying the behaviour of each semantic framework.

4.2.3 Characterisation from Generic Gradual Valuation

In this section, we will look at some postulates and properties taken from [13]. We will restate them within the context of our generic semantics based on their principles, and show whether $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$ comply. The first step in this endeavour is to recall the properties of Generic Gradual Valuations, given in Section 2.4.

For the sake of intelligibility, the properties for a generic gradual valuation v and the respective g and h functions are repeated clearly here.

- $v(a) \geq V_{min}$.
- $v(a) = \top_o$ if $\mathcal{R}^{-}(a) = \emptyset$.
- $v(a) = g(h(v(a_1), ..., v(a_n)))$, where $\mathcal{R}^{-}(a) = \{a_1, ..., a_n\}$.

- $h(x) = x$.
- $h() = V_{min}$.
- For any permutation $(x_{i1}, ..., x_{in})$ of $(x_1, ..., x_n)$, $h(x_{i1}, ..., x_{in}) = h(x_1, ..., x_n)$.
- $h(x_1, ..., x_n, x_{n+1}) \geq h(x_1, ..., x_n)$.

- If $x_i \geq x'_i$, then $h(x_1, ..., x_i, ..., x_n) \geq h(x_1, ..., x'_i, ..., x_n)$.

- $g(V_{min}) = V_{max}$.
- $g(V_{max}) < V_{max}$.
- g is non-increasing (if $x \leq y$, then $g(x) \geq g(y)$).

In our notation, these properties become the following.

V1 $M(a) \succeq \bot_o$.

V2 $M(a) = \tau(a)$ if $\mathcal{R}^-(a) = \emptyset$.

V3 $M(a) = \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^-(a)} M(a_i)$, where $\mathcal{R}^-(a) = \{a_1, ..., a_n\}$.

H1 $\curlyvee_{x \in \{a\}} M(x) = M(a)$.

H2 $\curlyvee_{x \in \emptyset} M(x) = \bot_a$, the neutral element of disjunction.

H3 For any permutation $(x_{i1}, ..., x_{in})$ of $(x_1, ..., x_n)$, $x_{i1} \curlyvee ... \curlyvee x_{in} = x_1 \curlyvee ... \curlyvee x_n$.

H4 $x_1 \curlyvee ... \curlyvee x_n \curlyvee x_{n+1} \succeq x_1 \curlyvee ... \curlyvee x_n$.

H5 If $x_i \succeq x'_i$, then $x_1 \curlyvee ... \curlyvee x_i \curlyvee ... \curlyvee x_n \succeq x_1 \curlyvee ... \curlyvee x'_i \curlyvee ... \curlyvee x_n$.

G1 $l_i \curlywedge \neg\bot_o = l_i$, with $l_i \in L_i$.

G2 $l_i \curlywedge \neg\top_o \prec l_i$, with $\bot_i \prec l_i \in L_i$.

G3 $l_i \curlywedge \neg l_2 \succeq l_i \curlywedge \neg l_1$, with $l_1 \succeq l_2$, $l_i, l_1, l_2 \in L_a$.

We trust the translation of **V1-3** and **H1-5** and are clear enough. The interesting changes between the work done in [13] and our current proposal lie in the function g. Recall that arguments should be as true as possible by default, but that there is an upper limit given by the crowd support. It would not make sense for an argument to become stronger than the crowd makes it out to be. These ideas are the driving forces behind the translations of **G1-3**.

Property **G1** deals with the behaviour of the semantics when an argument is unattacked, or attacked by defeated arguments. In these situations, the semantics should assign these arguments as true a value as possible, by virtue of their being unaffected by incoming attacks. Because crowd support is an upper bound to an argument's model, property **G1** states that $\bot_a$ is the neutral element for $\curlywedge\neg$, which would be the equivalent of g if it took two arguments.

Property **G2** also features a small change for the same reasons, but perhaps a review of what the property states in its original form is also in order. Notice that the domain of g is not V, but $W \supseteq V$. In practise, this property is stating that if an argument is attacked by a single perfect argument (or set of arguments of equivalent strength), then it should not be considered perfect itself, i.e. it must be partially defeated. In our terms, it states that an argument under those conditions must have a model which is lower than its crowd support. It must be weakened, but nothing is said about defeat.

Property **G3** is not so dependent on the notion of crowd support. It is simply a property about the antimonotonicity of $\curlywedge\neg$ with respect to the second argument.

Proposition 4.12. *Any $\mathcal{S}$ and model M satisfy properties **V1** and **V3**. If $\neg\bot_a$ is the neutral element of $\curlywedge$, then property **V2** is satisfied.*

Proof. By definition. □

Property 4.1. *The Subtraction Semantics satisfies properties **V1-3**, **H1-5** and **G1-3**.*

Proof. **V1-3** are satisfied by Proposition 4.12. **H1** and **H2** are trivially satisfied by the definition of $\curlyvee$. **H3** is satisfied by the commutativity of addition, and **H4** and **H5** by the increasing monotonicity of addition. **G1** is satisfied because $\bot_o$ is 0, the identity element of subtraction. **G2** is satisfied because $\neg\top_o = -1$, and thus $\max(0, l_i - 1) < 0 \leq l_i$. **G3** is guaranteed by the antimonotonicity of subtraction with respect to the second argument. □

Property 4.2. *The Product Semantics satisfies properties **V1-3**, **H1-5** and **G1-3**.*

Proof. **V1-3** are satisfied by Proposition 4.12. **H1** and **H2** are trivially satisfied by the definition of $\curlyvee$. **H3** is satisfied by the commutativity of multiplication: $l_1 \curlyvee l_2 = 1 - (1 - l_1) \cdot (1 - l_2) = 1 - (1 - l_2) \cdot (1 - l_1) = l_2 \curlyvee l_1$. For **H4** and **H5** we need to demonstrate the monotonicity of $\curlyvee$. Let $l, l_1, l_2 \in L_a$ s.t. $l_1 \preceq l_2$. Then,

$$\begin{aligned} l_1 \preceq l_2 &\Leftrightarrow -l_1 \succeq -l_2 \\ &\Leftrightarrow 1 - l_1 \succeq 1 - l_2 \\ &\Leftrightarrow (1 - l_1) \cdot (1 - l) \succeq (1 - l_2) \cdot (1 - l) \\ &\Leftrightarrow -(1 - l_1) \cdot (1 - l) \preceq -(1 - l_2) \cdot (1 - l) \\ &\Leftrightarrow 1 - (1 - l_1) \cdot (1 - l) \preceq 1 - (1 - l_2) \cdot (1 - l) \\ &\Leftrightarrow l_1 \curlyvee l \preceq l_2 \curlyvee l \end{aligned}$$

So, $\curlyvee$ is increasing monotonic and properties **H4** and **H5** are satisfied. **G1** follows from $l_i \curlywedge \neg\bot_a = l_i \cdot (1-0) = l_i \cdot 1 = l_i$. **G2** follows from $l_i \curlywedge \neg\top_a = l_i \cdot (1-1) = l_i \cdot 0 = 0 \prec \top_a = 1$. **G3** is the antimonotonicity of $l_i \curlywedge \neg l_1$, which is obtained as follows

$$\begin{aligned} l_1 \succeq l_2 &\Leftrightarrow -l_1 \preceq -l_2 \\ &\Leftrightarrow 1 - l_1 \preceq 1 - l_2 \\ &\Leftrightarrow l_i \cdot (1 - l_1) \preceq l_i \cdot (1 - l_2) \\ &\Leftrightarrow l_i \curlywedge \neg l_1 \preceq l_i \curlywedge \neg l_2 \end{aligned}$$

□

	V1	V2	V3	H1	H2	H3	H5	G1	G2	G3
$\mathcal{S}^{\text{-}}$	√	√	√	√	√	√	√	√	√	√
$\mathcal{S}^{\times}$	√	√	√	√	√	√	√	√	√	√

Table 4.1: Satisfaction of Properties from Generic Gradual Valuations

The results of this evaluation may be found in Table 4.1. As it turns out, there is no difference between the two semantics according to this particular characterisation. Both semantics satisfy the same operator-centric principles, which is slightly strange considering the distinction between L_a and both L_i and L_o that occurs in $\mathcal{S}^{\text{-}}$ but not in $\mathcal{S}^{\times}$.

After this description, we would like to point out that this characterisation in particular is very focused on the operators themselves, but does not give much insight as regards the expected models of a complex framework. It is definitely important in the sense that simple, well-founded frameworks (which typically depend only on direct applications of these operators) should always respect the undeniable intuitions and underlying principles of the presented properties.

In order to identify the source of the discrepancy between the two semantics, we felt the need to push forward and look at another characterisation.

4.2.4 Characterisation from Defeasible Reasoning with Variable Degrees of Justification

This section is dedicated to doing with [25] what we have done to [13] in Section 4.2.3. We will use properties of the ♦ operator in Pollock's work to further understand the behaviour of our semantics.

According to Pollock, the operator $a \blacklozenge b$ can be read as a is attacked by b. This statement, in our notation, can be seen as $M(a) = \tau(a) \curlywedge \neg M(b)$. We relate our operators to Pollock's by imposing simply that $\blacklozenge = \curlywedge\neg$.

Like before, we will re-state the original properties in a first stage and then follow up with a modification intended for our context.

- $\blacklozenge$ is continuous on the interval $[\bot_L, \top_L]$, or just $[\bot, \top]$ for clarity.
- If $\top > a > b > \bot$, then $a > a \blacklozenge b > \bot$. A non-tautological argument will diminish the strength of those it attacks without destroying them completely.
- If $\top > a > b > c > \bot$, then $a \blacklozenge b < a \blacklozenge c$ and $a \blacklozenge c > b \blacklozenge c$. $\blacklozenge$ is monotonic.
- If $\top \geq a \geq b \geq \bot$ then $b \blacklozenge a = \bot$.
- If $\top \geq a > \bot$, then $a \blacklozenge \bot = a$.
- If $\top > a$, $\top \geq b \geq \bot$ and $\top \geq c \geq \bot$, then $(a \blacklozenge b) \blacklozenge c = (a \blacklozenge c) \blacklozenge b$.

Let $l_1, l_2, l_3 \in L_o$. In our context, Pollock's properties become

P1 $l_1 \curlywedge \neg l_2$ is continuous on the interval $[\bot, \top] \times [\bot, \top]$.

P2 If $\top \succ l_1 \succ l_2 \succ \bot$, then $l_1 \succ l_1 \curlywedge \neg l_2 \succ \bot$.

P3 If $\top \succ l_1 \succ l_2 \succ l_3 \succ \bot$, then $l_1 \curlywedge \neg l_2 \prec l_1 \curlywedge \neg l_3$ and $l_1 \curlywedge \neg l_3 \succ l_2 \curlywedge \neg l_3$.

P4 If $\top \succeq l_1 \succeq l_2 \succeq \bot$ then $l_2 \curlywedge \neg l_1 = \bot$.

P5 If $\top \succeq l_1 \succ \bot$, then $l_1 \curlywedge \neg \bot = l_1$.

P6 If $\top \succ l_1$, then $(l_1 \curlywedge \neg l_2) \curlywedge \neg l_3 = (l_1 \curlywedge \neg l_3) \curlywedge \neg l_2$.

First of all, an understanding of the meaning of these properties is desirable. **P1** is fairly self-explanatory, and also desirable. It is desirable that a single vote does not have a far-reaching, significant impact on the models of other arguments. Indeed, a single vote should not completely alter the outcome of a discussion. One step towards this situation is taken by forcing continuity of the operators. At the very least, simple frameworks with continuous operators should have no suffer no such symptoms.

Property **P2** states that if an argument b attacks a stronger argument a, then a should not be completely defeated. It is important to note that nothing is said here about multiple attackers. It is perfectly plausible that several attackers can completely defeat an argument when a single one could not.

Property **P3** deals simply with the monotonicity ⧫ with respect to its first argument and antimonotonicity with respect to its second argument. If the crowd support of an argument increases, its model should too, and if attackers become stronger, then it should decrease.

Property **P4** says that an attacker which is stronger than the attacked argument will simply defeat that argument. Property **P5** places $\neg \bot_o = -0 = 0$ as the neutral element of $\curlywedge$. Finally, Property **P6** is the commutativity of $\curlywedge\neg$.

Notice that the above properties are relevant to our study because of the generalized De Morgan laws and commutativity of our operators. They guarantee that an argument's definition of model may be rewritten as a series of individual attacks considered separately from one another. Formally,

$$M(a) = \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^-(a)} M(a_i) = \tau(a) \curlywedge \neg M(a_1) \curlywedge ... \curlywedge \neg M(a_n)$$

We may now proceed confident that although the context of Pollock's ⧫ is rebutting attacks (implying cycle of only two arguments), his properties are perfectly applicable to our operators, which consider more than just rebutting attacks.

Property 4.3. *The Subtraction Semantics satisfies properties **P1-6**.*

Proof. Because by hypothesis $l_1 > l_2$, any subtraction of the form $l_1 - l_2$ will always result in a positive number. For this reason, it is possible to say that $l_1 - l_2 = \max(0, l_1 - l_2)$. For **P1**, $l_1 \curlywedge \neg l_2 = \max(0, l_1 - l_2)$, which is trivially continuous. For **P2**, $l_1 \curlywedge \neg l_2 = max(0, l_1 - l_2) > 0 = \bot$. Similarly, because $l_2 > 0$, then $l_1 > l_1 - l_2 \Leftrightarrow l_1 > \max(0, l_1 - l_2) \Leftrightarrow l_1 \succ l_1 \curlywedge \neg l_2$. For the first condition of **P3**,

$$\begin{aligned} l_2 > l_3 &\Leftrightarrow -l_2 < -l_3 \\ &\Leftrightarrow l_1 - l_2 < l_1 - l_3, \text{ and since } l_1 > l_2 \text{ and } l_1 > l_2, \\ &\Leftrightarrow \max(0, l_1 - l_2) < \max(0, l_1 - l_3) \\ &\Leftrightarrow l_1 \curlywedge \neg l_2 \prec l_1 \curlywedge \neg l_3 \end{aligned}$$

Similarly, for the second condition of **P3**,

$$\begin{aligned} l_1 > l_2 &\Leftrightarrow l_1 - l_3 > l_2 - l_3, \text{ and since } l_1 > l_3 \text{ and } l_2 > l_3, \\ &\Leftrightarrow \max(0, l_1 - l_3) < \max(0, l_2 - l_3) \\ &\Leftrightarrow l_1 \curlywedge \neg l_3 \prec l_2 \curlywedge \neg l_3 \end{aligned}$$

For **P4**, $l_2 < l_1 \Leftrightarrow l_2 - l_1 < 0$. Therefore, $\max(0, l_2 - l_1) = 0 \Leftrightarrow l_2 \curlywedge \neg l_1 = \bot$. For **P5**, $l_1 \curlywedge \neg\bot = l_1 - 0 = l_1$. Finally, for **P6**, $(l_1 \curlywedge \neg l_2) \curlywedge \neg l_3 = \max(0, \max(0, l_1 - l_2) - l_3)$. There are two possible situations now:

1. Suppose that $l_1 \geq l_2 + l_3$. in which case $\max(0, \max(0, l_1 - l_2) - l_3) = l_1 - l_2 - l_3 = l_1 - l_3 - l_2 = \max(0, \max(0, l_1 - l_3) - l_2) = (l_1 \curlywedge \neg l_3) \curlywedge \neg l_2$.

2. Suppose that $l_1 < l_2 + l_3$. in which case $\max(0, \max(0, l_1 - l_2) - l_3) = \max(0, \max(0, l_1 - l_3) - l_2) = 0$.

□

The Subtraction Semantics once again satisfies all properties set forth by Pollock. The following result tells us whether the same happens for the Product Semantics.

Property 4.4. *The Product Semantics satisfies properties **P1-3** and **P5-6**, and does not satisfy **P4**.*

Proof. The proofs are made over a composition of $\curlywedge$ and $\neg$ as follows $l_1 \curlywedge \neg l_2$, which in $\mathcal{S}^\times$ becomes $l_1 \cdot (1 - l_2)$. We will use this form directly.

- **P1**: $l_1 \curlywedge \neg l_2 = l_1 \cdot -l_2 = -l_1 \cdot l_2$, which is trivially continuous.

- **P2**: for any number x, and $y \in]0, 1[$, $x \cdot y < x$. Furthermore, for $z \in]0, 1[$, $\neg z = 1 - z \in]0, 1[$. Then, $l_1 \curlywedge \neg l_2 = l_1 \cdot (1 - l_2)$, and because $(1 - l_2) \in]0, 1[$, then $l_1 > l_1 \cdot (1 - l_2) \Leftrightarrow l_1 \succ l_1 \curlywedge \neg l_2$.

- **P3**: $l_2 > l_3 \Leftrightarrow -l_2 < -l_3 \Leftrightarrow 1 - l_2 < 1 - l_3$, and because of the monotonicity of the product, $l_1 \cdot (1-l_2) < l_1 \cdot (1-l_3) \Leftrightarrow l_1 \curlywedge \neg l_2 \prec l_1 \curlywedge \neg l_3$. Similarly, because $1-l_3 \in]0,1[$, $l_1 > l_2 \Leftrightarrow l_1 \cdot (1-l_3) > l_2 \cdot (1-l_3) \Leftrightarrow l_1 \curlywedge \neg l_3 \succ l_2 \curlywedge \neg l_3$.

- **P4**: the only zero element (recall that $\perp = 0$) is 0 itself, and therefore the only way that $l_2 \curlywedge \neg l_1 = \perp$ is if $1 - l_1 = 0$. However, l_1 may not be the perfect argument, and thus may take values other than 1. Potentially, $1 - l_1 > 0$, which violates this property. This property is *not* satisfied.

- **P5**: $l_1 \curlywedge \neg\perp = l_1 \cdot (1 - 0) = l_1 \cdot 1 = l_1$.

- **P6**: we must formally show that $\curlywedge\neg$ is commutative. We do so as follows,

$$\begin{aligned}(l_1 \curlywedge \neg l_2) \curlywedge \neg l_3 &= (l_1 \cdot (1 - l_2)) \cdot (1 - l_3)\\ &= l_1 \cdot (1 - l_3) \cdot (1 - l_2)\\ &= (l_1 \cdot (1 - l_3)) \cdot (1 - l_2)\\ &= (l_1 \curlywedge \neg l_3) \curlywedge \neg l_2\end{aligned}$$

□

A summary of the satisfaction of these properties by $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$ can be found in Table 4.2. The property which finally appears to identify the source of the deviation between the two semantics is **P4**. Let us look more carefully at its nature. We have seen before that it states that an attacker stronger than the attacked argument is capable of defeating it. This is precisely the core of the discussion about the Perfect Argument presented in Section 3.1.6. In that Section we motivated why only the perfect argument should be able to defeat other arguments. One of the corollaries of such a choice is that there is no imperfect argument capable of defeating another argument, and this goes against one of Pollock's principles.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**
$\mathcal{S}^{-}$	√	√	√	√	√	√
$\mathcal{S}^{\times}$	√	√	√		√	√

Table 4.2: Satisfaction of Properties from Defeasible Reasoning with Variable Degrees of Justification

The main difference between attacks in the Product Semantics and the Subtraction semantics is that the Subtraction Semantics considers the strength of an attack to be an absolute value. An attack will simply subtract the strength of the attacker to the strength of the defender, regardless of the strengths of each. The Product semantics, on the other hand, uses a relative notion of attack. The amount of strength that an attack takes away from the attacked argument is inherently proportional to its strength before the attack.

Perhaps an illustration of this concept is ideal, in order to understand the difference. Let us go back to the example in Section 3.4.3 and Figure 3.7. As seen before, the framework is well-founded, and thus has a unique model. Recall that $M(b) = M(c) = 0.25$, $\tau(a) = 1$ and $\tau(d) = 0.5$.

If the strength of an attack is an absolute value, then 1) a should remain undefeated because the conjunction of its attackers does not exceed its social support, and 2) d should be defeated because the conjunction of its attackers exceeds (more specifically, equals) its social support. Indeed, under the Subtraction Semantics, $M(a) = 0.5$ and $M(b) = 0$. The outcome of the debate is then that the beach is a much better place to be than the park. In fact, the park is not even an option, since it is assigned as being completely false. The 50 users to believe the park to be a good place to go in the Summer are utterly disregarded in the outcome of the debate. This is undesirable, as everyone's opinions should matter or have some sort of influence on the debate. As it stands, the outcome would be exactly the same without those votes.

For the Product Semantics, attacks are not considered to be absolute. They are relative to the attacker's strength. The model of b and c remains the same, but $M(a) = 0.5625$ and $M(b) = 0.28125$, and every single vote counted towards this result. The fact that the reduction carried out by an attack from an imperfect argument is always relative to the argument's strength makes it so that a vote can never be discarded. Every vote has some impact, even if very small.

In these last two sections we focused on identifying the core difference between our two proposed semantics because of the perceived different behaviours regarding multiple models. We looked specifically at ways to characterise our semantic frameworks (and semantics in general!) to that purpose. This systematic approach consisted of two sets of properties considered to be intuitively desirable.

The first set of properties were quite local and specific to the operators themselves. It focused on independent behaviour of the operators. All of our proposed operators complied to these principles, which gave us some confidence that we were not failing any basic prerequisites. However, this independent view of each operator was not enough to identify the reason of divergence between our two semantics.

We then looked to another set of principles which considered the operators as a whole. Besides providing even more solid footing for our operators, it was this holistic approach that was the key to figuring out the fundamental difference between our semantics. As it turns out, attacks can be considered to be absolute or relative to the attacked argument. The absolute approach makes it much easier for votes to be discarded as unnecessary and without effect to the final outcome. The relative approach, on the other hand, takes every single vote into account and makes it have an impact, no matter how small, in the rest of the framework. Because representability of all participants is a very important component of our system, this comparison is just another reason to prefer the Product semantics over the Subtraction semantics.

5 . Comparison

The purpose of this chapter is to study how our proposal relates to others found in the literature. The identification of embeddings from other proposals into ours or vice-versa is of extreme importance. It allows us to truly evaluate the validity, context, localisation and impact of this work.

In Section 5.1 we will attempt to find conditions for the coincidence between some of Dung's classic semantics [17] and our own.

In Section 5.2 we will do the same for [13]. This second comparison of is of more importance. It will define whether we are truly generalising the work done in [13] or simply modifying it to emcompass subjective reasoning more explicitly, maintaining the expressive power of Generic Gradual Valuations. Some of questions already have partial answers from other sections of the dissertation.

5.1 Dung's Abstract Argumentation Frameworks

This section is mostly a curiosity in trying to find whether some kind of relation between Dung's classic work [17] and ours might still subsist. Another source of interest is their incredibly different nature and purpose. Dung's work is extension (i.e. set) based, and thus binary in nature. Ours is labelling-based, without depending on sets of arguments to verify the acceptability of arguments.

Unsurprisingly, the result of this section is about well-founded frameworks, where the problem-inducing cycles are not present and everything behaves nicely.

Theorem 5.1. *Let AF* $= \langle \mathcal{A}, \mathcal{R} \rangle$ *be a well-founded Abstract Argumentation Framework,* $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ *be a Social Argumentation Framework. Let* $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ *and* V *such that*

- $\top_i = \top_o$
- $\forall_{x \in \mathcal{A}}\ \tau(x) = \top_i$.
- $l_i \curlywedge \neg \bot_a = l_i$*, with* $l_i \in L_i$ *(recall that* $L_i \subseteq L_o$*).*
- $l_i \curlywedge \neg \top_o = \bot_o$*, with* $l_i \in L_i$.
- $\top_o \curlyvee \top_o = \top_o$ *and* $\top_o \curlyvee \bot_a = \top_o$.

Then, $S = \{x_1, ..., x_n\}$ *is the unique extension of AF iff* $\mathcal{M} = \{M\}$ *and* $a \in S \Leftrightarrow M(a) = \top_o$ *and* $a \notin S \Leftrightarrow M(a) = \bot_o$.

Proof. Because *AF* is well-founded, then by Theorem 4.1, F has a single model. For the same reason, there is a single extension of *AF* which is preferred, stable, complete and grounded [17]. Let us focus on the grounded extension definition as the fix-point of a characteristic function $F_{AF}(S) = \{a \mid a \in \mathcal{A}, \forall_{b \in \mathcal{R}^{-}(a)} \exists_{c \in S} \text{ s.t. } c\mathcal{R}b\}$. F_{AF} is monotonic w.r.t. set inclusion, ensuring the existence of a grounded extension in every framework. We will use F_{AF} for a proof by induction on the membership on the unique extension. Let us denote by M the unique model of F.

For the $\Rightarrow$ direction of the proof:

Basis: $S = \emptyset$. Any $a \in F_{AF}(S)$ is unattacked and belongs to the unique extension of *AF*. By the definition of model,

$$\begin{aligned} M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} M(a_i), \text{ which with } \mathcal{R}^{-}(a) = \emptyset \\ &= \top_i \curlywedge \neg \bot_a \\ &= \top_o \end{aligned}$$

Induction Step: $\emptyset \subset S \subseteq \mathcal{A}$. Let $a \in F_{AF}(S) \setminus S$. Then, by the definition of $F_{AF}(S)$, $\forall_{b \in \mathcal{R}^{-}(a)} \exists_{c \in S}\ c\mathcal{R}b$. By induction hypothesis, $M(c) = \top_o$. The model for every b is as follows

$$\begin{aligned} M(b) &= \tau(b) \curlywedge \neg \curlyvee_{b_i \in \mathcal{R}^{-}(b)} M(b_i) \\ &= \top_i \curlywedge \neg \left(M(b_1) \curlyvee \ldots \curlyvee M(b_i) \curlyvee \ldots \curlyvee M(b_n)\right), \text{being } b_i = c \in S \\ &= \top_i \curlywedge \neg \left(M(b_1) \curlyvee \ldots \curlyvee \top_o \curlyvee \ldots \curlyvee M(b_n)\right) \\ &= \top_i \curlywedge \neg \top_o \\ &= \bot_o \end{aligned}$$

We have proven that the model of any $c \in \mathcal{R}^{-}(a), a \in S' \setminus S$ is $M(c) = \bot_o$. As for $a \in S' \setminus S$ itself,

$$\begin{aligned} M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{-}(a)} M(a_i) \\ &= \top_i \curlywedge \neg \left(\bot_o \curlyvee \ldots \curlyvee \bot_o\right) \\ &= \top_i \curlywedge \neg \bot_o \\ &= \top_o \end{aligned}$$

By induction, every $a \in F_{AF}(S)$, $M(a) = \top_o$. From this, we conclude that the model of every argument in the unique extension of *AF* is $\top_o$.

For the $\Leftarrow$ direction, we prove by absurdum. Suppose there is an argument $a \in \mathcal{A}$ such that $M(a) = \top_o$ and which is not in the unique extension, S. Because *AF* is well-founded, $a \notin S$

because $\exists_{b \in S}\ b\mathcal{R}a$. The model for a is thus

$$
\begin{aligned}
M(a) &= \tau(a) \curlywedge \neg \curlyvee_{a_i \in \mathcal{R}^{\cdot}(a)} M(a_i) \\
&= \top_i \curlywedge \neg \left(M(a_1) \curlyvee \ldots \curlyvee M(a_i) \curlyvee \ldots \curlyvee M(a_n)\right), \text{being } a_i = b \in S \\
&= \top_i \curlywedge \neg \left(M(a_1) \curlyvee \ldots \curlyvee \top_o \curlyvee \ldots \curlyvee M(a_n)\right) \\
&= \top_i \curlywedge \neg \top_o \\
&= \bot_o
\end{aligned}
$$

However, we have already proven that since $a \in S$, then $M(a) = \top_o$. This is a contradiction. Therefore, a must belong to the unique extension of *AF*.

For the second part of proof (i.e. $a \notin S \Leftrightarrow M(a) = \bot_o$), by the exact same reasoning as the $\Leftarrow$ direction above we prove the $\Rightarrow$ of this part. For the $\Leftarrow$ direction, by absurdum suppose that there is an argument $a \in S$ such that $M(a) = \bot_o$. By the $\Leftarrow$ direction above, if $a \in S$, then $M(a) = \top_o$, which is a contradiction as long as $\top_o \neq \bot_o$. □

This result relates the unique extension of a well-founded abstract argumentation framework with a family of semantic frameworks whose behaviour need only be specified for two values, $\top$ and $\bot$. Arguments assigned $\top$ by our semantics coincide with the ones belonging to the unique extensions. If nothing else, this result gives us a measure of security. We have not strayed far from the intuitions found in Dung's work.

There are surely more complex relations to be found, although further conditions for coincidence of semantics are bound to be more and more complicated. We conjecture, for example, that this result directly applies to frameworks without isolated cycles (i.e. cycles unattacked by outside arguments). If at least one argument of a cycle is attacked by an argument in the unique extension, then we can expect its successor in the cycle to be acceptable, and become part of the unique extension in a further application of F_{AF}.

This is a much more solid result that has not been investigated further due to time constraints. As a special case of the above result and as a curiosity we have the following curiosity.

Proposition 5.1. *$\mathcal{S}^{\times}$ satisfies the conditions of Theorem 5.1.*

Proof. Trivial □

Furthermore, even if there are isolated cycles of odd-length (which in certain cases have been proven to have unique models in $\mathcal{S}^{\times}$), then perhaps the notion of x-admissibility [20] would allow us to specify even tighter connections between Dung's work and the Product Semantics. X-admissibility is simply a criterion for the discretisation of degrees of truth into True/False, depending on whether the x threshold has been surpassed or not.

5.2 Generic Gradual Valuations

In this section we discuss some of the most important characteristics of our work. We will attempt to demonstrate whether our semantics coincides with Generic Gradual Valuations [13], and under what conditions it happens. There is no great difficulty, as our work is partially inspired by [13]. All we have to do is match our totally ordered sets with V and W, and our operators to the functions g and h.

Theorem 5.2. *Let AF and $F = \langle \mathcal{A}(AF), \mathcal{R}(AF), V \rangle$, such that $\forall_{a \in \mathcal{A}} V^+(a) = 1, V^+(b) = 1$. If v is a Generic Gradual Valuation of AF with g and h such that $h(l_1, ..., l_n) = h(l_1, h(l_2, ...h(l_{n-1}, l_n)))$, then there exists $\mathcal{S}$ and $M \in \mathcal{M}$ such that $\forall_{a \in \mathcal{A}} v(a) = M(a)$.*

Proof. Let us set $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$ such that

- $\forall_{a \in \mathcal{A}} \tau(a) = \frac{V^+(a)}{V^+(a)+V^-(a)} = 1 = \top_o = V_{max}$, $L_i = \{\top_o\}$,
- $L_o = V$,
- $L_a = W$,
- $\top_a = W_{max}$, $\bot_a = W_{min}$, $\top_o = V_{max}$, $\bot_o = V_{max}$,
- $\curlyvee = h$ (i.e. $l_1 \curlyvee l_2 = h(l_1, l_2)$), and thus $\bigcurlyvee_{x \in \{x_1, ..., x_n\}} x = x_1 \curlyvee ... \curlyvee x_n = x_1 \curlyvee (x_2 \curlyvee ...(x_{n-1} \curlyvee x_n)...) = h(x_1, h(x_2, ...h(x_{n-1}, x_n))) = h(x_1, ..., x_n)$,
- $\curlywedge\neg = g$. In other words, we set $l_1 \curlywedge \neg l_2 = g(l_2)$, by defining $l_1 \curlywedge l_2 = l_2$ and $\neg l_1 = g(l_1)$.
- $\neg\bot_o = \top_o$

Let $a \in \mathcal{A}$. There are only three possible cases:

- $\mathcal{R}^-(a) = \emptyset$. By the definition of v, g and h, $v(a) = V_{max}$. $M(a) = \tau(a) \curlywedge \neg \bigcurlyvee_{a_i \in \emptyset} M(a_i) = \top_i \curlywedge \neg\bot_o = \neg\bot_o = \top_o = V_{max} = v(a)$.
- $\mathcal{R}^-(a) = \{b\}$. By the definition of v, g and h, $v(a) = g(h(v(b))) = g(v(b))$. $M(a) = \tau(a) \curlywedge \neg \bigcurlyvee_{a_i \in \{b\}} M(a_i) = \top_o \curlywedge \neg M(b) = \neg M(b) = g(v(b)) = v(a)$.
- $\mathcal{R}^-(a) = \{a_1, ..., a_n\}$. By the definition of v, g and h, $v(a) = g(h(v(a_1), ..., v(a_n)))$.

$$\begin{aligned} M(a) &= \tau(a) \curlywedge \neg \bigcurlyvee_{a_i \in \mathcal{R}^-(a)} M(a_i) \\ &= \tau(a) \curlywedge \neg M(a_1) \curlyvee ... \curlyvee M(a_n) \\ &= \neg(M(a_1) \curlyvee (M(a_2) \curlyvee ...(M(a_{n-1}) \curlyvee M(a_n))...)) \\ &= \neg(v(a_1) \curlyvee (v(a_2) \curlyvee ...(v(a_{n-1}) \curlyvee v(a_n))...)) \\ &= \neg(h(v(a_1), h(v(a_2), ...h(v(a_{n-1}), v(a_n))...))) \\ &= \neg h(v(a_1), ..., v(a_n)) \\ &= g(h(v(a_1), ..., v(a_n))) \\ &= v(a) \end{aligned}$$

We have thus been able to construct a F and $\mathcal{S}$ for which the two semantics coincide. □

With this demonstration, we have shown that generic gradual valuations can be embedded within our system under extremely mild conditions. We may now state that our semantics is at least as expressive as generic gradual valuations. Whether we can embed our semantics within theirs is another matter altogether, which has already been studied when we attempted to reduce our frameworks to frameworks containing only perfect arguments. Before moving on, all embedding relations to generic gradual valuations are also transitively applicable to our framework.

As such, and without much detail,

Corollary 5.1. *The Social Gradual Valuation is capable of capturing the Rooted Labellings of Jakobovits and Vermeir [19].*

Proof. See [13], and by Theorem 5.2. □

Corollary 5.2. *The Social Gradual Valuation is capable of capturing the gradual valuations of Besnard and Hunter [11].*

Proof. See [13], and by Theorem 5.2. □

Further interesting results can be found in [13]. The two corollaries above were chosen to demonstrate the representative capabilities of the system.

One of the issues still open is whether [13] allows an infinite number of arguments at the onset. When talking about the global valuation procedures, the removal of cycles implies the addition of an infinite number of arguments to the framework. However, nothing is said about the initial state of the framework, and hence our inability to compare. Furthermore, global valuations are radically different from generic gradual valuations, so we do not believe its usage of infinite number of arguments implies their existence in generic gradual frameworks.

Two-way embeddings implying equivalence are usually some of the more expressive and important results in certain studies. At the very least, it is important to prove embeddings under some set of restrictions. With Theorems 4.4 and 5.2 we have achieved a good understanding of the relative expressive capabilities of these two proposals.

By restricting our semantics quite heavily, thereby ruling out the two concrete semantics that we have presented and discussed, we are capable of reducing social argumentation frameworks to equivalent but possibly infinite abstract argumentation frameworks. This is the object of work in [13]. On the other hand, under very mild assumptions, we can restrict [13] in a way that we can represent their semantics. However, we argue that this restriction is merely superficial. Our entire semantics could easily be rewritten in the form of functions instead of operators without any implications in the rest of the work. We simply chose this notation to help the reader use his intuitions with fuzzy logics and fuzzy logical operators.

We thus believe our approach at the generic level to be strictly more expressive than those currently found in the literature.

6 . Conclusions

In this chapter we intend to draw attention to the motivation behind the project that serves as the context for this dissertation. We will quickly review the problem we intend to solve and how our framework semantics and address that problem. A recap of the properties of the system as well as relations to other formalisms follows, along with a list of contributions.

From the multitude of online debating systems and forums where users engage in discussions, it is clear that a desire to engage in debates is pervasive around the Internet. Some specific websites such as http://www.debate.org attempt to address this issue by providing a debating platform where the user base itself decides the winner of a debate. The driving force of these systems is, quite simply, its user base. It single-handedly decides the winners.

This dissertation is inserted within the context of a project that aims at creating an online debating system that harnesses the power of a formalisation of such debates and addresses a series of perceived shortcomings in existing systems. We will not repeat them here, as they have already been introduced elsewhere.

The core concept of this online debating system is that two levels of expertise are accommodated. On one side, experts engage in debates regarding several topics, creating a reusable formal knowledge base structured by the arguments proposed. On the other side, less expert users are given the opportunity to vote on arguments in a pro/con way. This participation does not generate content, but opinion: it gives us an insight into how the crowd perceives each argument.

Assuming that a system can transform online information into this knowledge base and opinion, the proposed objective is to define a semantics capable of providing the subjective but formally justifiable outcome of a debate. This is what we set out to do. In Chapter 3 we introduced the formal representation of an online debate with votes. Deeper into the chapter, we defined a very generic model-based semantics, and finally two fully specialised semantics. The contributions of that particular chapter were:

- An overview of the decisions that led us to the current proposal.
- The formalisation of the structure of a debate as a Social Argumentation Framework $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$.
- An extremely generic model-based semantics dependent on a semantic framework $\mathcal{S} = \langle L_i, L_a, L_o, \tau, \neg, \curlywedge, \curlyvee \rangle$.
- Two concrete semantics, with the intent of serving as the backbone of the online debating system.

Chapter 4 is intended to provide the means of understanding the behaviour of the generic and concrete proposals of Chapter 3. It is essential in that there are many intuitively desirable properties which the system must satisfy. We studied the behaviour of the generic semantics in frameworks with certain restrictions. The contributions of this initial part of the chapter were:

- The study of the impact of the existence of cycles on the uniqueness of models for a framework.

- A description of the behaviour of a the generic semantics in an uncontroverial framework, given its evolving state.

- A proof that any social argumentation framework can be represented by possibly infinite abstract argumentation frameworks under certain restrictive semantics.

This overview is important insofar as it does not show any unacceptable erratic behaviour. In fact, we are able locate and isolate the source of possible undesirable situations. There is a preliminary analysis of the behaviour of the generic semantics in uncontroversiality, and finally a result about the expressiveness of our frameworks and semantics relative to other proposals. It should have given the reader an overview of how the system is expected to behave in the general case.

The second part of Chapter 4 deals with concrete semantics. These are to be used as implemented solutions of the envisioned system, so a deeper understanding of them is necessary. The contributions of this part of the dissertation were:

- A description of cycle-centric conditions for the non-uniqueness of models for $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$.

- The validation of the principles underlying our operators by comparison with commonly accepted properties found in the literature [13, 25].

- The identification of a core divergence in the principles behind $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$ regarding the notion of attack, and motivation behind the preference of $\mathcal{S}^{\times}$ over $\mathcal{S}^{-}$.

These contributions are used as further justification that our work does not stray from Argumentation Theory's most important postulates. There is one particular situation where there is a divergence which is perfectly justifiable given our context.

Chapter 5 aims at establishing relations between our semantics and the literature. For this, we take Dung's [17] classic work as well as [13], which is the closest to our current proposal that we know of. The contributions of that chapter were:

- The proof of an embedding of Dung's unique extensions of well-founded abstract argumentation frameworks into a special case of our semantics and social argumentation framework.

- The proof of an embedding of generic gradual valuations to our generic semantics under very relaxed conditions.

- The embedding of proposals by [19] and [11] by transitivity through [13].

Despite the great differences in their nature, we are still able to prove a simple relation to Dung's classic work for the uncontroversial case of well-founded frameworks. We conjecture that further relations can be shown, pending further investigation. The main discovery of the chapter is undoubtedly the relation with generic gradual valuations, providing our semantics with a wealth of further results.

This chapter follows, and the next one is a review of a few open questions that are still of considerable importance for this work.

In summary, we have defined the initial formalisation of a system we believe to be capable of providing an answer to our envisioned online debating system, effectively paving the way for a formal system capable of subjective reasoning based on popular opinion. Throughout this dissertation, we have proven properties that enable us to further understand the behaviour the generic semantics for such a system in many different cases.

Given their possible role as the online system's choice semantics, an effort was also made to understand the Product and Subtraction semantics more specifically, as well as the differences between them. We do this in order to better justify a decision between the two. For this, a wealth of properties taken from the literature was used. We argue that at this stage the Product Semantics is more desirable for two main reasons. First, only perfect arguments can defeat other arguments, virtually ensuring that every vote has some amount of weight in the outcome, and thus that every vote is represented. Secondly, our study indicates that it will have unique models more often than the alternative. It is thus our choice for the implementation of such an online system, if it were to happen.

Finally, we analyse the expressiveness of our proposal by comparing it to others such as generic gradual valuations [13]. We argue why the addition of votes makes for more expressive semantics, and conclude that their proposal is essentially embeddable within ours.

With more research invested in the topics listed under Further Work (Chapter 7), we hope make available a solid solution for the envisioned online debating system and more generally for subjective formal reasoning.

7 . Future Work

This chapter intends to describe future research that naturally follows from this dissertation. The good thing about this work is that it raises more questions than it answers. Aside from furthering the results on certain topics, there are whole new areas do explore. We will present them by order of importance.

First and foremost, there are two conjectures (4.2 and 4.3 that need to be proven in order to guarantee a unique model. Above all else, these conjectures hold the key to the applicability of the Product Semantics to the envisioned online system. We will need to analyse the non-linear equation systems obtained from frameworks, and understand whether they always have a solution or whether they have multiple solutions. It may happen that the conjectures are only true for certain types of frameworks, in which case we hope that the analysis of the non-linear equation systems will provide us with some insights as to the structure of the frameworks with a unique model.

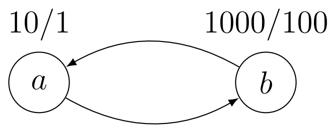


Figure 7.1: Skewed mutually attacking arguments.

The inclusion of the notion of argument relevance is an extremely interesting topic for several reasons. Consider Figure 7.1. Any of our semantic frameworks will assign $\tau(a) = \tau(b) = \frac{10}{11}$. Based on this data, it will conclude that both arguments have equal weight, and will assign a model based on that conclusion.

Our intuition warns us that this does not "feel" right. Two arguments, b with a hundred times more votes than a have the same weight as long as their ratio is the same. It is quite debatable that we should allow this. One may still counter with the idea that as an online system where many users participate, these discrepancies between the number of votes tend to level with time. This is not quite satisfactory, as it is undeniable that some arguments are simply more interesting than others to the crowd, and will call for more votes. The same can be said for entire debates, which happen independently, until an attack between the two is identified and they become connected. The two debates may have generated a completely different interest from the crowd, and their merge will undoubtedly lead to the situation described above.

The inclusion of relevance may serve another purpose. We have seen that in occasions with cycles and where arguments have equal social supports, an infinity of models tends to appear. The inclusion of relevance would make the conditions of existence of multiple models for two arguments having equal inherent values (both crowd support and relevance) that much more remote.

It also raises more interesting questions which are not of a technical nature. The attack relation may be chosen to affect acceptability as well as relevance. Does an argument which generated very little interest become more relevant if it is attacked (or attacks) a very relevant argument? There are good arguments in favour and against this.

Another very interesting avenue of research is allowing votes on attacks. Because the debates will happen in natural language, any modelling of the knowledge represented therein will inherently have certain omissions, flaws, or inconsistencies. With arguments being abstract, the absence of a clearly defined logic precludes the existence of premises and derivable conclusion within an argument. Without this clear statement, it is possible that certain attacks between arguments themselves become a matter of debate. For example, recall the initial example of voting by the Portuguese crowd. Perhaps the wind is more of a problem when you're in the beach than when you're in a park. As of yet, we cannot represent this difference in the weight of each attack.

One way of dealing with this issue is by allowing arguments to attack attacks, as in [22]. More in tune with the envisioned online system is simply allowing attacks to be voted on as well. The validity of the attack according to the crowd is then taken into account, for example by pondering the attack based on some metric defined similarly to τ. As described in this dissertation, our semantics assumes that all attacks are completely valid and make perfect sense, thus being "perfect attacks". We expect this extension of weights in attacks to be quite easy - we can simply ponder each attacker strength by the relevant attack weight.

Within the context of the Reduction Theorem (4.4), tt would be of some importance to clarify under what conditions one could obtain a finite representative framework. It is already clear that if any argument is inherently flawed, we need an infinite number of arguments but the remaining situations remain uncharacterised. This would strengthen the relation between our model-based semantics and [13].

During this dissertation, we identified that the main cause of there being multiple models was the existence of cycles. Instead of finding very particular conditions where cycles do not exist (or are extremely limited structurally) with the purpose of avoiding multiple models, a radically different approach might be suitable. We could accept that our semantics provides us with a multitude of models. One could then characterise and study that set of models, by looking at the structure of the cycles and crowd support of each argument, and arrange for a way to turn that set into a single, more desirable model.

In Figure 4.4, for example, it was not hard to identify the source of the problem as being $(1, 1)$. We went even further, and discovered that there is actually a value for the model of a as the social supports of a and b tend towards 1. This limit was, in this particular instance, a perfect way of reducing the number of models from infinite to just one. This type of analysis might be generalisable to more situations, and should be looked into with much more attention. We can attempt to use this method for more complex cycles where crowd support may take arbitrary values.

Related to the previous point, the study of cycle behaviour for $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$ is still preliminary. In particular, we have shown sufficient conditions for the existence of multiple models in a framework. A description of the structure of some non well-founded frameworks that can be proven to have a unique model would be highly desirable. It would provide us with stronger guarantees that the envisioned online system would present the users with a single outcome.

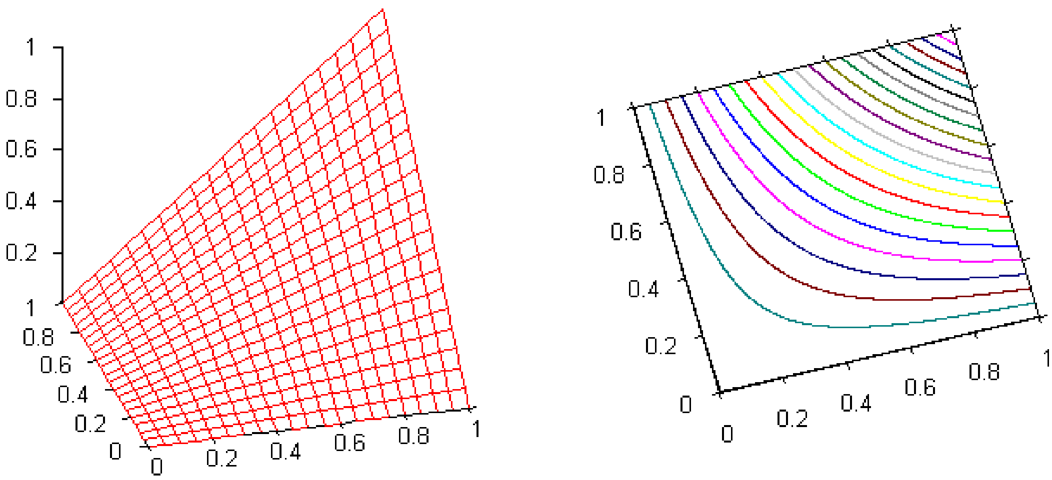


Figure 7.2: Graph of the Product T-norm, taken from Wikipedia.org

Besides $\mathcal{S}^{-}$ and $\mathcal{S}^{\times}$, there are alternative semantics that deserve further study. In particular, a T-Norm based semantic framework based on the Hamacher T-norm would warrant further study. The Product T-norm and Hamacher T-norm have very similar graphs, and the impact of such minimal differences for example in the existence and uniqueness of models deserves more attention. Judging from the graphs of the two T-norms in Figures 7.2 and 7.3, the Hamacher T-norm is much better suited for situations where arguments have low crowd support. It differentiates small changes near 0 much better than the Product T-norm which appears to normalise values in that area of its domain.

It is also possible that some of the other non-continuous T-norms relate particularly well with extension-based semantics, and so the Drastic and Łukasiewicz T-norms should not be simply discarded.

Finally, an extremely interesting research avenue could be in applying this work to negotiation in multi-agent systems with inconsistent knowledge bases. Imagine several agents need to

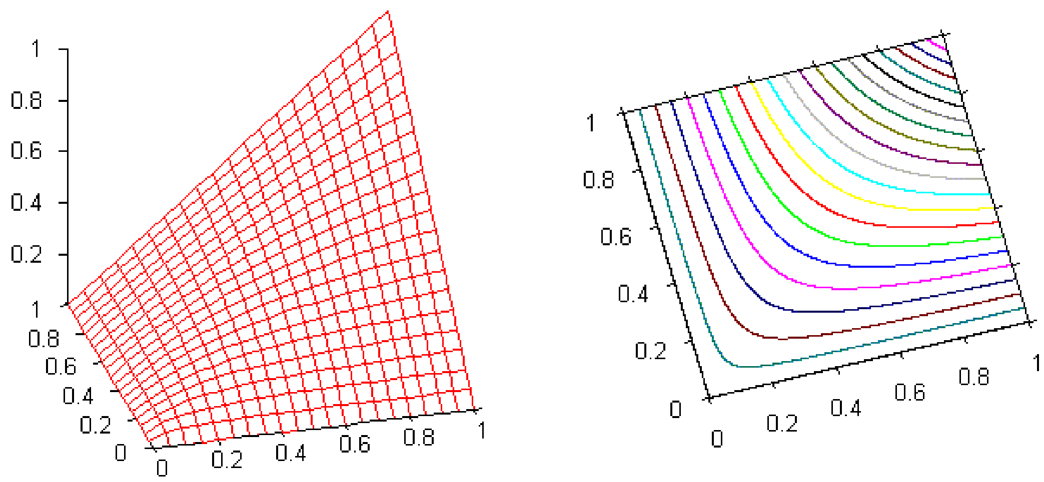


Figure 7.3: Graph of the Hamacher T-norm, taken from Wikipedia.org

make a decision, and each has a logical set of beliefs. Then, we can use formal proofs to construct structured arguments, which are not ruled out by our decision not to use them explicitly them. An agent can propose a structured argument for his preferred decision to a certain topic. Typically, agents would simply use rebutting attacks for this decision. They might use undercutting attacks, but the inconsistent knowledge bases would completely confuse the ensuing debate and make it impossible to make any sensible decision.

By applying our work to this scenario, agents might be able to vote on arguments which are consistent or inconsistent with their personal knowledge and beliefs. Therefore, even though there are inconsistent knowledge bases, these semantics could provide a definite formal outcome for the negotiation.

We hope to have motivated the reader to all the work that lay ahead, and why that work is relevant both within and without the context of this dissertation's project.

Bibliography

[1] Jürgen Dix Alejandro J. García and Guillermo R. Simari. Argument-based logic programming. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 153–172. Springer, 2009.

[2] Leila Amgoud. Argumentation for decision making. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 301–320. Springer, 2009.

[3] Leila Amgoud, Jean-François Bonnefon, and Henri Prade. An argumentation-based approach to multiple criteria decision. In Lluis Godo, editor, *ECSQARU*, volume 3571 of *Lecture Notes in Computer Science*, pages 269–280. Springer, 2005.

[4] Leila Amgoud, Claudette Cayrol, and Marie-Christine Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In James P. Delgrande and Torsten Schaub, editors, *NMR*, pages 1–9, 2004.

[5] Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58, 1950.

[6] Katie Atkinson and Trevor Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artif. Intell.*, 171(10-15):855–874, 2007.

[7] Pietro Baroni and Massimiliano Giacomin. Semantics of abstract argumentation systems. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 24–44. Springer, 2009.

[8] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, June 2003.

[9] Trevor Bench-Capon and Katie Atkinson. Abstract argumentation and values. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 45–64. Springer, 2009.

[10] Trevor Bench-Capon, Katie Atkinson, and Alison Chorley. Persuasion and value in legal argument. *J. Log. and Comput.*, 15(6):1075–1097, 2005.

[11] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artif. Intell.*, 128(1-2):203–235, 2001.

[12] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In Lluis Godo, editor, *ECSQARU*, volume 3571 of *Lecture Notes in Computer Science*, pages 366–377. Springer, 2005.

[13] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Graduality in argumentation. *J. Artif. Int. Res.*, 23(1):245–297, 2005.

[14] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Coalitions of arguments in bipolar argumentation frameworks. In *CMNA*, volume 3571, pages 14–20, 2007.

[15] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolar abstract argumentation systems. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 65–84. Springer, 2009.

[16] Carlos Viegas Damásio and Luís Moniz Pereira. Antitonic logic programs. In *LPNMR '01: Proceedings of the 6th International Conference on Logic Programming and Nonmonotonic Reasoning*, pages 379–392, London, UK, 2001. Springer-Verlag.

[17] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[18] Thomas F. Gordon and Douglas Walton. Proof burdens and standards. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 239–259. Springer, 2009.

[19] Hadassa Jakobovits and Dirk Vermeir. Robust semantics for argumentation frameworks. *J. Log. Comput.*, 9(2):215–261, 1999.

[20] Jeroen Janssen, Martine De Cock, and Dirk Vermeir. Fuzzy argumentation frameworks, 2008.

[21] Nikos I. Karacapilidis and Dimitris Papadias. Computer supported argumentation and collaborative decision making: the hermes system. *Inf. Syst.*, 26(4):259–277, 2001.

[22] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, 173(9-10):901–934, 2009.

[23] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[24] John L. Pollock. How to reason defeasibly. *Artif. Intell.*, 57(1):1–42, 1992.

[25] John L. Pollock. Defeasible reasoning with variable degrees of justification. *Artif. Intell.*, 133(1-2):233–282, 2001.

[26] Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7(1), 1997.

[27] Henry Prakken and Giovanni Sartor. Reasoning with precedents in a dialogue game. In *ICAIL*, pages 1–9, 1997.

[28] Iyad Rahwan and Kate Larson. Argumentation and game theory. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 321–340. Springer, 2009.

[29] Ralf Schweimeier and Michael Schroeder. A parameterised hierarchy of argumentation semantics for extended logic programming and its application to the well-founded semantics. *Theory Pract. Log. Program.*, 5(1-2):207–242, 2005.

[30] Frieder Stolzenburg, Alejandro Javier García, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Computing generalized specificity. *Journal of Applied Non-Classical Logics*, 13(1):87–, 2003.

[31] Bart Verheij. Automated argument assistance for lawyers. In *ICAIL*, pages 43–52, 1999.

[32] Bart Verheij. Deflog: on the logical interpretation of prima facie justified assumptions. *J. Log. Comput.*, 13(3):319–346, 2003.

[33] Bart Verheij. The toulmin argument model in artificial intelligence. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 219–238. Springer, 2009.

[34] D. N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1996.

[35] Douglas Walton. Argumentation theory: A very short introduction. In Guillermo R. Simari Iyad Rahwan, editor, *Argumentation in Artificial Intelligence*, pages 1–22. Springer, 2009.