Debate in a multi-agent system : multiparty argumentation protocols

Dionysios Kontarinis

Debate in a multi-agent system: multiparty argumentation protocols

PhD Thesis of **Dionysios KONTARINIS**
defended on 21 November 2014

Laboratoire d'Informatique Paris Descartes (LIPADE)
Ecole Doctorale EDITE de Paris

Jury composition:

Prof. Pavlos MORAITIS	Université Paris Descartes	(Director)
Dr. Elise BONZON	Université Paris Descartes	(Co-supervisor)
Prof. Nicolas MAUDET	Université Pierre et Marie Curie	(Co-supervisor)
Prof. Sébastien KONIECZNY	CNRS, CRIL	(Reviewer)
Prof. Antonis KAKAS	University of Cyprus	(Reviewer)
Dr. Sylvie DOUTRE	Université Toulouse 1	(Examiner)
Dr. Gabriella PIGOZZI	Université Paris Dauphine	(Examiner)

Acknowledgements

I would like to thank the following people for helping me, in one way or another, during my thesis. My supervisors, Elise, Nicolas and Pavlos, for offering me this great opportunity, for their guidance, and for always being available to discuss my ideas and questions. My parents, brother and grandmothers, for their unconditional love and support. Katerina for her understanding and help. My friends Manolis, Petros, Dionysis, Giannis, Anestis, Mehdi, Nabila, Julien, Charles, Italo and the guys at the Foyer Hellénique, for always keeping my morale high. Also, Leendert, Serena and Alan, with whom it was a pleasure to collaborate. Finally, the people who helped me during my stay in Paris: Sœur Winnie, Marie-Claire Vallaud, Xenia Chryssochoou, Jos Aelvoet, Maria Gravari-Barbas and Father Panagiotis Xenitellis.

Contents

1 Introduction **7**

I Computational Argumentation **15**

2 Basics of Argumentation **17**
2.1 Abstract Argumentation . . . 18
2.1.1 Abstract systems with attack relations (only) . . . 18
2.1.2 Abstract systems with preferences and weights . . . 25
2.1.3 Bipolar Argumentation Frameworks . . . 31
2.1.4 Other types of relations . . . 33
2.2 Merging Multiple Viewpoints . . . 33
2.2.1 Focusing on the notion of attack . . . 33
2.2.2 How is disagreement over attacks possible? . . . 34
2.2.3 Merging argumentation systems . . . 34

3 Expertise in Argumentative Debates **37**
3.1 Modelling the Debate's Participants . . . 37
3.1.1 Defining Expertise . . . 38
3.2 Modeling an Argumentative Debate among Experts . . . 39
3.2.1 The Gameboard . . . 40
3.3 Obtaining the Debate's Conclusions . . . 42

II Abstract Argumentation Dynamics **45**

4 Background on Abstract Argumentation Dynamics **47**
4.1 Proof Theories as Argument Games . . . 49
4.2 Dynamics of Labels . . . 50
4.3 Dynamics and Change . . . 52
4.3.1 Effects of Change on Abstract Argumentation Systems . . . 52
4.3.2 The enforcing problem . . . 55

5 Contribution to Dynamics of Abstract Argumentation **61**
5.1 Abstract Argumentation System with Modifiable Attacks (ASMA) . . . 62
5.2 Goals and (Minimal) Successful Change . . . 63
5.2.1 Relations among sets of Successful Moves and Target Sets . . . 65
5.2.2 Some properties on the content of target sets . . . 69
5.2.3 Modifiable arguments: do they increase expressiveness? . . . 71
5.3 A Rewriting Procedure for Target Set Computation . . . 72
5.3.1 The Maude system and the intuition behind our program . . . 72
5.3.2 Program structure - Rewriting Rules . . . 73
5.3.3 The Rewriting Procedure (RP) . . . 74

5.4 Target Set Evolution . . . 82
5.4.1 Playing outside target sets . . . 83
5.4.2 Playing in target sets . . . 86
5.5 Controversy of Argumentation Systems . . . 87
5.5.1 Debate Phase: The experts express their opinions . . . 88
5.5.2 Evaluation phase: measuring the controversy . . . 89
5.5.3 Phase 3: Asking the opinion of an additional expert . . . 91

III Protocols for Argumentative Dialogues 101

6 Background on Argumentative Dialogues 103
6.1 Basic Elements of Argumentative Dialogues . . . 105
6.1.1 Persuasion dialogues . . . 105
6.1.2 Dialogue system for argumentation . . . 106
6.1.3 Agent Strategies . . . 110
6.2 Multilateral Argumentative Dialogues . . . 111
6.3 A Typology of Protocols for Multilateral Dialogue Games . . . 114
6.3.1 Agent Groups . . . 115
6.3.2 Rules for speaker order (turn-taking) . . . 116
6.3.3 Locutions (permitted utterances) . . . 117
6.3.4 Agent Strategies . . . 118
6.4 Evaluating Argumentative Dialogues . . . 120
6.4.1 Evaluating persuasion dialogues . . . 121
6.4.2 Evaluating strategies in persuasion dialogues . . . 122

7 Contribution to Argumentative Debate Protocols 125
7.1 Converging quickly to a persistent issue . . . 127
7.2 A Debate Protocol Focusing on Target Sets . . . 129
7.2.1 Modeling the debate's setting . . . 129
7.2.2 Defining the protocol . . . 132
7.2.3 Strategies using target sets . . . 133
7.2.4 The experimental setting . . . 138
7.2.5 Analyzing the results . . . 140
7.2.6 Conclusion on strategies and heuristics . . . 143
7.3 Debate Protocols Using Bipolar Argumentation Frameworks . . . 143
7.3.1 Brief reminder of bipolar argumentation . . . 143
7.3.2 Merged Bipolar Argumentation Framework . . . 145
7.3.3 Designing protocols - focus on the agents' goals . . . 146
7.3.4 Illustrative examples and properties of π_0, π_1 . . . 152
7.3.5 Conclusion on bipolar protocols . . . 161
7.4 Protocol Evaluation . . . 161

8 Conclusion 165
8.1 Directions of Future Research . . . 167
8.1.1 Amendments and extensions . . . 167
8.1.2 Contribution to some broader research directions . . . 169

A Table of Basic Notations 181

B Maude's listing 183

Chapter 1

Introduction

Interest in the formal study of multiparty (also called multilateral) dialogues [DV04] has recently increased. Two reasons contributing to this fact are the following. Firstly, advances in the domain of Multi-Agent Systems [Woo09] have lead to more research on efficient methods for agent communication. Secondly, dialogues among users are omnipresent in social media (e.g. blogs and microblogs such as *Twitter* and *Tumblr*; social networking sites such as *Facebook*[1]; content communities such as *YouTube* and *DailyMotion*), but also in various web-sites, such as news and sports sites.

Figure 1.1: Part of a dialogue in Facebook, taken from the page of the non-governmental organisation *Debating Europe*.

The dialogue in Figure 1.1 is featured on the Facebook page of the non-governmental organisation *Debating Europe*.[2] The form of the dialogue is quite simple, it essentially consists of a list of comments. Participation of many users is supported, and each user can post his comment(s).

[1] http://www.facebook.com
[2] http://www.debatingeurope.eu

The comments are illustrated in chronological order and users have the possibility to *like* previous comments.[3]

In media, such as Facebook, users can discuss freely on any topic they want. But some dialogues are very different than others. Imagine for example a dialogue among friends which has no specific motivation, other than recreation, and another dialogue where the participants debate on which political party has the best program. The authors in [WK95] have proposed a typology of dialogue types, based on the type of the participants' goals, as well as on the means they use in order to achieve them. Our work focuses on a specific type of dialogues, called *persuasion dialogues*. In these dialogues the participants have contradicting viewpoints on some issue and each participant tries to convince the others to adopt his viewpoint.

In order to persuade others, participants may exchange *arguments* [POT69]. An argument is roughly a claim and supporting information for that claim. Argumentative dialogues in social media such as Facebook face a number of challenges. We begin by identifying two basic ones.

1. **Difficult identification of arguments (and their relations):**
   Users construct arguments in natural language, instead of using a stricter formalism. Often, a comment intended to be an argument, also contains irrelevant information. Moreover, in some cases multiple arguments may be sent in a single comment, not clearly separated. Therefore, identifying arguments and their relations, in such dialogues is not an easy task. Works such as those in [SCG13, CV12] address this issue, by using techniques able to identify arguments and their relations from text in natural language.

2. **Unstructured dialogue:**
   Illustrating the users' comments the one after the other, in chronological order, may be sufficient for simple interactions, but when they become longer, an important issue is raised: it becomes increasingly difficult to understand the relations between the comments. For example, it may be difficult to understand if a given argument is an intended reply to a previous argument, and to which one.

An important step towards improving the quality of argumentative dialogues among human users is mapping arguments [Dav11] to graphical representations. In some cases, we may have graphical representations for other concepts, such as issues, ideas and statements. For example, Issue-Based Information Systems (IBIS) [KR70] were defined to support coordination and planning of political decision processes, by guiding the identification, structuring, and settling of issues.

There exist several tools for argument mapping. Some examples are: CompendiumNG [4], Araucaria[5] [RR04], Rationale [6],Argunet.org [7]. Tools for argument mapping can be valuable to anyone who wants to study the relations among arguments related to a given issue. But they are also an essential feature of multi-party debates. In these debates, the participants need a way to represent all the arguments which have been put forward, as well as their relations.

Multi-party debates over the Internet may take place on *debate platforms*. Such platforms provide support to their users, who can debate with others about various topics. A debate in these platforms is usually represented in the form of a graph, collectively constructed by the debating users. Typically, the users are allowed to introduce arguments on the graph (but also other entities such as issues and positions), as well as relations among them. Sometimes, they can also express their opinions by casting votes on the graph's elements. In some simpler cases, the debate is just a collection of arguments, partitioned into those supporting and into those opposing a given position. Some debate platforms are the following: Debategraph[8], Quaestio-it.com[9] [ET14],

[3] We note that debates on the official site of *Debating Europe* have a very similar setting.
[4] http://www.compendiumng.org
[5] http://araucaria.computing.dundee.ac.uk
[6] http://rationale.austhink.com
[7] http://www.argunet.org
[8] http://debategraph.org
[9] http://www.quaestio-it.com

TruthMapping.com[10], LivingVote.org[11], Debate.org[12], and DebatePedia[13]. In Figure 1.2 part of a debate on Debategraph is illustrated.

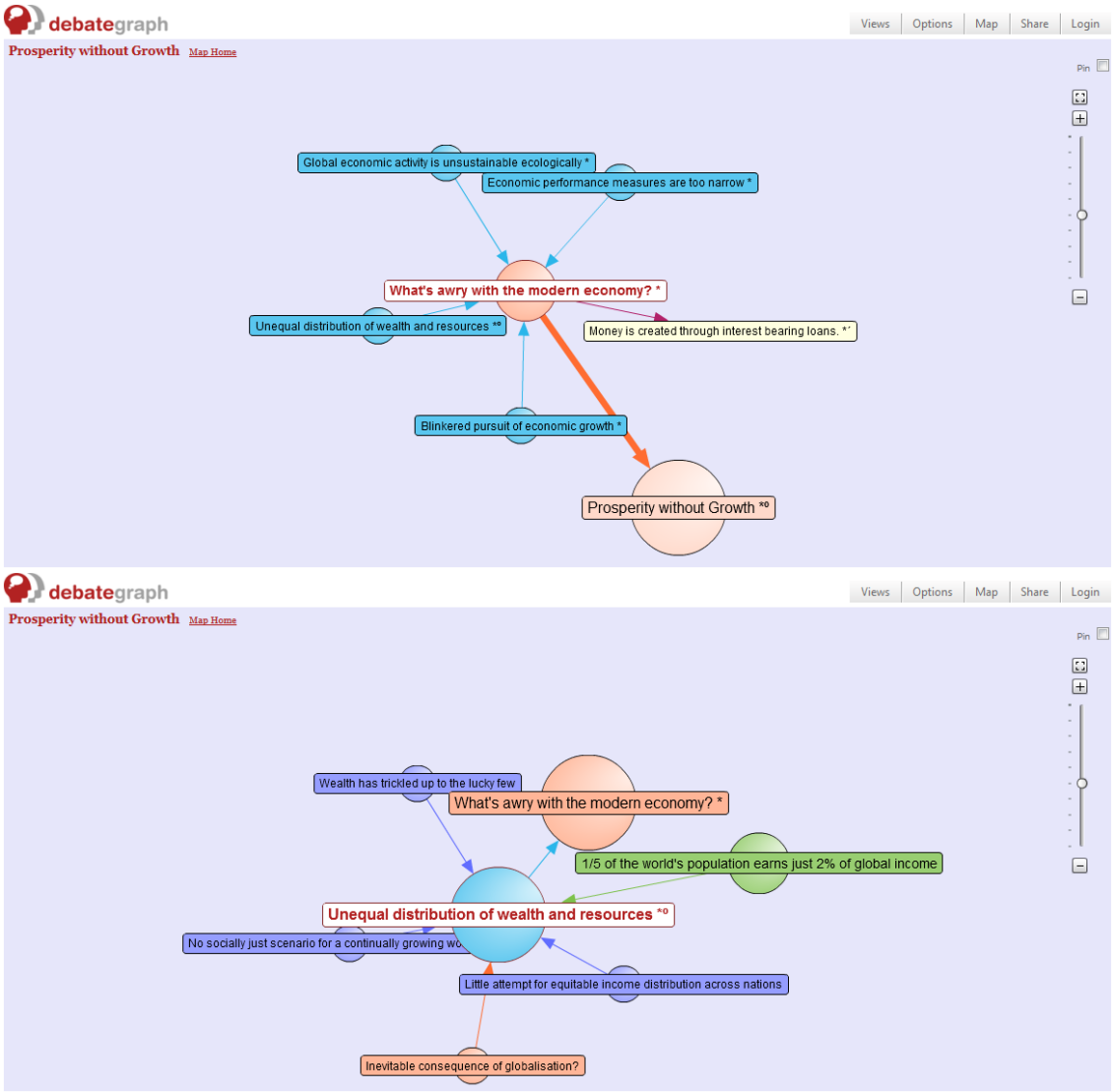


Figure 1.2: Two snapshots from the same debate in Debategraph. Different colours correspond to different types of entities (e.g. arguments, issues, positions). Arrows represent different types of relations among entities. An arrow's thickness reflects its support from the users.
Top: The focus is on the issue "What is awry with modern economy?". A number of positions (blue circles) *respond* to this issue.
Bottom: After clicking on the position "Unequal distribution of wealth and resources" (shown in the top image), the focus shifts to it. Dark blue circles are the position's components. Green circles are supporting arguments.

Debate platforms which use argument mapping address the basic issues of multilateral argumentative dialogues:

[10] http://truthmapping.com
[11] http://www.livingvote.org
[12] http://www.debate.org/
[13] http://idebate.org

- The content of the arguments is explicitly stated. It may be a text in natural language, a link towards a web-page, or even an image. The same holds for other types of entities that some systems allow, such as positions and issues. This can be seen in Figure 1.2 which illustrates a part of a debate in Debategraph.

- Argument mapping leads to the construction of graphs. These graphs allow the users to easily identify arguments, and their relations, even when their numbers increase significantly.

- Users are able to state their opinions on arguments and relations. For example in Debategraph, a user can assign a number to each argument and relation, reflecting his opinion on it. Notice for example the thick orange arrow in Figure 1.2, which has received positive feedback from the users.

In [SLPM10] the authors provide a review of the state of the art of computer-supported argumentation.[14] While open platforms have addressed some critical issues of multilateral debates among human users, there is still a number of additional issues, among which:

1. **Reaching conclusions in debates:**
   An inconvenience of debates in most open platforms is that, once the debate finishes, the conclusions are not always clear. This is due to the fact that most open platforms focus on the debate's representation, but they leave to the users the task of drawing the conclusions. In many cases this can be problematic. For example, users may be prejudiced and "read" the debate as they desire. Or perhaps the debate may be really complicated and the users may be unable to draw any conclusions. We argue that a mechanism for drawing conclusions would be a valuable addition to such platforms. Such a mechanism could provide additional information to the users, for example it could highlight the most important points of the debate, or it could evaluate whether the debate's conclusions can be questioned.

2. **Unfocused debates:**
   Usually, a debate is centered around a particular argument (or position), which we shall call *issue* of the debate. A problem arises from the fact that most debate platforms do not put restrictions on the possible utterances of the users. This may result in users jumping from one part of the debate to another, regardless of whether their utterances are related to the issue, and to what extent. This is problematic, as the debate may lose its focus on the issue, as explained in [Pra05]. The problem may be aggravated when there are pressing time constraints, and the debate's conclusion must be quickly drawn, but also when there are no time constraints, and the users easily shift their focus on other subjects.

3. **Lack of user coordination:**
   In debates where the number of users is arbitrary, and perhaps they can dynamically come and leave, coordination is a major issue. By coordination here we do not refer to how the users communicate and decide on their actions, but we refer to how the debate's central authority specifies what they are allowed to do, and when. For example, in what order should the users be asked to contribute? Can multiple users send utterances at the same time? Or perhaps a fixed order is preferable? If a user is asked for his contribution and he refuses to provide one, should he be asked again later? These questions have no easy answers, so a simple solution would be to consider almost no constraints, and to let the users contribute whenever they want. This solution is usually adopted in open platforms. But if there is pressure to obtain a conclusion (e.g. in presence of time constraints, or when the debate will lead to an important decision), this solution may be unsatisfactory. In these cases, the debate's central authority may prefer to coordinate the users more strictly.

4. **No consideration of user profiles:**
   Different users may have expertise in different topics, and a system which aggregates the

[14] They review literature not only on collaborative, but also on individual argumentation systems.

users' opinions must take this into account, in order to reach rational conclusions. Additionally, knowing the users' reputations [KW82] may be useful, as even experts should not always be trusted. Aggregating the users' opinions is an important issue in multilateral debates, and it is not sufficiently addressed in most existing platforms.

5. **Ambiguity of the meaning of votes:**
   Positive and negative votes on arguments and attacks have no predefined meaning, in most existing platforms. For example, a negative vote by a user on an argument could have two different meanings: (1) that he believes the argument is badly formed, or (2) that he believes the argument is well-formed, but not very convincing. This may lead to confusion, and to problems regarding the aggregation of votes.

We consider that the quality of debates using argument mapping can be improved. Let us first focus on the problem of drawing conclusions. Computational argumentation studies how, given some arguments and their relations, meaningful conclusions about the arguments can be drawn. For example, how can *accepted* and *rejected* arguments be computed?

One type of argumentation is *abstract argumentation* [Dun95]. In this type of argumentation we abstract from the structure and content of arguments, which are simply seen as abstract entities. Over a set of arguments A, an attack relation R is defined, indicating conflicts between arguments. The arguments and the attack relations can be represented in the form of a graph, which is called *argumentation graph*. Arguments are nodes on the graph, and attacks are arcs between nodes. Even in such a seemingly simple framework, the computation of accepted and rejected arguments is not straightforward. For example, if two arguments mutually attack each other, then which one should be accepted? As we will see in Chapter 2, different acceptability semantics for abstract argumentation have been defined, which correspond to different ways conclusions can be obtained from an argumentation graph. Abstract argumentation has been enriched in different ways, in order to take into account more information on the arguments and their relations. For example, support relations over arguments have been introduced, giving rise to Bipolar Argumentation Frameworks [ACLSL08]. From the study of debates in platforms we can verify the common use of support relations in everyday argumentation (check the only argument in Figure 1.2 which is a supporting argument). Also, preference relations over arguments can be used, as in [ADM08]. This can help us decide which argument to accept when, for example, two arguments are mutually attacked. Sometimes, preference relations are represented in the form of numerical weights, giving rise to weighted argumentation systems [DHM$^+$11]. These systems can be particularly useful in multi-party debates, because the weights attached to arguments or attacks could represent the support they receive from the users.

For the different types of frameworks, there exist different ways to compute acceptable arguments. For example, we can compute sets of arguments which defend themselves against their attackers (by counterattacking them) or we can attribute numerical valuations to arguments, indicating how much each argument is accepted or rejected. Computational argumentation has provided already many possible answers to the question of how to obtain meaningful conclusions, given a set of arguments and their relations. Our work will use computational argumentation, in the setting of multi-party debates.

From the above we see that computational argumentation can provide answers to the first issue mentioned before: the unclear conclusions of debates. But, what about the other issues? For example how can unfocused and uncoordinated debates be treated? In order to answer these questions, we may use elements of works which study argumentative dialogues, such as [WK95, Pra05]. In [Pra05, PMSW07] the problem of keeping a dialogue focused on a given issue is addressed (while at the same time allowing some liberty of expression to the participants). The notion of *relevant move* is used, which is an assertion that directly affects the issue under discussion, thus advancing the debate. The problem of directly using the results of [Pra05] in a debate platform, is that it focuses on two-party dialogues. Therefore, the problem is not addressed in the more complex, multi-party setting. In such a setting it may be necessary to coordinate the

participants in more elaborate ways, in order to ensure fairness (every user must be given chances to contribute) and advance the debate towards its conclusion.

Moving to the multi-party setting, lately, some works have studied the aggregation of opinions expressed in the form of argumentation systems [CMDK+07, CP11]. For example, in [CMDK+07] the problem addressed is roughly the following: given a set of abstract argumentation systems which share some arguments and attacks, but also differ in some other arguments and attacks, how can we aggregate them into one or more argumentation systems? If every initial argumentation system is the viewpoint of a user, then the aggregated argumentation system must represent a meaningful collective viewpoint. This work studies how, given all the users' systems as input, an aggregation satisfying desired properties can be computed. This immediate aggregation of opinions is very different than the aggregation which takes place in the course of a debate. In the latter, users do not state all their arguments at once, and they may even hide particular arguments, for various reasons: they may contain valuable information, or they may be used against them later in the debate. So, in the general case, during a debate users refrain from disclosing all the information they have.

Finally, there are works which study multi-party dialogues, as for example [AM02], but not many of them define specific debate protocols and analyze their properties. A work which does this is found in [BM11]. Like in [CMDK+07, CP11], the users are assumed to have private argumentation systems representing their viewpoint over the debate, but the difference is that the users contribute to the debate by stating, step-by-step, arguments and attacks. The debate is focused on a specific argument (issue) and it is regulated by a specific protocol, which enforces the turn-taking and only allows relevant moves to be played. Also, the authors propose a simple way to partition users in two opposing groups, something which is considered in previous works like [Pra05], but not studied for multi-party dialogues.

There are still many issues which have not been addressed in multi-party argumentative debates. For example, are there more elaborate ways to aggregate the users' opinions than the one used in [BM11]? What if the users want to use acceptability semantics which differ from the one used in [BM11]? In that case, which protocols could keep the debate focused and also properly coordinate the users? How could we evaluate the conclusions of the debates? Also, how could the users employ strategies, helping them to achieve their goals? We will try to address some of these issues in this work.

Let us present an outline of this work's contribution to multi-party argumentative debates.

Representing and using expertise in debates:
The first issue we address is how users with different topics of expertise may take part in an argumentative debate. We take into account that the arguments of a debate may refer to different topics, so we will need a way to represent: (1) the users' expertise and (2) the topics of the arguments in a debate. As in [CMDK+07, BM11] we assume that the experts may have different opinions on some elements of the debate. For example, between two given arguments, some users may believe that an attack holds, while others may believe the opposite. Such disagreements are resolved, during the debate, by letting the users express their opinions in the form of votes. The votes of the users are aggregated, at every step of the debate, leading to the construction of a single argumentation system which represents the debate. The aggregation of the votes relies on a notion of impact, reflecting how important a specific expert's opinion is, on a specific part of the debate. Finally, we address the issue of drawing meaningful conclusions from the aggregated system.

Decision-support for the central authority (mediator) of a debate:
As we have seen before, the conclusions drawn from multi-party debates are often unclear. But, if

a formal way of aggregating the opinions is used, as well as a formal way to compute the accepted arguments, then are the debate's conclusions always clear and unquestionable? We will argue that this is not always the case. Even then, a debate may be deemed controversial, for several reasons.

- *Strong disagreement among users on some points*: For example, on some specific elements of the shared system, the users' opinions may be split.
- *Possibility of radical changes, if the debate continues*: If a modification takes place in some contested elements of the shared system (e.g. an attack gets deleted), then what are the repercussions on the arguments' acceptability? Could some accepted arguments become rejected, and vice-versa?
- *Arguments with undecided status*: In many cases, there are arguments whose status cannot be declared neither accepted, nor rejected (e.g. two mutually attacking arguments). The existence of such *undecided* arguments may not satisfy debating users who require clear conclusions.

Our task is to provide decision-support for a debate's mediator. First, we propose ways in which a debate's mediator can measure the controversy of a debate (an estimation of how much someone should rely on the debate's conclusions). Then, if the mediator estimates that a debate is very controversial and its conclusions are unreliable, we help him decide which additional experts should be asked to contribute to the debate. The tricky part is that we assume that these users' expertise is known, but that their actual opinions on the debate are not. Therefore, we must be careful, as some users may turn the debate even more controversial than before.

Decision-support for the participants of a debate:
In a debate, each participant usually has an opinion on the issue under discussion, and he wants it to be adopted, by the other participants, at the end of the debate. This type of dialogical interaction is called *persuasion* in [WK95]. In order to guide the participants who are pursuing their goals (of persuading others) we define and study a type of dynamic argumentation system. It is an abstract argumentation system containing attacks which can be added and removed (as the result, for example, of participants voting on them).

For such a system, we study the modifications which bring about different effects (e.g. the acceptance or rejection of a given argument). Then, we particularly focus on the study of *minimal* change (in the sense of minimal sets of attack additions and removals) bringing about an effect. We define and implement an algorithm which computes all the minimal changes (under the name of *target sets*) leading to specific types of effects. Then, different properties of target sets are studied. Next, we focus on the usefulness of target sets in debates. In a sense, target sets indicate the most economical and simplest way to achieve a goal, so intuitively it seems that they can be valuable in debate strategies. We check whether target sets can be used to define dominant strategies in debates. Finally, we analyze in what sense focusing on target sets is advantageous for the participants of a debate, by relying on properties describing the target sets' evolution.

Definition and analysis of debate protocols:
Based on the above, we proceed to the definition and analysis of new multi-party debate protocols. A debate protocol specifies a number of elements, and among them: the turn-taking, the permitted moves at every point, and the winner of the debate. In order to make specific choices for these elements, we must answer some important questions. Will the users be partitioned in groups, based on their goals? If yes, in what way? How can a debate keep its focus on the issue, and at the same time, offer flexibility of expression to the users? How can the central authority of the debate coordinate the users and ensure fairness, while keeping the debate focused? One critical element to answer the above questions is the nature of the argumentation system shared by the users. We explore different possibilities, and we define protocols based on two types of systems. Firstly, we consider systems which allow for attack and support relations between arguments and where argument acceptability is based on a numerical valuation. Secondly, we use argumentation systems where argument acceptability depends on the arguments' inclusion in some extensions

(sets of arguments) which collectively defend themselves from attacks. These different types of systems will lead to the definition of different types of debate protocols.

Subsequently, we study the properties of our protocols, such as termination, determinism and convergence to specific "collectively rational" outcomes. For every protocol, we examine if the users can act strategically in order to increase their chances of winning, and in what ways. We focus on strategies which are based on the situation of the shared system, and we assume, for simplicity, that no coordination takes place among the users. Then, we conduct an important number of experiments (simulations of debates). The results obtained from the experiments are used to evaluate the users' strategies, with respect to different criteria.

Let us briefly present the content of the different chapters. Chapter 1 contains the introduction. Chapter 2 provides some background on abstract argumentation systems. In Chapter 3 we propose a way to aggregate different experts' opinions during a debate, in the form of a single weighted argumentation system. Chapter 4 provides some background on the dynamics of abstract argumentation systems. In Chapter 5 we define a type of abstract argumentation system which can structurally change, by the addition and removal of attacks. We study how (minimal sets of) attack additions and removals can turn an argument accepted or rejected. Also, we identify and address controversial debates. Chapter 6 provides some background on argumentative dialogues. Finally, in Chapter 7 we define, analyze and evaluate different protocols for multi-party argumentative debates.

The following diagram is meant to help the reading of this document. It illustrates which chapters are based on which others. Chapters 3, 5, 7 (in bold) contain original work.

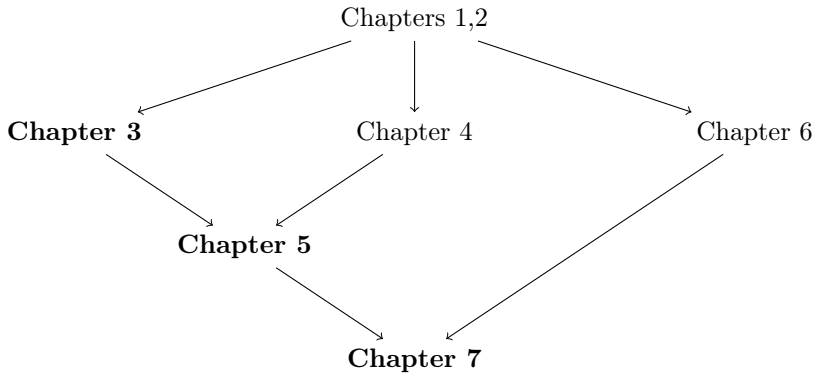


Figure 1.3: The suggested reading order of the chapters. $X \to Y$ suggests to read X before Y.

Finally, in Appendix A we provide a table containing the basic notations used in this work.

Part I

Computational Argumentation

Chapter 2

Basics of Argumentation

Argumentation is a form of reasoning which resembles the way humans reason when trying to answer non-trivial questions. The history of argumentation is very long [Rya84], as it has been both the object of research, as well as an essential tool, of scientists across many different disciplines (e.g. Philosophy, Logic, Law and Artificial Intelligence), as it studies how valid conclusions can be drawn, in presence of some information which can be contradicting and incomplete. Nowadays, there is a range of directions where argumentation seems able to play an important role. Let us provide some examples, which show that argumentation may be helpful in settings with software and/or human agents. First, it can be part of the reasoning module of software agents, especially in new, decentralized multi-agent applications. Also, it can help software agents explain their decisions to human agents, and provide the interface in dialogues between them. Moreover, as we saw in the introduction, argumentation is omni-present in the Web and in social media, though still usually unstructured and affronting several challenges. Findings in formal argumentation can be used to improve these interactions.

The basis of argumentation is the notion of **argument**, an entity which is (very often) considered to be formed by some premises, and by a conclusion which is supported by those premises. The second basic notion, indispensable for arguing in the presence of conflicting information, is the notion of **attack**. An attack from argument x to argument y indicates a conflict between either: the conclusion of x and the conclusion of y (rebutting attack), or the conclusion of x and a premise of y (undercutting attack). The notion of attack is an important focal point of this work, as we will see that it may be the object of controversy. Except from the basic attack relation, other types of relations have been introduced in argumentation systems, such as preference relations and defeat relations [ADM08, Mod09], or support relations [ACLSL08, CLS05c]. Also, some works have studied more complex (n-ary) relations among arguments [BW10, CMDM06].

Argumentation can be used by a single agent in order to reason about his beliefs, goals and possible actions. But, an even more complex and interesting form of argumentation is when arguments are coming from a number of different agents. In a multi-agent context, agents with different beliefs, preferences and goals will usually put forward different arguments and they will have diverging opinions on arguments and attacks between them. Walton and Krabbe in [WK95] proposed a categorization of the different types of agent dialogues, based on the goals which the agents try to achieve. They identified several types of dialogues: persuasion dialogues, negotiation dialogues, deliberation dialogues, information seeking dialogues, inquiry dialogues and eristic dialogues. Argumentation may prove helpful in many types of dialogues, for example in negotiation dialogues [KM06], as well as in deliberation dialogues leading to a decision [KM03]. In our work we study **persuasion dialogues**. These dialogues are focused on a specific issue (e.g. a particular statement, or argument) upon which the agents disagree (e.g. some believe it must be accepted, while others believe it must be rejected). Every agent debates with the goal of convincing the others about the correctness of his point of view on the issue. Not surprisingly, argumentation is the centerpiece of persuasion dialogues, and [Pra05, Pra06] provide good overviews.

2.1 Abstract Argumentation

Abstract argumentation is a type of argumentation which has received a great deal of interest since the work of Dung in [Dun95]. The idea behind it is that, given an argumentation system, we abstract from the content and the structure of its arguments. The result is that, henceforth, the arguments are simply considered as abstract entities, and they can be represented as the nodes of a graph.

In Subsection 2.1.1 and Subsection 2.1.2 we provide an overview of abstract argumentation systems which consider only one type of relation: the so-called *attack* relation over arguments. Then, in Subsection 2.1.3, we will see another type of relation which can be defined, the *support* relation over arguments.

2.1.1 Abstract systems with attack relations (only)

The simplest type of abstract argumentation system contains just a set of abstract arguments and an attack relation over arguments.

Definition 1 *[Dun95] An* **abstract argumentation system** *is a tuple $AS = \langle A, R\rangle$, where A is a set of abstract arguments and $R \subseteq A \times A$ is a binary attack relation over arguments. aRb means that a attacks b (or, in other words, that b is attacked by a). Given an argumentation system, its corresponding* **argumentation graph** *is the digraph whose nodes are the arguments in A and whose arcs are the attacks in R.*

Based on an argumentation graph, we need a reasoning process to decide about the status of any specific argument (or group of arguments). We can identify two main types of reasoning processes: The first type of process is based on extensions (sets of arguments) which satisfy some desired properties, as proposed in [Dun95]. A variation of the above process, analyzed in [Cam06], uses the notion of argument labelling. The second type of process is based on the computation of a gradual valuation of the arguments, as in [BGW05, CLS05b]. We shall first focus on the extension-based and labelling-based approaches, and then on the approaches using gradual argument valuation. Before starting our overview of acceptability semantics in argumentation, let us mention that in *Logic Programming*, the task of defining acceptability semantics has been addressed in works such as [KMD94].

Extension-based acceptability for abstract argumentation

In Dung's framework, the *acceptability of an argument* depends on its membership to some sets, called extensions. The reason is that an argument's evaluation does not depend on the argument itself, but on its relations to the other arguments. We begin by stating two key notions. The first is the notion of conflict-free set of arguments. The second is the notion of acceptability of an argument with respect to a set of arguments.

Definition 2 *[Dun95] Let $AS = \langle A, R\rangle$ be an argumentation system and let $C \subseteq A$.*

- *The set C is* **conflict-free** *iff $\nexists a, b \in C$ such that aRb.*
- *An argument $a \in A$ is* **acceptable with respect to** *C iff $\forall x \in A$: $xRa \Rightarrow \exists y \in C$ such that yRx.*

Several types of semantics for extensions have been defined in [Dun95], and many more since. An overview can be found in [BG09]. In the following we present some of the most commonly used semantics.

Definition 3 *[Dun95] Let $AS = \langle A, R\rangle$ be an argumentation system and let $C \subseteq A$ be a conflict-free set of arguments.*

- *C is an* **admissible extension** *if and only if each argument of C is acceptable with respect to C.*
- *C is a* **preferred extension** *if and only if it is a maximal (with respect to $\subseteq$) admissible extension.*
- *C is a* **complete extension** *if and only if $\forall x \in A$: x is acceptable with respect to $C \Rightarrow x \in C$.*
- *C is a* **stable extension** *if and only if $\forall x \notin C$, $\exists y \in C$ such that yRx.*
- *C is a* **grounded extension** *if and only if it is the minimal (with respect to $\subseteq$) complete extension.*

In the following, let the elements of the set $S = \{Adm, Pref, Comp, St, Gr\}$ respectively denote the admissible, preferred, complete, stable and grounded semantics. In this work we will focus on these semantics, though we note that a number of additional ones have been proposed in the literature. In [BCG11] an overview of many proposed semantics can be found. We shall denote the set of extensions corresponding to semantics S, by $\mathcal{E}_S$. For any semantics in S, there are two main types of argument acceptability: *credulous* and *skeptical*.

Definition 4 *Let $AS = \langle A, R \rangle$ be an argumentation system, and let $a \in A$. Argument a is called:*

1. **Credulously accepted** *with respect to AS under semantics S, denoted $S_\exists(a, AS)$, if and only if a belongs to at least one extension of AS under the S semantics.*
2. **Skeptically accepted** *with respect to AS under semantics S, denoted $S_\forall(a, AS)$, if and only if a belongs to all the extensions of AS under the S semantics.*
3. **Rejected** *with respect to AS under semantics S, denoted $S_\nexists(a, AS)$, if and only if a does not belong to any extension of AS under the S semantics.*

We conclude with some remarks from [Dun95]. Firstly, the connections between the previously defined acceptability semantics are illustrated in Figure 2.1.

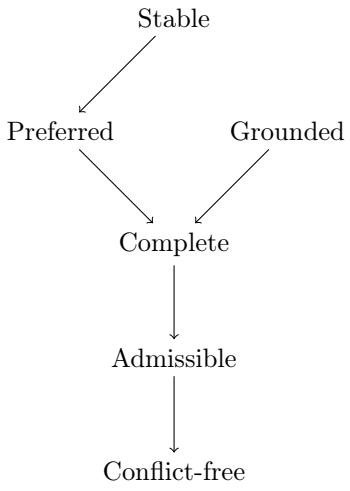


Figure 2.1: The connections between some basic acceptability semantics. $X \rightarrow Y$ should be translated as "if an extension is X, then it is also Y".

Also, it holds that the empty set is always an admissible extension, so skeptical acceptability under admissible semantics is a vacuous notion. Moreover, there always exists a *unique* grounded

extension (which may be empty), so there is no difference between credulous and skeptical acceptability for grounded semantics. Finally, an argument $a \in A$ belongs to the grounded extension if and only if it is skeptically accepted under the complete semantics.

Example 1 *Let us see the different extensions of the abstract argumentation framework* $\langle A, R\rangle$*, where* $A = \{a, b, c, d, e, f, g\}$ *and* $R = \{(a, b), (b, c), (b, d), (c, d), (c, e), (d, c), (e, f), (f, g), (g, e)\}$.

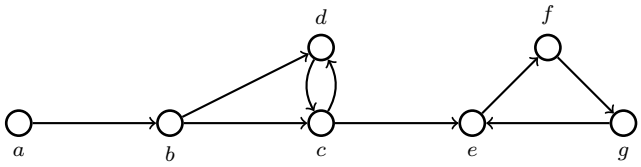


Figure 2.2: The abstract argumentation framework $\langle A, R\rangle$

This example is interesting as the argumentation graph contains two important types of subgraphs, an even-length attack cycle[1] *(c-d-c), and an odd-length attack cycle*[2] *(e-f-g-e). We begin with the preferred semantics. There are two preferred extensions, more specifically:* $\mathcal{E}_{Pref} = \{\{a, d\}, \{a, c, f\}\}$*. Notice that no argument of the odd-length cycle (e-f-g-e) can be added into* $\{a, d\}$*, and still give an admissible extension. As far as complete extensions are concerned, it holds that* $\mathcal{E}_{Comp} = \{\{a\}, \{a, d\}, \{a, c, f\}\}$*, because each one of these three sets includes every argument which is acceptable with respect to that set. Stable extensions attack all the arguments which do not belong to them, therefore we have* $\mathcal{E}_{Stab} = \{\{a, c, f\}\}$*. Finally, the grounded semantics is in a sense the most skeptical one, as the grounded extension includes only arguments which are found in all complete extensions. Therefore,* $\mathcal{E}_{Gr} = \{\{a\}\}$.

As far as credulous and skeptical acceptability is concerned, focusing e.g. on preferred semantics, it holds that: Argument a is skeptically accepted under the preferred semantics ($Pref_{\forall}(a, AS)$*). Also, arguments c, d and f are credulously accepted under the preferred semantics (*$Pref_{\exists}(c, AS)$*,* $Pref_{\exists}(d, AS)$*,* $Pref_{\exists}(f, AS)$*). Finally, arguments b, e and g are rejected under the preferred semantics (*$Pref_{\nexists}(b, AS)$*,* $Pref_{\nexists}(e, AS)$*,* $Pref_{\nexists}(g, AS)$*).*

The following property from [Dun95] states that the set of arguments credulously accepted under the admissible semantics are the same as those credulously accepted under the preferred, or the complete semantics.

Property 1 [Dun95] *Let* $AS = \langle A, R\rangle$ *be an abstract argumentation system, and let* $a \in A$*. Then it holds that* $Adm_{\exists}(a, AS) \Leftrightarrow Pref_{\exists}(a, AS) \Leftrightarrow Comp_{\exists}(a, AS)$.

Proof 1 *This proof has four points:*

1. $Adm_{\exists}(a, AS) \Rightarrow Pref_{\exists}(a, AS)$*: If a belongs to an admissible extension, then a also belongs to a maximal w.r.t.* $\subseteq$ *admissible, and thus preferred, extension.*

2. $Pref_{\exists}(a, AS) \Rightarrow Adm_{\exists}(a, AS)$*: As every preferred extension is admissible, if a belongs to a preferred extension, a also belongs to an admissible one.*

3. $Pref_{\exists}(a, AS) \Rightarrow Comp_{\exists}(a, AS)$*: As every preferred extension is complete, if a belongs to a preferred extension, a also belongs to a complete one.*

4. $Comp_{\exists}(a, AS) \Rightarrow Pref_{\exists}(a, AS)$*: Given that a belongs to a complete extension, that extension is also admissible, so a is in an admissible extension. Thus, from point (1) above, a belongs to a preferred extension.*

[1] An even-length cycle of two attacks is called *Nixon diamond* in the literature.
[2] Odd-length cycles have no admissible subset of arguments.

The case of skeptical acceptability is a bit different. Every argument skeptically accepted under complete semantics is also skeptically accepted under preferred semantics, but the inverse does not hold in the general case.

Property 2 [Dun95] *Let $AS = \langle A, R\rangle$ be an abstract argumentation system, and let $a \in A$. It holds that $Comp_\forall(a, AS) \Rightarrow Pref_\forall(a, AS)$. The inverse does not hold, in the general case.*

Proof 2 *This proof has two points:*

1. *$Comp_\forall(a, AS) \Rightarrow Pref_\forall(a, AS)$: As every preferred extension is also a complete one, if a belongs to all the complete extensions, then it also belongs to all the preferred ones.*
2. *$Pref_\forall(a, AS) \not\Rightarrow Comp_\forall(a, AS)$: There follows a counter-example: $A = \{a, b, c, d\}$, $R = \{(a, b), (b, a), (a, c), (b, c), (c, d)\}$. Here, d is skeptically accepted under preferred semantics, as $\mathcal{E}_{Pref} = \{\{a, d\}, \{b, d\}\}$, but it is not skeptically accepted under complete semantics, as $\mathcal{E}_{Comp} = \{\{\}, \{a, d\}, \{b, d\}\}$.*

Labelling of arguments (and attacks)

Another approach for evaluating the status of arguments, directly related to the extension-based approach, is argument labelling. In this approach, a label is assigned to each argument of a system $AS = \langle A, R\rangle$. Three possible values for labels are often used: *in*, *out*, and *undec*, as in [Cam06], though some approaches consider additional values [JV99]. If the label of argument $a \in A$ is *in* (resp. *out*, *undec*), this is denoted by $lab(a) = in$ (resp. $lab(a) = out$, $lab(a) = undec$). We note that *in* stands for *accepted*, *out* stands for *rejected*, and *undec* stands for *undecided*.

We shall denote a labelling of system AS as follows: $\langle lab^{AS}(in), lab^{AS}(out), lab^{AS}(undec)\rangle$, where $lab^{AS}(in) \subseteq A$ (resp. $lab^{AS}(out) \subseteq A$, $lab^{AS}(undec) \subseteq A$) is the subset of arguments having the *in* (resp. *out*, *undec*) label.

The two basic postulates in [Cam06] that a labelling must satisfy in order to constitute a **reinstatement labelling** are the following:

- $\forall a \in A$: $lab(a) = out$ if and only if $\exists b \in A$ s.t. $lab(b) = in$ and bRa.
- $\forall a \in A$: $lab(a) = in$ if and only if $\forall b \in A$ s.t. bRa it holds that $lab(b) = out$.

For any given argumentation system, there may exist one or more possible reinstatement labellings. As said before, there is a direct link between the extension-based approach for acceptability and the labelling-based approach. This link is studied in detail in [Cam06]. Here we only provide some examples:

- Reinstatement labellings with no *undec* labels, coincide with stable extensions.
- Reinstatement labellings where *in* (or *out*) labels are maximal, coincide with preferred extensions.
- A reinstatement labelling where *undec* labels are maximal, coincides with the grounded extension.

Example 1 (cont.) *Let us consider again the system $\langle A, R\rangle$. Three possible reinstatement labellings of this system are illustrated in Figure 2.3.*

Let us first see if there is a reinstatement labelling where undec is empty. Indeed, this is the case of the labelling $L_1 = \langle\{a, c, f\}, \{b, d, e, g\}, \{\}\rangle$. We see that such a labelling corresponds[3] to the stable extension $\{a, c, f\}$. Next, let us check the reinstatement labellings with maximal in (and maximal out) labels. There are two such reinstatement labellings: $L_1 = \langle\{a, c, f\}, \{b, d, e, g\}, \{\}\rangle$

[3] If we consider just the *in* arguments.

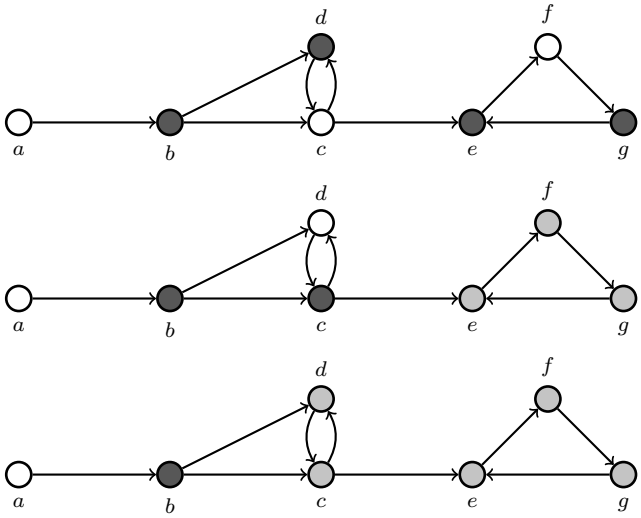


Figure 2.3: Three possible reinstatement labellings of $\langle A, R\rangle$. The *in* (resp. *out*, *undec*) arguments are drawn in white (resp. black, grey). The first two labellings correspond to the two preferred extensions, while the third labelling corresponds to the grounded extension.

and $L_2 = \langle\{a, d\}, \{b, c\}, \{e, f, g\}\rangle$. *Indeed, as commented above, they correspond to the preferred extensions* $\{a, c, f\}$ *and* $\{a, d\}$. *Finally, the reinstatement labelling which maximizes the undec labels is* $L_3 = \langle\{a\}, \{b\}, \{c, d, e, f, g\}\rangle$. *It corresponds to the grounded extension* $\{a\}$.

Villata et al. in [VBvdT11] have introduced the *attack semantics* where the focus is shifted towards the acceptability status of attacks. Instead of attributing labels to arguments, the authors of [VBvdT11] attribute labels to attacks. Roughly:

- An attack (a, b) is 1 when a is labelled *in*.
- An attack (a, b) is ? when a is labelled *undec*.
- An attack (a, b) is 0 when a is labelled *out*.

An attack is called *successful* when it is 1 or ?, and *unsuccessful* when it is 0.

Example 2 *Let* $AS = \langle A, R\rangle$ *be an abstract argumentation system, with* $A = \{a, b, c\}$, $R = \{(a, b), (b, c), (c, b)\}$.

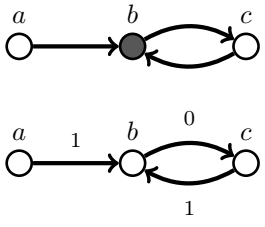


Figure 2.4: Above: A labelling of the arguments of AS (*in* arguments are in white, *out* arguments in black). Below: A labelling of the attacks of AS.

There is a single reinstatement labelling for the system AS which is $\langle\{a, c\}, \{b\}, \{\}\rangle$. *In attack semantics, the attacks* $(a, b) \in R$ *and* $(c, b) \in R$ *are successful, whereas* $(b, c) \in R$ *is unsuccessful.*

Gradual valuation of arguments

We will mainly focus on [CLS05b], where the authors present a general framework where every abstract argument, depending on the attacks it receives, is assigned a gradual valuation. We specify that the arguments have no initial (intrinsic) valuations. When the attacks they receive are taken into consideration, then they are assigned gradual valuations. In a gradual valuation, given an argument a, its valuation $v(a)$ is an element of a totally ordered set which has a minimum and a maximum element (usually an interval of the real numbers, e.g. $[0, 1]$).

Actually, they propose two different types of gradual valuation:

- **Local valuation**: where the valuation of an argument depends on the valuations of the arguments which are set to attack it (called *direct attackers*).
- **Global valuation**: where the valuation of an argument depends on the set of *attack branches* leading to it.

The basic difference between local and global valuation is the following: Let x be an argument of the argumentation graph. A defence (resp. attack) branch of x is a path of attacks leading to a, and containing an even (resp. an odd) number of attacks. Under the perspective of the *local valuation*, if some (new) defence branches for x are inserted on the graph, then the valuation of x will *decrease*. The explanation is that, even if those new branches are defence branches, the existence of new attackers against x should decrease its valuation. On the other hand, under the perspective of the *global valuation*, the addition of some defence branches for x has the opposite effect, it increases the valuation of x. Here the explanation is that x has successfully defended itself from a number of attackers, therefore its valuation should be increased.

In our work, we will focus on local valuation, but for the sake of completeness of our overview, let us first highlight the basics of global valuation, as presented in [CLS05b].

In the case of global valuation, an argument's valuation depends on two things: on the *number* and on the *size* of its defence and attack branches. In [CLS05b] the valuation of an argument is a tuple of the form $[(d_1, d_2, \ldots, d_n), (a_1, a_2, \ldots, a_m)]$ where $(d_1, d_2, \ldots, d_n)$ are the lengths of its defence branches, and $(a_1, a_2, \ldots, a_m)$ are the lengths of its attack branches.

In order to compare the valuations of two arguments, the authors propose the following approach: (i) Compare the numbers of their defence branches, and the numbers of their attack branches (therefore we have two criteria, which must be aggregated). In case the two arguments contain the same number of defence branches and the same number of attack branches, then (ii) compare the quality of the defence and attack branches, by taking into consideration their sizes (so we have, again, two criteria which must be aggregated). Based on this comparison of argument valuations, an argument x can be said to be *better*, *worse*, or *incomparable* to another argument y.

Finally, the authors address the complex case of graphs which contain cycles. Their proposed solution is to identify an infinity of attack and/or defence branches, which are then used in order to compute the arguments' valuations.

Local valuation of arguments is an approach more commonly used than global valuation. As mentioned before, in this case an argument's valuation depends only on the valuations of its direct attackers. More specifically, in [CLS05b] the authors propose a *generic gradual valuation* as follows:

Definition 5 *[CLS05b]* **(Generic gradual valuation)**
Let $\langle A, R\rangle$ be an argumentation system. Also, let $(W, \geq)$ be a totally ordered set with a minimum element (V_{Min}) and let a subset $\mathcal{V}$ of W, which contains V_{Min} and has a maximum element V_{Max}. A **valuation** *is a function $v : A \to \mathcal{V}$ such that:*

1. *$\forall a \in A$, $v(a) \geq V_{Min}$.*
2. *$\forall a \in A$, if a has no attackers, then $v(a) = V_{Max}$.*

3. $\forall a \in A$, *if* a *has the attackers* $a_1, \ldots, a_n$, *then* $v(a) = g(h(v(a_1), \ldots, v(a_n)))$.[4]

Where $h : \mathcal{V}^* \rightarrow W$ *such that:*[5]

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

And $g : W \rightarrow \mathcal{V}$, *such that:*

- $g(V_{Min}) = V_{Max}$
- $g(V_{Max}) < V_{Max}$
- g *is non-increasing (if* $x \leq y$ *then* $g(x) \geq g(y)$*)*

Some of the properties of this generic gradual valuation are given in [CLS05b]. We underline that the valuation is maximal for arguments without attackers, and that the valuation of an argument is a non-increasing function of the valuation of its direct attackers.

Once the arguments' valuations have been computed, the question is which arguments should be considered acceptable. There are two main choices:

- *Acceptability based only on valuations:*
  We indicate two possible choices: (i) accept every argument with a valuation higher than some threshold value (e.g. 0.5), or (ii) accept every argument whose valuation is greater than the average argument valuation.
- *Acceptability based on extensions (and valuations):*
  Regarding acceptability based on extensions (and valuations), in [CLS05b] the authors propose two options: The first option is to compute extensions in Dung's classical way, and then separate the arguments into *uni-accepted* (or skeptically accepted), *exi-accepted* (or credulously accepted), *cleanly-accepted* [6] and *not-accepted*. Finally, the arguments of the same class can be compared and ordered, based on their valuations. The second option is to compute extensions using a notion of defence which takes into account the arguments' valuations. A *well-defended argument* is an argument not receiving any attacks from arguments with a greater valuation. So, three classes of arguments can be identified: not-attacked, attacked and well-defended, attacked and not well-defended. As before, the arguments of the same class can be compared and ordered, based on their valuations.

Finally, a work which proposes a different way to compute valuations of abstract arguments is found in [MT08]. A game-theoretic notion of argument strength is used. The main idea is that, given an abstract argumentation system $\langle A, R \rangle$ and an argument $a \in A$, two players (a proponent and an opponent of a) play a game as follows. Each player considers different ways to choose a subset of arguments. The opponent of a may choose a subset of arguments "against" a, and the proponent of a may choose a subset of arguments "for" a. Given a choice by the players, a numerical *payoff* is computed for each player, based on how many of his arguments got attacked by the other player, and on how many arguments of the other player he attacked. The valuation of a is computed, based on the different possible payoffs.

[4] The intuition behind this equation in that function h computes the combined "strength" of a sequence of attackers, while function g computes the valuation of an argument, given the "strength" of the sequence of its attackers.

[5] $\mathcal{V}^*$ denotes the set of all the finite sequences of elements of $\mathcal{V}$.

[6] An argument is cleanly-accepted if it has no attacker which belongs to some extension.

2.1.2 Abstract systems with preferences and weights

We have just presented three ways to compute argument acceptability, given an argumentation system featuring only a binary attack relation. But when we reason with abstract argumentation systems, we face an inconvenience which is troublesome in some cases: it is not possible to take into consideration the strength (importance) of the different arguments and attacks, as they are just the nodes and the arcs of a digraph. This may lead us sometimes to debatable conclusions. Here we present some of the propositions in the literature whose goal is to attach this type of information into an abstract argumentation system. Usually this is achieved with the help of a **preference relation** over arguments and/or attacks. In some works these preferences are expressed in the form of numerical values attached to arguments and/or attacks, called **weights**. These systems are called **Weighted Argumentation Systems (WAS)**[7]. The authors in [DHM$^+$11] offer a good overview of the approaches existing in the literature, to which we will add some further references and remarks.

Preferences over Arguments

We first focus on approaches which express argument strength with the help of a *preference* relation over arguments (without using numerical weights). One of the first works to use such a preference relation was [PS97]. In that work, preference over arguments depends on their structure, as arguments are formed from *strict rules* and *defeasible rules* (over which another preference relation is defined). In a similar setting, in [AC02] the notion of Preference-based Argumentation Framework (PAF) is defined as follows:

Definition 6 *[AC02] A* **Preference-based Argumentation Framework (PAF)** *is a triple* $\langle A, R, Pref \rangle$ *where* A *is a set of arguments,* $R \subseteq A \times A$ *is a binary attack relation over arguments, and* $Pref$ *is a (partial or complete) preordering over* A*.*

In that work, similarly to [PS97], (as well as [AMP00, GS04]) an argument's strength depends on the strength of its building blocks, while argument strengths are essential in order to compute a defeat relation D over the arguments. Also, in [KM03] the authors use relative argument strengths, based on priorities over rules (which are constituents of arguments). The *Gorgias-C* system [NK09] can be used to implement such settings. Also, argument structure plays a crucial role for defining argument acceptability in the ASPIC system [ABC$^+$06]. Finally, in [Pra10] the author analyzes several issues regarding argument acceptability in settings with structured arguments.

There are also works which define and use a notion of argument strength, while abstracting from the arguments' structure. An important example, similar to the PAFs, are the Value-based Argumentation Frameworks (VAFs) [BC03], where every abstract argument promotes a specific *social value* (e.g. peace, economic growth, justice, freedom). Since every agent has a personal preference over these values, he also has a personal preference over the arguments. Then, similarly to the previous works, a binary defeat relation over arguments takes into consideration both the notion of (abstract) attack, and the notion of preference over arguments. More specifically, in [BC03] a Value-based Argumentation Framework (VAF) is defined as follows:

Definition 7 *[BC03] A* **Value-based Argumentation Framework (VAF)** *is a 5-tuple* $\langle A, R, V, val, valprefs \rangle$ *where* A *is a set of arguments,* R *is a set of binary attacks over arguments,* V *is a non-empty set of values,* val *is a function which maps from elements of* AR *to elements of* V*, and* $valprefs$ *is a preference relation (transitive, irreflexive and asymmetric) over* V*.*

Example 1 (cont.) *Let us enrich the previous abstract framework* $\langle A, R \rangle$*, by attaching two values (*v_1 *and* v_2*) to its arguments. We define the* $VAF = \langle A, R, V, val, valprefs \rangle$ *where:*
$A = \{a, b, c, d, e, f, g\}$,

[7] There is no single definition of Weighted Argumentation System, because (as we will see) some weighted systems only have weighted attacks, others only have weighted arguments, and finally others have both weighted attacks and weighted arguments

$R = \{(a,b),(b,c),(b,d),(c,d),(c,e),(d,c),(e,f),(f,g),(g,e)\}$,
$V = \{v_1, v_2\}$,
$val(a) = val(f) = v_1$, $val(b) = val(c) = val(d) = val(e) = val(g) = v_2$,
$valprefs = \{(v_1, v_2)\}$.

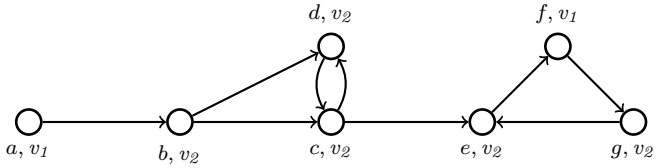


Figure 2.5: The digraph showing the attack relation over the arguments in the VAF. Every argument has a value in $V = \{v_1, v_2\}$.

In [BC03], an argument $a \in A$ is said to **defeat** *an argument $b \in A$ if and only if both aRb and not $valprefs(val(b), val(a))$. The attack relation is shown in Figure 2.5 and the defeat relation in Figure 2.6. Notice that there is no arc from e to f, since $(v_1, v_2) \in valprefs$. As a result, following [BC03], the V-preferred extensions of this framework are: $\{a, c, f\}$ and $\{a, d, e, f\}$. Notice that the extension $\{a, d, e, f\}$ did not exist in the initial (non-valued) framework.*

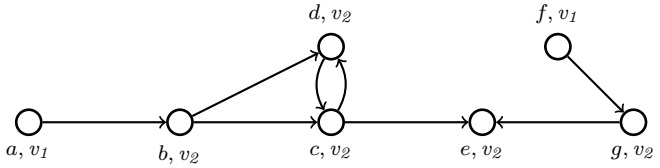


Figure 2.6: The digraph showing the defeat relation over the arguments in the VAF. Notice that e does not defeat f, because the value promoted by f (v_1) is preferred to the value promoted by e (v_2).

Moreover, Extended Argumentation Frameworks (EAFs) proposed in [Mod09], provide a way to define (and reason about) preferences over arguments. In EAFs, arguments are allowed to attack other attacks.

Definition 8 *[Mod09] An* **Extended Argumentation Framework (EAF)** *is a triple $\langle A, R, RR\rangle$ where A is a set of arguments, $R \subseteq A \times A$ is a binary attack relation (over arguments), $RR \subseteq A \times R$ is a binary attack relation (from arguments to attacks), and if $(X,(Y,Z)),(X',(Z,Y)) \in RR$, then (X,X'), $(X',X) \in R$.*

The intuition behind the last point of the definition of an EAF is that, if argument X attacks (Y,Z), then this is like argument X stating a preference for Z over Y, and similarly, if argument X' attacks (Z,Y), then this is like argument X' stating a preference for Y over Z. In this case, the arguments X and X' are considered to attack each other, thus (X,X'), $(X',X) \in R$.

Example 3 *Let $a, b \in A$ be two arguments which attack each other. A preference of a over b could be expressed as follows: An argument c is introduced and it is set to attack the attack (b,a). The above is captured by the following $EAF = \langle A, R, RR\rangle$, where: $A = \{a,b,c\}$, $R = \{(a,b),(b,a)\}$ and $RR = \{(c,(b,a))\}$. As a result, the only extension of the EAF is $\{c,a\}$.*

Finally, preferences over arguments can be used not only for the computation of extensions, but also afterwards, in order to refine the choice of extensions. This can be particularly useful in the context of decision-making problems (two examples of such works are found in [ADM08, DMA09]),

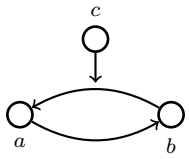
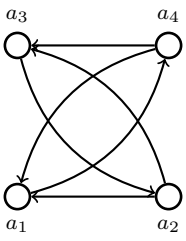


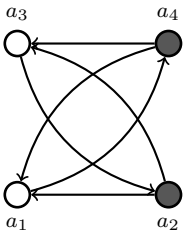
Figure 2.7: The above EAF has a single preferred extension which is $\{a, c\}$.

where it seems that some choices (extensions) must be discarded, even if they are reasonable. In [DMA09] the authors provide the following example:

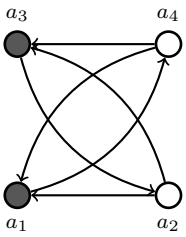
Example 4 *[DMA09] Let $\langle A, R, Pref\rangle$ denote an argumentation system with a preference relation over arguments, where: $A = \{a_1, a_2, a_3, a_4\}$, $R = \{(a_1, a_2), (a_2, a_1), (a_1, a_4), (a_4, a_1), (a_3, a_2), (a_2, a_3), (a_3, a_4), (a_4, a_3)\}$, and finally $Pref = \{(a_2, a_1), (a_4, a_3)\}$ is a binary preference relation over arguments.*



(a) The defeat relation $D \subseteq R$.



(b) The first stable extension is $\{a_1, a_3\}$



(c) The second stable extension is $\{a_2, a_4\}$

Figure 2.8: The system $\langle A, R, Pref\rangle$ and its two stable extensions.

As we see in Figure 2.8, there are two stable extensions: $\{a_1, a_3\}$ and $\{a_2, a_4\}$. In [DMA09] the authors argue that $\{a_2, a_4\}$ is the extension which should be preferred over the other one. The intuition behind this statement is that there exist arguments in $\{a_2, a_4\}$ which are preferred to some arguments in $\{a_1, a_3\}$, while there are no arguments in $\{a_1, a_3\}$ which are preferred to some arguments in $\{a_2, a_4\}$.

Preferences over Attacks

The authors of [MGS08] propose argumentation frameworks with varied-strength attacks (AFV). An AFV is defined as a triple $\langle A, Atts, PrefAtts\rangle$ where A is a set of arguments, $Atts$ is a set of binary attack relations (all of them defined over A), and $PrefAtts$ is a binary preference relation defined over $Atts$. An attack relation $R_1 \in Atts$, compared to an attack relation $R_2 \in Atts$, can be:
(i) stronger, if $(R_1, R_2) \in PrefAtts$ and $(R_2, R_1) \notin PrefAtts$, denoted $R_1 \gg R_2$,

(ii) weaker, if $(R_1, R_2) \notin PrefAtts$ and $(R_2, R_1) \in PrefAtts$, denoted $R_1 \ll R_2$,
(iii) equivalent in force, if $(R_1, R_2) \in PrefAtts$ and $(R_2, R_1) \in PrefAtts$, denoted $R_1 \approx R_2$, or
(iv) of unknown difference force, if $(R_1, R_2) \notin PrefAtts$ and $(R_2, R_1) \notin PrefAtts$, denoted $R_1 ? R_2$.
The notion of **defence** is updated, identifying several types of defending arguments (strong, weak, normal, unqualified). As a result, the notion of acceptability is also updated, based on the different types of defending arguments.

Example 5 *Let us consider the argumentation system illustrated in Figure 2.9, but transformed into an AFV, as follows. Let $\langle A, Atts, PrefAtts\rangle$ be an AFV such that: $A = \{a, b, c, d, e, f, g\}$ is a set of arguments, $Atts = \{R_1, R_2\}$ is a set of two binary attack relations (illustrated below), and $PrefAtts = \{(R_1, R_2), (R_1, R_1), (R_2, R_2)\}$ is a binary preference relation over the attack relations.*

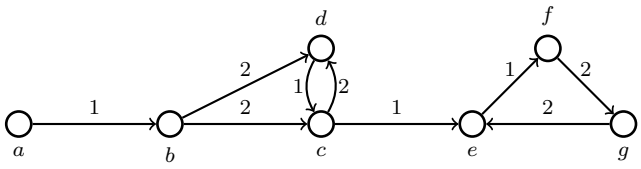


Figure 2.9: The above system $\langle A, Atts, PrefAtts\rangle$ is an AFV.

Following [MGS08] we say that the attack scenario $[\{a, c, f\}, \{\gg, \ll, \approx\}]$ is an admissible scenario, because the set of arguments $\{a, c, f\}$ is conflict-free and every attack against an argument of $\{a, c, f\}$ is answered by a counter-attack from $\{a, c, f\}$ which is stronger ($\gg$), or weaker ($\ll$), or equivalent in force ($\approx$). On the other hand, the attack scenario $[\{a, c, f\}, \{\gg, \approx\}]$ is not an admissible scenario, as (for example) argument c receives the attack dR_1c and $\{a, c, f\}$ does not counter-attack d with a stronger or equivalent in force attack (but only with a weaker one, cR_2d).

The meaning of argument and attack weights

Now let us turn our attention to works using numerical weights in order to model preference. The authors of [DHM$^+$11] highlight the different underlying reasons for attaching weights to arguments and/or attacks of an abstract argumentation system. Such weights convey useful information and according to [DHM$^+$11] they can indicate:

1. **The result of a voting process**: A number of recent works, e.g. [LM11, EML13, KBMM12], are based on this approach which couples Argumentation with Social Choice. The idea is that there exists an argumentation system which is visible to (and perhaps constructed by) a number of agents who are allowed to express their opinions on its arguments and/or attacks, through voting. The votes can be then translated into numerical weights of arguments and/or attacks. In [LM11, EML13, KBMM12] a vote on an argument or attack is either positive or negative. As far as the meaning of a vote's polarity is concerned: In [KBMM12] a positive (resp. negative) vote on an attack indicates that it should be considered valid (resp. invalid), while in [LM11, EML13] a positive (resp. negative) vote on an argument (or attack) indicates that its effect should be increased (resp. decreased).

2. **Different degrees of belief**: An agent can attach a weight to an argument or attack, based on how strongly he initially believes that it holds (e.g. using probabilities as in [Hun12, Hun13]), without taking into consideration the attacks set against it (we note that in [Mod09] attacks can target not only arguments, but other attacks as well). In the case of an attack, the authors in [DHM$^+$11] interpret its weight as the degree of inconsistency between its two arguments.

3. **Different relative strengths**: An agent ranks the arguments, or attacks, based only on their relative strengths (not being interested in their objective strengths). Two examples of this approach are found in [KM03], and in [MGS08], although we underline that, in these works, attacks are not attributed numerical weights.

We underline that in these cases weights are assigned to arguments *a priori*, before considering the arguments' interactions.

Weights on Arguments

One particular type of argumentation frameworks which use weights are the *Probabilistic Argumentation Frameworks*, proposed in [LON12]. In a Probabilistic Argumentation Framework, there is a function which attributes a probability value to every argument, and a second function which attributes a probability value to every attack. Since we will not focus now on probabilities of attacks, let us give the following definition of *probabilistic argument graph* [LON12, Hun13], which only uses probabilities of arguments.

Definition 9 *[LON12, Hun13] A* **probabilistic argument graph** *is a triple* $\langle A, R, w\rangle$ *where* $\langle A, R\rangle$ *is a Dung-style argumentation framework, and* $w : A \to [0, 1]$.

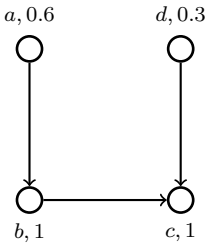
There is no consensus in the literature on a specific meaning for this probability, neither on how it should be computed. We can say that, in general, it is the probability that an argument holds *a priori* or, under another perspective, the probability that it belongs to the argumentation graph. In [Hun13], where an overview of the proposed approaches is provided, the probability of an argument is computed from the probabilities of its premises. Also, two approaches for deciding on argument acceptability are presented: (i) *The epistemic approach*, where argument probabilities are computed "in accordance with" the graph's structure. Therefore, if an argument with a relatively big probability attacks another argument, then the attacked argument obligatorily will have a relatively small probability. (ii) *The constellations approach*, taken from [LON12], which computes the probability than an argument belongs to some extension, based on the probabilities of all the possible subgraphs of the initial argumentation graph. Let us give an example of the constellations approach:

Example 6 *Let* $\langle A, R, w\rangle$ *be a probabilistic argument graph, where* $A = \{a, b, c, d\}$, $R = \{(a, b), (b, c), (d, c)\}$, *and* $w(a) = 0.6$, $w(b) = w(c) = 1$, $w(d) = 0.3$.

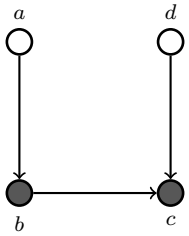
The constellations approach begins from the graph $\langle A, R, w\rangle$ *and computes a set of possible "worlds" (abstract argumentation systems). In every such system, only a subset of the arguments from A appears. In this example, there are 4 different abstract systems we may end-up with (notice that arguments c and d are found in every system, as their probability is* 1*). The initial graph, as well as the 4 abstract systems we may end-up with, are shown in Figure 2.10.*

The probabilities attached to the arguments of $\langle A, R, w\rangle$ *shall help us compute the probabilities of each possible abstract system. In this example, we assume that the probabilities of appearance of a and d are independent. Also, we shall focus on the probability of accepting argument c. Let* A' *denote the set of arguments of an abstract system we may end-up with. Then:*

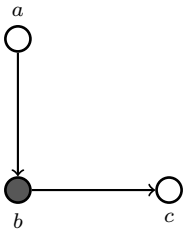
- *Probability* $(A' = \{a, b, c, d\}) = 0.6 \times 0.3 = 0.18$
  In this scenario $lab(c) = out$.
- *Probability* $(A' = \{a, b, c\}) = 0.6 \times 0.7 = 0.42$
  In this scenario $lab(c) = in$.
- *Probability* $(A' = \{b, c, d\}) = 0.4 \times 0.3 = 0.12$
  In this scenario $lab(c) = out$.
- *Probability* $(A' = \{b, c\}) = 0.4 \times 0.7 = 0.28$
  In this scenario $lab(c) = out$.



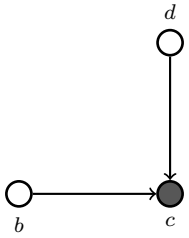
(a) The probabilistic argument graph $\langle A, R, w \rangle$.



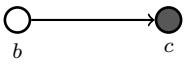
(b) Both a and d are on the system: $lab(c) = out$.



(c) a is on the system, d is not: $lab(c) = in$.



(d) d is on the system, a is not: $lab(c) = out$.



(e) Neither a nor d are on the system: $lab(c) = out$

Figure 2.10: The probabilistic argument graph $\langle A, R, w \rangle$, and its 4 corresponding abstract systems.

Concluding, $lab(c) = in$ with a probability of 0.42, and $lab(c) = out$ with a probability of 0.58.

Next, in the social context of multi-agent debates, the authors of [LM11] consider that, during a debate, the agents create an argumentation graph by putting forward their arguments and attacks, and by voting on these arguments. They introduce the notion of Social Abstract Argumentation Framework (SAF) which is defined as follows:

Definition 10 *[LM11] A* **Social Abstract Argumentation Framework (SAF)** *is a triple $\langle A, R, V \rangle$ where A is a set of arguments, $R \subseteq A \times A$ is a binary attack relation, and $V : A \to \mathbb{N} \times \mathbb{N}$ is a total function mapping each argument to its number of positive and its number of negative votes.*

In [LM11] every argument has an initial and a final valuation. An argument's *initial* valuation depends on the number of positive and negative votes it has received. A positive (resp. negative) vote on an argument increases (resp. decreases) its initial valuation. An argument's *final* valuation depends on: (i) its initial valuation and (ii) the valuations of its direct attackers. The final valuation of argument $a \in A$ is denoted $M(a)$, where $M : A \to L$ is a mapping from the set of arguments into the set L, a totally ordered set with a top and a bottom element, e.g. the set of real numbers $[0, 1]$.

Therefore, in [LM11], an argument's valuation represents its *social acceptance*, *not its logical correctness*. We also mention that, in [EML13] which is an extension of this work, agents can vote both on arguments, and on attacks.

Weights on attacks

Let us begin with the work in [DHM$^+$11], where the source of the attacks' weights is not taken into consideration and an argumentation system with weighted attacks is defined as follows:

Definition 11 *[DHM$^+$11] A* **weighted argument system with weighted attacks** *is a triple* $W = \langle A, R, w\rangle$ *where* $\langle A, R\rangle$ *is a Dung-style abstract argumentation system, and* $w : A \to \mathbb{R}_{\geq}$ *is a function assigning positive real valued weights to attacks.*

The authors use extension-based semantics for deciding on the acceptability of arguments, and they consider that an attack's weight indicates how "problematic" it is to not consider it. A central notion of this work is the notion of **inconsistency budget**. Given an inconsistency budget $n \in \mathbb{N}$ we are allowed to delete attacks whose sum of weights does not exceed the budget n. As there are many choices, we can end up with a set of different argumentation systems (where extension-based acceptability is then applied). Therefore, given a budget n, even by using semantics like the grounded which give a unique extension in the classical setting, we may obtain multiple extensions.

The work in [CMKMO12] is based on [DHM$^+$11], to which the authors make some additions. Firstly, they define an aggregation operator $\oplus$ of a vector of weights which is not obligatorily the sum (as in [DHM$^+$11]), but it can be, for example, the *max*, the *leximax*, or the *leximin*. Based on the $\oplus$ operator, they propose a way to relax extensions by deleting attacks, as in [DHM$^+$11]. Also, based on that operator, they compute the collective (aggregated) attack against an argument and they propose refinements of the notions of defence and acceptability.

The authors in [KL11] propose a way to compute weighted attacks, based on a given preference relation over arguments.

Finally, the authors of [EML13], which is an extension of [LM11], consider that agents can vote not only on arguments, but also on attacks. This leads to systems which have both weighted arguments and weighted attacks.

2.1.3 Bipolar Argumentation Frameworks

The most basic relation between arguments is, undoubtedly, the attack relation. But there exist approaches that take into account another type of binary relation over arguments, which can be seen as the opposite of the attack relation: the **support** relation. Intuitively, we say that argument a supports argument b (denoted aSb) if believing in the conclusion of argument a gives grounds to believe more in argument b (and therefore in its conclusion).

The support relation does not stem from the attack relation (neither vice-versa), in the sense that if a supports b, then it does not follow that a attacks an attacker of b. For example, if the conclusion of a is a premise of b, then we could say that a supports b. In that case, it could be argued that stating the support relation aSb is not really helpful, because the content of the supporting argument a could be integrated with argument b. Nonetheless, it is true that supporting arguments are omnipresent in everyday argumentation (this can be verified, for example, at the on-line debate platform *Debategraph* [8]). Additionally, it is quite possible for an argument a to support various other arguments, therefore by considering argument a as a separate entity on an argumentation graph, we could minimize the repetition of information.

Bipolar Argumentation Frameworks without weights

Definition 12 *[ACLSL08, CLS05a] A* **Bipolar Argumentation Framework (BAF)** *is defined as a triple* $\langle A, R, S\rangle$ *where: A is a set of arguments, R is an attack relation over arguments, and S is a support relation over arguments. Its corresponding argumentation graph is called* **bipolar graph**.

[8] www.debategraph.org

In the literature, there exist various ways of defining acceptability semantics for argumentation systems featuring support. For example, the works in [BGvdTV10, CLS05c] analyze some possibilities. We will focus on a way for defining such acceptability semantics which is proposed in [ACLSL08] (and is based on the ideas of [CLS05b]) called *local gradual valuation*.[9]

Definition 13 *[ACLSL08, CLS05a] Let $\langle A, R, S\rangle$ be a bipolar argumentation framework. Also, let $\mathcal{V}$ denote a totally ordered set with a minimum element (V_{Min}) and a maximum element (V_{Max}). A* **local gradual valuation** *on $\langle A, R, S\rangle$ is a function $v : A \to \mathcal{V}$ such that $v(a) = g(h_S(v(c_1), \ldots, v(c_p)), h_R(v(b_1), \ldots, v(b_n)))$, with:*

- *Function h_R: $\mathcal{V}^* \to H_R$ valuating the quality of the attack on a.*
- *Function h_S: $\mathcal{V}^* \to H_S$ valuating the quality of the support on a.*
- *Function $g : H_S \times H_R \to \mathcal{V}$ with $g(x, y)$ increasing on x and decreasing on y.*

$\mathcal{V}^*$ denotes the set of the finite sequences of elements of $\mathcal{V}$, including the empty sequence.
H_S and H_R are ordered sets.

Also, the function h ($h = h_R$ or $h = h_S$) must respect the following constraints:

1. *If $x_i \geq x'_i$ then $h(x_1, \ldots, x_i, \ldots, x_n) \geq h(x_1, \ldots, x'_i, \ldots, x_n)$.*
2. *$h(x_1, \ldots, x_i, \ldots, x_n, x_{n+1}) \geq h(x_1, \ldots, x_i, \ldots, x_n)$.*
3. *$h() = \alpha \leq h(x_1, \ldots, x_i, \ldots, x_n)$ for all $x_1, \ldots, x_i, \ldots, x_n$.*[10]
4. *$h(x_1, \ldots, x_i, \ldots, x_n) \leq \beta$ for all $x_1, \ldots, x_i, \ldots, x_n$.*[11]

Some basic characteristics of this gradual valuation are the following:

1. There exist many possible instances of the above generic valuation.[12]
2. The valuation of an argument is a function of its direct attackers and of its direct supporters.
3. If the quantity of the supporting (resp. attacking) arguments increases, then the quality of the support (resp. attack) increases.
4. If the quality of the support (resp. attack) increases, then the valuation of the argument increases (resp. decreases).

Based on the above (local) gradual valuation of arguments, a simple way to define argument acceptability would be to accept all the arguments having a valuation greater than some threshold. For example, if the valuations fall into the interval $[-1, 1]$, then that threshold could be 0.

Bipolar Argumentation Frameworks with weights

A work using both weighted arguments and weighted relations (attacks and supports) is found in [BGW05]. The authors define an argumentation system where every argument has an numerical *weight*, representing its strength. Furthermore, every attack and support has a numerical *transmission factor*, which indicates its effectiveness. We note that arguments may attack (or support), not only arguments, but also attacks and supports. Given an argumentation system, the initial weight of every argument, and the initial transmission factor of every attack and support, is *updated*, on the basis of: (i) its initial value, (ii) the weights of its attackers and supporters, and (iii) the transmission factors of the attacks and supports towards it.

Finally, the authors in [ET12] define the notion of *Extended Social Abstract Argumentation Framework (ESAAF)* which considers votes on arguments.

[9] In [ACLSL08] an alternative choice is also proposed, the *global gradual valuation*.
[10] α denotes the minimum value of an attack (or support).
[11] β denotes the maximum value of an attack (or support).
[12] In Chapter 7 we will use a specific instance.

Definition 14 *[ET12] An* **Extended Social Abstract Argumentation Framework** *is a 4-tuple $\langle A, R, S, V\rangle$, where $\langle A, R, S\rangle$ is a (finite) BAF and $V : A \to \mathbb{N} \times \mathbb{N}$ is a function mapping arguments to the number of their positive and negative votes.*

Then, an argument's valuation in an ESAAF depends on its positive and negative votes, as well as on the attack and support relations in the system.

2.1.4 Other types of relations

Apart from argumentation systems with attack, support and preference relations, there have been defined systems featuring more complex, n-ary relations, where argument acceptability is defined differently. We just mention the Abstract Dialectical Frameworks [BW10], and the Constrained Argumentation Frameworks [CMDM06]. In these types of frameworks, expressions of the following type can be modelled: "Argument a is accepted, if and only if arguments b, c are accepted, or, if arguments d, e are rejected". In our work we will not consider these types of n-ary relations, but they may indeed be an interesting choice if we decide to model more complex debates.

2.2 Merging Multiple Viewpoints

When numerous agents engage in a debate, it is quite possible for them to consider different arguments, as well as different attacks between them. This section presents the methods proposed in the literature which aggregate, in meaningful ways, different viewpoints over arguments and attacks. It also provides some possible reasons explaining the disagreements over the attack relations.

First, we analyze the notion of attack and we explain how there can be room for disagreement over attacks between the agents. Second, we provide a number of methods appearing in the literature, which aggregate different private argumentation systems into a single, common argumentation system (or into a set of systems), taking into account all the agents' viewpoints.

2.2.1 Focusing on the notion of attack

Though our work focuses on abstract argumentation, we now provide some examples of structured arguments, in order to explain why, in real-life argumentation, agents may disagree over the validity of attacks between arguments. An example of work where an argument is defined as a minimal set of premises (called *support*) from which the conclusion can be inferred, is found in [AC02].

Definition 15 *[AC02] Given a propositional knowledge base Σ, which may be inconsistent, an argument of Σ is a pair (H, s) where $H \subseteq \Sigma$ such that:*

- *H is consistent.*
- *$H \vdash s$ (where $\vdash$ denotes classical inference)*
- *H is minimal (for set inclusion).*

H is called **the support** *and s is called* **the conclusion** *of the argument.*

Some pairs of atoms of Σ are contradictory. This notion of contradiction is the basis of the attack relation between arguments. Let $x = (H_x, s_x)$ and $y = (H_y, s_y)$ be two arguments. We say that x attacks y if and only if, either: (i) s_x contradicts a premise of H_y, or (ii) s_x contradicts the conclusion s_y.

2.2.2 How is disagreement over attacks possible?

Let us now analyze some important characteristics of the attack relation, which will help us understand why different agents may have different viewpoints over the attacks of an argumentation system.

An argument which contains some implicit premises is called **enthymeme** [Wal08]. Given that an enthymeme's set of premises is incomplete, its conclusion does not directly follow from them. As underlined in [Wal08, Hun07, Mod13] real-life arguments are very often enthymemes. Enthymemes are very useful, because by avoiding to repeat what might be common knowledge, agents gain time and their arguments become more compact and understandable. From another perspective, agents can use enthymemes in order to maneuver strategically in a dialogue, avoiding to disclose any information that may be used against them, or avoiding to reveal valuable information.

Let $y = (H_y, s_y)$ be an enthymeme, having an implicit premise. Agent ag_1 believes that the implicit premise is p_1, while agent ag_2 believes that the implicit premise is p_2 instead. Now, let $x = (H_x, s_x)$ be another argument whose conclusion contradicts p_1, but it does not contradict p_2. In this case, if both agents consider arguments x and y, then ag_1 believes that attack xRy holds, while ag_2 believes that xRy does not hold.

Another type of disagreement over an attack takes place when an argument's conclusion does not directly contradict another argument's premise or conclusion. Let again $x = (H_x, s_x)$ and $y = (H_y, s_y)$ be two arguments. We assume that the conclusion s_x does not directly contradict neither a premise in H_y, nor the conclusion s_y. But, s_x together with some additional information would have actually contradicted either a premise in H_y, or the conclusion s_y. As a result, an agent who can (resp. cannot) identify this missing information will consider that the attack xRy holds (resp. does not hold).

The two above types of disagreement over attacks were based on the notion of enthymeme. This is not the only way to obtain contradictory viewpoints over an attack.

We have already seen that in Preference-based Argumentation Frameworks [Mod09, ADM08], when argument a attacks argument b (aRb), then depending on whether $Pref(a, b)$ or $Pref(b, a)$, the defeat relation aDb holds or not. Therefore, if two agents have different opinions on the preference of a over b, they may have different opinions on the defeat relation between these arguments. We may have similar situations when using Value-based Argumentation Frameworks [BC03], or Weighted Argumentation Systems [DHM$^+$11].

2.2.3 Merging argumentation systems

Let there be a set of agents, each one having his own private argumentation system. The agents may not only consider different sets of arguments but, as we just saw, they may also disagree on the attack relations between them. In such a context, a natural question is the following: How can we compute the aggregation of these systems, in the form of one or more (merged) argumentation systems? Notice that this type of aggregation assumes that an entity knows all the information appearing in the agents' private systems, and proceeds to their aggregation. As we will see later, this is quite different from the aggregation taking place during a debate. In the latter case the aggregation is done progressively, as long as agents keep contributing to the debate.

Let us present some methods in the literature for merging argumentation systems. Initially, in [CMDK$^+$07] the authors remark that instead of merging the agents' sets of acceptable arguments (the final step of their reasoning process), it makes more sense to merge their viewpoints on the arguments and on the attacks between them (the initial step of their reasoning process). They also take into consideration the fact that agents may have different sets of arguments in their private systems. In order to compute meaningful aggregations in these cases, they introduce the notion of *Partial Argumentation Framework*. Given two arguments x and y, and a binary attack relation R, a Partial Argumentation Framework makes the distinction between the three following cases:

1. Certainty that xRy holds.

2. Certainty that xRy does not hold.

3. Uncertainty whether xRy holds or not.

Definition 16 *[CMDK$^+$07] A (finite)* **Partial Argumentation Framework** *over A is a quadruple $\langle A, R, I, N\rangle$ where:*

- *A is a finite set of arguments,*
- *R, I, N are binary relations over A:*
  - *R is the* **attack relation**.
  - *I is called the* **ignorance relation** *and is such that $R \cap I = \{\}$.*
  - *$N = (A \times A) \setminus (R \cup I)$ is called the* **non-attack relation**.

Let $AS_1, \ldots, AS_n$ be some argumentation systems. Then, $P(AS_1, \ldots, AS_n)$ is called a *profile*. The **expansion** of system AS_i, with $i \in [1, \ldots, n]$, with respect to $P(AS_1, \ldots, AS_n)$, is a Partial Argumentation Framework denoted $PrtAF_i$, which preserves all the arguments and attacks of AS_i, and "expands" it by taking into consideration all the other systems in $P(AS_1, \ldots, AS_n)$. Many types of expansions can be defined, and the authors mainly focus on the *consensual expansion*. The consensual expansion has two main characteristics: (1) the set of arguments of every $PrtAF_i$ is the union of all the systems' arguments and (2) an attack (or non-attack) relation is added into $PrtAF_i$ if all the systems which contain the corresponding pair of arguments agree on that relation.

Given a profile $P(AS_1, \ldots, AS_n)$ a merging process consists of two steps:

1. **Expansion**: from every AS_i, its corresponding Partial Argumentation Framework $PrtAF_i$ is computed.

2. **Fusion**: From the set of Partial Argumentation Frameworks computed in the expansion step, a set of abstract systems are computed, such that they all have minimum distances from the set of Partial Argumentation Frameworks (different definitions of distance can be used here).

Summing-up, the approach in [CMDK$^+$07] relies on an "expansion" and a "fusion" step. There exist several ways to expand an AS, as well as to fusion a set of PAFs. Finally, the result of the merging process may be a set of abstract systems (not necessarily a single abstract system).

The work in [TBS08] focuses on a simpler setting, as all the argumentation systems share the same arguments. The agents may have different opinions on the existence of attacks, but there is no ignorance relation over arguments, as opposed to [CMDK$^+$07]. The goal is to aggregate all the different systems, into a single system. The authors study different aggregation methods, which are based on **voting**, and more specifically *majority voting* and *qualified majority voting* (where some agents can veto the inclusion of an attack). Then, the authors check whether the proposed voting methods satisfy desired properties.

The works in [GR12, CLS11] propose a different way to merge abstract argumentation systems. The difference is that the merged system is a unique *weighted* argumentation system. Its computation is done more quickly than in [CMDK$^+$07], as there are no separate expansion and fusion steps. Also, having a unique merged system may be useful for practical reasons. In [GR12, CLS11] the agents have, in the general case, different sets of arguments and different sets of attacks. The merged system contains the union of all the arguments, and the union of all the attacks of the agents. Let us now see the differences between these works.

Firstly, in [CLS11] each agent votes on each pair of arguments in his system: he casts either a positive vote (if he has the attack in his system), or a negative vote (if he does not have it). A formula aggregates these votes, computing a numerical value in the interval $[0, 1]$, so the merged

system is a WAS with weighted attacks. Additionally, the authors propose a way to extend this approach in bipolar argumentation frameworks, considering that supports can compensate for attacks, and proposing a way to compute values of relations in the interval $[-1, 1]$ (where negative values indicate attack and positive values indicate support).

Secondly, in [GR12], the authors propose another way to merge a profile of argumentation systems into a single WAS. As we said above, the agents do not share the same arguments, or attacks. Therefore, given a component of the framework (an argument or an attack) an agent can have the following opinions about it: (i) positive (he has the argument in his system and it is accepted; or he has the attack in his system), or (ii) neutral (he does not have the argument in his system; or he does not have both the attack's arguments), or finally (iii) negative (he has the argument in his system and it is rejected, or he has both the attack's arguments, but not the attack). The authors propose two types of aggregation of the agents' opinions, such that a component's weight is found in the interval $[0, 1]$: The *credulous* aggregation (taking into account only the agents who are positive or negative towards the component) and the *skeptical* aggregation (taking into account all the agents).

The works in [RT10, CP11] also propose ways to merge different viewpoints. In contrast to the previously described works, these ones assume that all the agents share the same argumentation system. Therefore, there is no disagreement, neither on arguments, nor on attacks. The agents use argument labelling in order to decide on the arguments' acceptability, and this is the source of disagreement. We remind that a reinstatement labelling of a system is not unique, so the labellings of the agents can differ. In general, the authors propose ways to compute a **collective labelling**, given a profile of labellings (with *in*, *out* and *undec* being the labels' possible values).

Firstly, in [RT10] the authors propose a number of desired properties for a labelling aggregation operator (for example unanimity is satisfied when, for every argument, its collective label is identical to the label it has in all the agents' systems). They also focus on labelling aggregation operators which are strategy-proof (no agent has incentive to declare a different labelling than his own). Also, they provide some impossibility results: they identify sets of properties which cannot be satisfied together by a labelling aggregation operator.

Secondly, in [CP11] the authors focus on the issue of collective labelling, and they define three labelling aggregation operators, which are based on conflict-free, admissible and complete labellings [Cam06].

The first is the *skeptical aggregation operator* which computes a labelling "in accordance" with every agent's labelling, in the sense that every *in* and *out* argument of the merged system has the same label in all the agents' systems. The second is the *credulous aggregation operator* which computes a labelling "not conflicting" with any agent's labelling, in the sense that for every *in* and *out* argument of the merged system, there is no agent's system where that argument has the opposite label. Finally, the *super-credulous operator* expands the labelling computed by the credulous operator.

Also, in [CPP11] the authors study a couple of operators which aggregate different labellings into a collective labelling. They check whether the agents can manipulate the procedure by lying, and whether the results of the aggregation are (Pareto) optimal.

Finally, the works in [LM11, EML13] do not study the problem of merging argumentation systems, but of merging opinions (expressed in the form of positive and negative votes) on specific arguments and attacks of an argumentation system. The difference, compared to the previous works, is than the agents can choose the arguments and attacks on which they will give their opinions. The result of the aggregation is a system with weighted arguments in [LM11], and with weighted arguments and attacks in [EML13].

Chapter 3

Expertise in Argumentative Debates[1]

As explained in Chapter 1, this work aims to serve as a contribution for the enhancement of real-life argumentative debates. In this chapter our goal is to define the basic elements of a framework where a number of agents, with expertise in different domains, take part in an argumentative debate. We will not provide any specific debate protocols yet. This will be the subject of Chapter 7. From the rich literature on argumentation frameworks, an overview of which has been presented in Chapter 2, we carefully choose the elements we need in order to successfully model a debate among experts. The main questions that must be addressed are the following:

1. How are the debate's participants modelled, and especially their expertise?
2. How is the debate modelled, and how can the experts influence it?
3. How are the debate's conclusions obtained?

In the following we propose some answers to the above questions.

3.1 Modelling the Debate's Participants

Our framework for aggregating different experts' opinions which is presented in this chapter, will be later used in the context of multilateral debates. These debates usually focus on a specific "issue", which can be a statement or, as in our work, an argument.[2]

We assume that every agent, based on his own knowledge, preferences and reasoning capabilities, is able to construct a private (Dung) abstract argumentation system. An agent's private system obligatorily includes the argument-issue, otherwise the agent would not be motivated to take part in the debate. It also includes all the arguments which the agent finds relevant to the issue. Another assumption we make is that every agent, for every pair of arguments (a, b) appearing in his system, has a clear opinion on the relation which holds between them: he either believes that the attack aRb holds, or that it does not hold.[3] So, as far as the agents' systems are concerned, there is no binary relation representing indecisiveness (as opposed to the work in [CMDK$^+$07]), and the attacks are not weighted.

Based on his private system, an agent is able to make his mind about the acceptability status of the issue. We assume that all the agents are using the same method to compute argument acceptability, which can be either extension-based, labelling-based, or based on numerical argument

[1] Our publication relevant to this chapter is: [KBMM12].

[2] Notice that a position may sometimes be the conclusion of several arguments. In that case the debate must focus on all these arguments, not on a single one. In this work we always focus on one argument, but the ideas, protocols and algorithms we present can be amended, in order to focus on a set of arguments. We believe that this is a non-trivial task which merits further study.

[3] In Chapter 7 we will also consider support relations.

valuation. Usually, as for example in [BM11], the acceptability status of the issue in such contexts is two-valued (either accepted, or rejected), for simplicity reasons. In these cases the debating agents are typically partitioned into two sets, PRO (containing the agents who accept the issue) and CON[4] (containing the agents who reject the issue). During the debate, every agent tries to convince the others that his opinion on the issue's status is correct. Later, we shall see how exactly the agents can pursue their goals in a debate.

We now introduce some needed notation. Let N denote the set of agents taking part in a debate. Every agent $i \in N$ is equipped with a private (Dung) abstract argumentation system, denoted $AS_i = \langle A_i, R_i \rangle$. Let x denote the issue of the debate. Then $\forall i \in N$ it holds that $x \in A_i$ because, as we said before, the issue is found in all the agents' private systems. A priori, not all the agents consider the same sets of arguments, so for agents $i, j \in N$ with $i \neq j$, in the general case $A_i \neq A_j$. That said, in many cases we shall make the simplifying assumption that the agents do share the same sets of arguments. Whenever we make this assumption, it will be explicitly stated. As far as the attack relation is concerned, in Chapter 2 we have presented different reasons which can cause disagreement over the validity of an attack. Disagreement over attacks is a key point of our work and we consider that for agents $i, j \in N$ with $i \neq j$, in the general case $R_i \neq R_j$.

3.1.1 Defining Expertise

A common feature of such debates is that their corresponding argumentation graphs can expand quickly, as new arguments are introduced by different agents. This usually leads to a shift of the focus of the discussion into points which are farther away from the original issue. As a result, while the issue might be related to a topic (e.g. it might be an argument related to Economics), other arguments of the graph may be related to various other topics (e.g. International Relations, Law, Ethics, Ecology). Moreover, given that different agents have expertise in different topics, it seems natural to give more weight to an agent's opinion whenever that agent is expert in the topics he talks about. For example, the opinion of a judge on the nature of a new law, should have a bigger impact on the debate than the opinion of an economist. The inverse should happen, when they provide their opinions on possible effects that law may have in the economy.

Another important practical issue is whether the different experts can be trusted in an argumentative dialogue [PTS+11]. In our work we will not elaborate on the issue of trust. We make the simplifying assumption that the experts are trustworthy and we just focus on their expertise.

We assume that agents only use abstract arguments, whose structure is left unspecified, and there is no semantic analysis of their contents. Nevertheless, each argument is *tagged* with keywords specifying which topics the argument is related to. An example of this approach is found in [TT11].

In a debate, we assume that there exists a set of potential *topics*, denoted T, which is known and fixed a priori. Every argument a that an agent may put forward, is related to a subset of these topics, which can be empty or contain as many topics as wished.

Definition 17 *Let T be the set of topics. The set of* **topics of an argument** *$a \in A$ is given by function $top(a) \subseteq T$.*

Example 7 *Consider a debate system designed to support discussion among PC members about papers to accept or reject for a conference. There is a list of keywords that PC members can choose to indicate their area of expertise,* e.g. *T = {comp, kr, ml, cog}, where comp stands for "complexity", kr stands for "knowledge representation", ml stands for "machine learning", and cog stands for "cognitive science". A first reviewer ($PC1$) argues that the paper is good because it presents an interesting representation formalism, very elegant, and very much plausible from the cognitive point of view (argument* a*, with $top(a) = \{kr, cog\}$). A second reviewer ($PC2$) challenges this on the basis that the formalism is too expressive, so the reasoning tasks would be intractable (argument* b*, with $top(b) = \{comp\}$), and that the formalism contains some imperfections that*

[4] In the literature, instead of CON, this set is often denoted OPP.

should be worked on before publication (argument d*, with* $top(d) = \{kr\}$*). A third reviewer (*$PC3$*) challenges argument* b *by saying that he is aware of related formalisms and problems in machine learning for which very good approximation algorithms work in practice, so it is not unlikely that the same could happen with this one (argument* c*, with* $top(c) = \{comp, ml\}$*).*

Figure 3.1: A preliminary attempt to model the debate among the experts $PC1$, $PC2$ and $PC3$. The arguments of the above system are attached to various topics, from a set of topics $T = \{comp, kr, ml, cog\}$.

Next, from topics attached to arguments, we deduce how topics are attached to attacks. We make the simple assumption that an attack between two arguments obligatorily refers to the topics of both arguments.

Definition 18 *Let* T *be the set of topics, and let* R *be a set of attacks. The set of* **topics of attack** $(a, b) \in R$ *is given by the function*[5] $top(a, b) = top(a) \uplus top(b) \subseteq T \uplus T$.

As attacks are binary, this is simply a multiset where topics appearing in the attacking and attacked arguments appear twice. For a potential attack, it is thus possible to distinguish three levels of "relevance" for topics:

- *prominent* topics (attached to both arguments) denoted $prom(a, b) \subseteq T$,
- *relevant* topics (attached to one argument) denoted $rel(a, b) \subseteq T$, and
- *irrelevant* topics (not attached to either argument) denoted $irr(a, b) \subseteq T$.

Example 7 (cont.) *We have* $top(c, b) = \{comp, comp, ml\}$*, thus* $prom(c, b) = \{comp\}$*,* $rel(c, b) = \{ml\}$*, and* $irr(c, b) = \{cog, kr\}$*. We also have* $top(b, a) = \{comp, kr, cog\}$*, so* $prom(b, a) = \{\}$*,* $rel(b, a) = \{comp, kr, cog\}$*, and* $irr(b, a) = \{ml\}$*.*

Let us now model the agents' expertise. We simply assume that every agent $i \in N$ is expert in a subset of the topics of T. A practical question would be how the agents' expertise is known and verified in a debate. We will not elaborate on this issue, we simply assume that a central authority of the debate asks the participants to declare (and prove) their expertise in a subset of topics of T, before the beginning of the debate.

Definition 19 *The* **expertise of agent** $i \in Ag$ *is given by a function* $exp(i) \subseteq T$.

Having presented how the agents' personal beliefs and expertise are modelled, we now propose a way to model the multi-agent debate.

3.2 Modeling an Argumentative Debate among Experts

We shall now present the central structure which forms the backbone of a debate. This structure is an argumentation system, called *Gameboard*. The use of a central structure is standard way to represent dialogues and debates, as seen in [BM11, Pra05, Pra06]. The innovative part of our work consists in defining how the agents may influence the Gameboard, according to their expertise, as well as in studying its dynamics (in Section 5.5).

[5] $\uplus$ indicates the multiset union

3.2.1 The Gameboard

The agents debate with the help of a type of *blackboard* [Nii86] which is called *Gameboard*, denoted GB. The GB is a type of Weighted Argumentation System, "visible" to all the agents during the whole course of the debate. Its role is to aggregate the agents' opinions and represent the current *state of the debate*. Every time that an agent asserts an opinion, the GB is updated. As the debate proceeds in timesteps, we denote by GB^t the Gameboard at timestep t. It can be argued that every rational agent who looks at the debate and has no predefined personal opinions, would draw conclusions based solely on the information available on the GB. Let us now see what types of actions the agents can perform during the debate:

- **Introduce arguments and attacks on the** GB:
  During the debate, the agents are allowed to introduce arguments on the GB and attacks between arguments which are already there. This is of course a standard type of action in argumentative dialogues, as seen for example in [BM11, Pra05, Pra06]. We note that similarly to [BM11], and in contrast for example to [Pra05] (where dialogues give rise to trees of arguments), the insertion of arguments and attacks on the GB may lead to the construction of cycles. Thus the obtained argumentation graph is not a tree, in the general case.

- **Vote on attacks of the** GB:
  We have insisted on the fact that agents may have conflicting opinions on the attacks between arguments. Disagreement over attacks of the GB is addressed by letting the agents *vote* on them. We adopt a quite simple approach: every agent is able to vote on an attack, by casting either a positive, or a negative vote. In the context of our work, a positive (resp. negative) vote by agent $i \in N$ on attack (a, b) has a specific meaning: it indicates that agent i believes that argument a attacks (resp. does not attack) argument b.

The first point above (introduction of arguments and attacks on the GB) is quite straightforward. For the moment, let us focus on the second point. Experts may express their opinions on attacks of the GB, by casting positive or negative votes on them.

Definition 20 *Let R be a set of attacks. A* **vote** *is a tuple $\langle (a, b), s, i \rangle$ where $(a, b) \in R$ is the attack concerned by the vote, $s \in \{-1, +1\}$ is the polarity (sign) of the vote, and $i \in N$ is the voter.*

When an agent expresses his opinion for or against an attack (in our context by voting), the impact of that vote depends on his expertise over the topics of this attack. Intuitively, the opinion of an expert in the topics of the voted attack should have more importance, than the opinion of a non-expert in those topics. However, this general principle needs to be made much more precise, as illustrated in the next example.

Example 7 (cont.) *Let us assume that reviewers have to choose two keywords (topics of expertise) from a list. PC1 has expertise in $\{kr, cog\}$, PC2 has expertise in $\{kr, comp\}$, and PC3 has expertise in $\{comp, ml\}$. We focus on attack (c, b). As PC3 puts forward argument* c *and attack (c, b), he voted by default positively on the attack. PC1 voted against this attack because he thinks it is not valid, while PC2 voted in favour of it. As reviewers have different topics of expertise, we can summarize their votes, as well as their expertise, by means of vectors (where prominent topics are grouped together, and the same holds for relevant and for irrelevant topics) as follows:*

$$
\begin{array}{lllll}
\text{Vote of PC3 on (c,b) :} & \langle\langle comp : +1\rangle, & \langle ml : +1\rangle, & \langle kr : 0, cog : 0\rangle\rangle \\
\text{Vote of PC2 on (c,b) :} & \langle\langle comp : +1\rangle, & \langle ml : 0\rangle, & \langle kr : +1, cog : 0\rangle\rangle \\
\text{Vote of PC1 on (c,b) :} & \langle\langle comp : 0\rangle, & \langle ml : 0\rangle, & \langle kr : -1, cog : -1\rangle\rangle
\end{array}
$$

But how can we aggregate different votes and obtain a meaningful result? The question is difficult because, in order to aggregate them, we must consider two things: first, that a single vote concerns several topics, and second that various votes must be aggregated. There are some key assumptions that need to be made explicit:

1. **Independence of expertise**:
   Suppose an expert in *kr* votes for (b, a), then another expert in *comp* votes the same way. Does this have the same impact as one expert in both topics voting once for (b, a)?

2. **Compensation among topics**:
   Should we allow compensation among levels of topics? For instance, should two votes on a relevant topic be as important as one vote on a prominent one? Should irrelevant topics be considered in the first place?

There are many ways to aggregate the votes. Some interesting ideas can be found in [DGPS01]. Here we make the following choices: we suppose that independence of expertise holds; we allow compensation among topics (considering prominent topics to be twice as important as relevant topics); and finally we disregard votes on irrelevant topics. Initially, we define the notion of *impact* of a vote, which indicates how important the vote is, compared to others. Every vote's impact has a numerical value, in order to facilitate the aforementioned comparisons. Our definition of impact has two advantages. First, the more topics of an attack an agent is expert in, the greater his impact is on the attack. Second, expertise in the prominent topics of an attack leads to a greater impact than expertise in its relevant topics.

Definition 21 *Let R be a set of attacks and let $i \in N$ be an agent. The* **impact of** i **on** $(a, b) \in R$ *is denoted $imp_i(a, b)$ and defined by $imp_i(a, b) = 2 \times |exp(i) \cap prom(a, b)| + |exp(i) \cap rel(a, b)|$. If $imp_i(a, b) = 0$ we say that i is a* **dummy voter** *on (a, b).*

Example 7 (cont.) *Since $exp(PC2) = \{kr, comp\}$ and $top(c, b) = \{comp, comp, ml\}$, it holds that $imp_{PC2}(c, b) = 2 \times 1 + 0 = 2$, as $PC2$ is expert in one prominent topic of the attack (and in no relevant topics).*

Based on the above definition of impact, we now define a method to aggregate the experts' votes. As we already said, given all the votes cast on an attack (and their impacts), there exist various ways to aggregate them. In this work we want the aggregation to capture two things:

- Whether the collective opinion on (a, b) is positive or negative, and *how* positive or negative it is. This will be captured by a single numerical value, denoted $w(a, b)$.
- Whether the collective opinion is based on a sufficient number of votes.

This information on an attack is captured in its *evaluation vector*.

Definition 22 *Let R be a set of attacks. The* **evaluation vector of** $(a, b) \in R$ *is denoted $\vec{v}(a, b) = \langle w(a, b), mw(a, b)\rangle$, where $w(a, b) \in \mathbb{Z}$ is called the* **weight** *of (a, b)), and $mw(a, b) \in \mathbb{N}$ is called the* **max-weight** *of (a, b).*

The **weight** of (a, b) is the aggregated impact of all the votes cast on (a, b), and the **max-weight** of (a, b) is the aggregated impact of all the votes cast on (a, b), with the difference that all votes with negative polarities are considered to have positive polarities instead.

Let us define how an evaluation vector is computed. Let R denote a set of attacks. At the beginning, before any votes on attacks are cast, it holds that $\forall (a, b) \in R$, $\vec{v}(a, b) = \langle 0, 0\rangle$. We now iteratively define how the evaluation vector $\vec{v}(a, b) = \langle w(a, b), mw(a, b)\rangle$ is updated after expert $i \in N$ casts a vote $\langle (c, d), s, i\rangle$.

$$upd(\vec{v}(a, b), \langle (c, d), s, i\rangle) = \begin{cases} \vec{v}(a, b), \text{ if } (a, b) \neq (c, d) \text{ or } imp_i(a, b) = 0 \\ \langle w(a, b) + (s \times imp_i(a, b)), mw(a, b) + |top(a, b)|\rangle \text{ otherwise} \end{cases}$$

Notice that a vote on attack (a, b) can only change the evaluation vector of this specific attack, and only if the voter has some relevant expertise in its topics. The weight $w(a, b)$ aggregates the positive and negative votes on (a, b) by using a simple sum. Therefore positive and negative votes can be balanced. Of course, there exist more elaborate choices for the aggregation, but we preferred to promote simplicity. Finally, the value of $mw(a, b)$ is always equal to the product of the number of (non-dummy) voters on (a, b) and the cardinality of $top(a, b)$.

Example 7 (cont.) *The evaluation vector of (c, b) after the vote of PC3 is $\vec{v}(c, b) = \langle 3, 3\rangle$. It becomes $\vec{v}(c, b) = \langle 3 + 2, 3 + 3\rangle = \langle 5, 6\rangle$ after the vote of PC2, and remains unchanged after the vote of PC1 who has no expertise in the topics of this attack (and thus is a dummy voter on (c, b)). Now, we assume that PC2 votes for the attack (b, a), whereas PC1 and PC3 vote against it; and that PC2 is the only expert who expresses his opinion on (d, a). We obtain $\vec{v}(b, a) = \langle -1, 9\rangle$, $\vec{v}(c, b) = \langle 5, 6\rangle$ and $\vec{v}(d, a) = \langle 2, 3\rangle$. The dotted arrow denotes an attack with negative weight.*

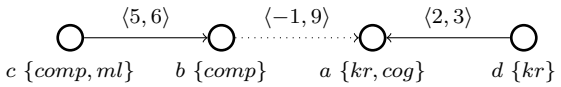


Figure 3.2: The attacks (b, a), (c, b) and (d, a) in the above system have their evaluation vectors attached to them. Attacks with positive (resp. non-positive) values of $w(a, b)$ are drawn with normal (resp. dotted) arrows.

Before closing this section, we can finally formally define the notion of Gameboard. The GB, at timestep t, is denoted GB^t and it is a Weighted Argumentation System which contains all the arguments put forward by the agents until t (denoted A^t_{GB}), the attacks defined by the Cartesian product $R^t_{GB} = A^t_{GB} \times A^t_{GB}$, and an evaluation vector $\vec{v}(a, b)$ for every attack $(a, b) \in R_{GB}$.

Definition 23 *The* **Gameboard**, *at timestep $t \in \mathbb{N}$, is denoted by $GB^t = \langle A^t_{GB}, R^t_{GB}, Eval^t\rangle$ where: A^t_{GB} is the set of arguments which have been introduced in the debate until t, $R^t_{GB} = A^t_{GB} \times A^t_{GB}$ is the set of attacks between arguments in A^t_{GB}, and finally $Eval^t = \{\vec{v}(a, b) \mid (a, b) \in R^t_{GB}\}$ is the set of the evaluation vectors of the attacks in R^t_{GB}.*

When we refer to no specific timestep, or when the timestep we refer to is evident from the context, we will drop the superscripts t from the notation.

Having explained how the experts' opinions are aggregated on the Gameboard, the next issue which must be addressed is how conclusions can be drawn from it.

3.3 Obtaining the Debate's Conclusions

Once the Gameboard has taken its final form, in the sense that no agent wishes to introduce a new argument or vote on an attack, there exist many ways in which the debate's conclusions can be drawn.

We remind that we are interested in the acceptability status of a single argument, which is the debate's issue, and that the Gameboard is a type of Weighted Argumentation System with weighted attacks. In Chapter 2 we have presented several propositions in the literature for computing the acceptability status of an argument in a WAS. Our choice in this work is motivated by the following considerations:

1. An agent's positive (resp. negative) vote on $(a, b) \in R_{GB}$ reflects his opinion that a attacks (resp. does not attack) b. Therefore, if we follow the agents, it seems reasonable to separate

the attacks of the GB into two sets: those which are collectively considered "valid" (to hold) and those which are collectively considered "invalid" (not to hold).

2. Real-life debates should have a simple and understandable way to draw conclusions. For example, the use of grounded semantics ensures that a unique grounded extension will always be computed.[6]

Taking the above into account, we made the following choice: if an attack's weight on the GB is strictly positive (resp. negative or zero), then this attack is considered collectively valid (resp. not valid). Alternatively, another (relatively small) threshold value could be used, instead of zero. This motivates us to define, given a GB, another argumentation system, called *non-weighted counterpart* of GB.

Definition 24 *Let $GB = \langle A_{GB}, R_{GB}, Eval\rangle$ be a Gameboard.*[7] *The* **non-weighted counterpart argumentation system** *of GB is an abstract argumentation system defined as follows: $GB_{cp} = \langle A_{cp}, R_{cp}\rangle$, with $A_{cp} = A_{GB}$ and $R_{cp} = \{(a,b) \mid w(a,b) > 0\}$.*

Based on the counterpart system GB_{cp}, we can use many methods to decide on the acceptability status of the arguments. In order to promote simplicity, we mainly use the grounded semantics.

Example 7 (cont.) *Below we see the GB and its counterpart system GB_{cp}.*

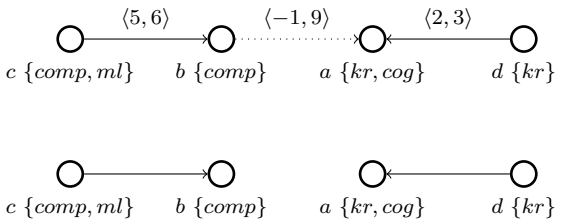


Figure 3.3: The GB (above) and its counterpart system GB_{cp} (below). The latter contains two attacks, the ones whose weights were strictly positive.

The grounded extension of GB_{cp} is $\{c, d\}$. If argument a ("the paper is good because it presents an interesting representation formalism, very elegant, and very much plausible from the cognitive point of view") is the issue of the debate, then it is rejected.

Let us mention that our choice to compute argument acceptability by using the counterpart system, does not mean that for an attack (a, b), the absolute value of $w(a, b)$ and the value of $mw(a, b)$ are useless. Later, in Section 5.5, they will be used to estimate how confident we can be on the conclusions of a debate.

In this chapter we have presented a method to aggregate opinions coming from a set of agents with different types of expertise, and with different beliefs about argument relations. This aggregation takes place step-by-step, as long as the experts keep exchanging opinions about the validity of argument relations. Given that a decision-maker should be able to draw specific conclusions from such an exchange, we have proposed a way to construct (and to gradually amend) a single argumentation system (the GB) which captures the actual state of the discussion between the experts. Then, from the GB, we have proposed a specific method to obtain conclusions regarding argument acceptability.

Given the fluid nature of the GB (any new expert opinion may influence it), in Chapter 5 we will focus on modeling potential changes of the GB, on computing their repercussions on argument

[6] An additional advantage of the grounded extension is that it is not computationally expensive.
[7] We do not consider a specific timestep t.

acceptability, and on evaluating how confident we should be on our actual conclusions. Subsequently, in Chapter 7, we will focus on how to coordinate agents who engage in such discussions, and on defining and evaluating agent debate strategies.

Part II

Abstract Argumentation Dynamics

Chapter 4

Background on Abstract Argumentation Dynamics

The term *argumentation dynamics* refers to the ways that the components of an argumentation system (its arguments and relations) influence each other, a crucial element for the analysis of structural change of an argumentation system. We have seen that the different semantics for argument acceptability are defined on the basis of how arguments are connected on the argumentation graph, therefore the study of argumentation dynamics can help us identify properties of the existing semantics, classify them on the basis of these properties, and define new semantics which satisfy desired properties. An example is the work in [BG09] which provides a good overview of the different semantics of abstract argumentation, but also elaborates on their dynamics. For instance, it defines the intuitive *directionality principle* which requires that in a system $AS = \{A, R\}$ an unattacked subset of arguments $U \subseteq A$ is unaffected (as far as belonging to extensions is concerned) by the remaining part of AS.

Another important aspect of studying argumentation dynamics is its usefulness when *change* in an argumentation system is considered. If we know how the elements of a system influence each other, then it is easier to analyze an evolving system, and to predict its behaviour.

Practically, there are various reasons which can lead to the change of an argumentation system. Firstly, in a single-agent setting, an argumentation system can represent the knowledge an agent has over a specific issue. Usually, agents are placed in an ever-changing environment, where they can come across new information, something that may make them update or revise their previous information. Sometimes the agent's arguments have specific content and structure, and they are constructed from an underlying logical language L. In these cases, an agent may have a knowledge base consisting of a set of propositions in that language. If that agent adds, deletes or transforms some propositions in his knowledge base, this can lead to the construction of new arguments (which could not initially be constructed), and to the deletion of other arguments (if propositions are deleted from his knowledge base). Naturally, these changes may influence the acceptability status of many arguments the agent is able to construct. But also, revision may take place in a setting where arguments have no specific structure. Imagine for example an abstract argument representing an argument in natural language that an agent hears, and decides to add it into his argumentation system. In this work, we abstract from the structure and the content of arguments, and we just focus on dynamics of abstract argumentation systems. On an abstract argumentation system, since its only components are arguments and binary attacks over arguments, change can consist in one of the following things, or in a combination of them:

- Addition of an argument, or set of arguments.
- Removal of an argument, or set of arguments.
- Addition of an attack, or set of attacks.

- Removal of an attack, or set of attacks.

Secondly, in a multi-agent setting, change is even more frequent. Imagine for example the case of a multilateral dialogue. The agents who are not satisfied with the current state of the dialogue, try to put forward new arguments and attacks which will alter the dialogue's outcome in their favour. So, if the debate is represented in the form of a single argumentation system, evolving during the debate, then the agents who understand argumentation dynamics, have an important edge in enforcing their desired result.

We will focus on the dynamics of systems using extension-based and labelling-based acceptability. Initially, we will present a couple of works which study how an argumentation system's elements influence each other, and which do not consider structural change of systems. The work of Modgil and Caminada in [MC09] presents ways in which the following question can be addressed: Given an abstract system and an argument in it (let us call it *issue*), how can we find its acceptability status without computing all the extensions? In order to do this, we must understand the dynamics of the system: which are the arguments that influence the issue's status, and in what way.

Then, as far as the study of the (structural) change of an abstract argumentation system is concerned, we propose the separation of questions in two main types:

1. **For a given change of a system, which may its effects be?**
   Here general types of change are studied. Let us give two examples. Which are the possible effects of a single argument's addition or removal? Which are the possible effects of deleting a single attack (a, b), when a is labelled *out*?

2. **For given effects on a system, which changes may bring them about?**
   Compared to the previous type of questions, this is its inverse. This type of questions is usually referred to as the *enforcing problem*. We have an initial system AS, and a goal which we strive to achieve. The goal can be, for example, the acceptance of a specific argument, or the existence of a specific extension (under given semantics). Then, we search for all the possible changes on the initial system AS which lead to the satisfaction of the goal. An even more specific problem is the computation of the **minimal change** on AS achieving the goal, or at least the computation of some upper and lower *bounds* of the change which is needed (e.g. the maximum/minimum number of structural changes). The enforcing problem is particularly interesting in multilateral debates, where the debating agents must know how to persuade the others on the acceptability of their positions. Interestingly, it can be viewed from another, abductive, viewpoint: starting from a system AS, an agent may reason abductively and begin from the fact that argument x is accepted. Then, the agent searches for all the possible explanations of x's acceptance (modified "versions" of the initial system AS).

Let us also say a word on system *equivalence*. Two systems AS_1 and AS_2 are called *equivalent* if and only if they have the same extensions, under given semantics. There are more specific types of equivalence, for example AS and AS' are called *strongly equivalent* in [OW11] if and only if they have the same extensions (under given semantics), and also for any (identical) change on AS and AS', they will still have the same extensions. We can ask questions concerning equivalence, which may fall into any of the two above types, for example: can a particular type of change on AS lead to an equivalent system? What are the possible ways to obtain a system equivalent to AS?

Let us now present some works from the literature on the dynamics of abstract argumentation. We do not aim at providing a comprehensive overview of this literature, instead we focus on the works which have provided a basis for our research, as well as on works whose objectives were similar to ours.

4.1 Proof Theories as Argument Games

We start with the work of Modgil and Caminada in [MC09]. A question they address is how to find an argument's status without computing all the possible extensions of the system it belongs to. [1] This may be helpful in practice, because computing all the extensions under given semantics may be computationally expensive, when the system contains many arguments and attacks. The authors suggest that *proof theories* can be used for this task. These proof theories have the form of argument games between two players. Different games are defined for different Dung semantics. Given an argumentation graph, a procedure (proof theory) focuses locally on the *issue*, on its attackers, and then, recursively, on their attackers.

Let us begin by a brief overview of the main argument games defined in [MC09]. Given an abstract argumentation system $AS = \{A, R\}$, and an argument $a \in A$, the authors ask the following questions:

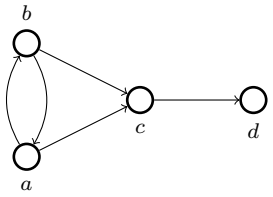
- Does a belong to the grounded extension of AS?
- Does a belong to some preferred extension of AS?

Though we are situated in a single-agent context, the proposed argument games have a dialogical nature. Each argument game takes place between two imaginary opponents who exchange arguments: a PRO player who supports the issue, and a CON player who objects to it. During the game, the two opponents put forward arguments and attacks which appear on AS and, in the process, they collectively construct a tree of arguments. Initially, that tree only contains a root, which is the issue. At every step of the game, an argument is *in* if it has no attackers, or if all its attackers are *out*; an argument is *out* if it has at least one *in* attacker. The players take turns, and add arguments and attacks to the tree: after every move, the label of the root must change. These games, which are of course no real dialogues, have been proved to compute correctly the acceptability status of arguments, under different semantics, as we will now see.

Two games are defined for the two previous questions: The first is the *grounded game* and the second is the *preferred game*. At the end of the grounded (resp. preferred) game, if the issue's status is *in* on the constructed tree, then the conclusion is that the issue belongs to the grounded (resp. to at least one preferred) extension.[2] Otherwise, the opposite holds. The two games have slightly different rules. For example, in the grounded game PRO cannot repeat an argument, something which he can do in the preferred game. Also, in the preferred game the arguments introduced by PRO should not attack each other (this check is redundant in the grounded game).

Let us provide an example of the grounded and preferred games:

Example 8 *Let $AS = \langle A, R\rangle$ be an abstract argumentation system with $A = \{a, b, c, d\}$, $R = \{(a, b), (a, c), (b, a), (b, c), (c, d)\}$. Let $d \in A$ be the issue.*



We begin with the preferred game. PRO should be able to win because d does belong in some preferred extension. Actually, it belongs in two preferred extensions, as $\mathcal{E}_{Pref}(AS) = \{\{a, d\}, \{b, d\}\}$. The intuition behind the preferred game is that PRO tries to construct an admissible extension

[1] A similar approach can be found in [BH01], with the difference that arguments and attacks in that work are not abstract, instead they are based on classical logic.

[2] We remind that these semantics are found in Definition 3.

containing d. PRO can repeat an argument, while CON cannot (otherwise termination would not be ensured). In this case, the preferred game can be won by PRO in two different ways, shown below: PRO can either use argument a or argument b in order to defend d. In both cases, he is able to construct an admissible extension containing d.

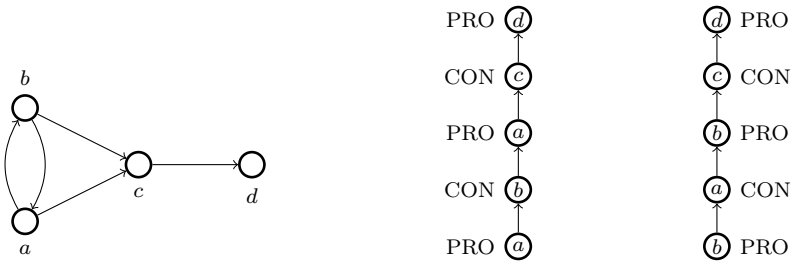


Figure 4.1: Two possible runs (disputes) of the preferred game. Both runs lead to trees consisting of a single branch. In both runs, PRO wins the dispute, as he has the last word. The fact that PRO is able to win a run, means that d belongs in some preferred extension. (Actually, d belongs in two preferred extensions: $\{d, a\}$ and $\{d, b\}$).

As far as the grounded game is concerned, as Figure 4.2 shows, the situation is quite different: PRO loses the game. The reason is that when he uses argument a to defend the issue, CON uses argument b (and vice-versa). Then, PRO has no move at his disposal, as the rules of the grounded game prevent him from re-introducing the same argument. Therefore, PRO loses. Indeed, it holds that $\mathcal{E}_{Gr}(AS) = \{\{\}\}$.

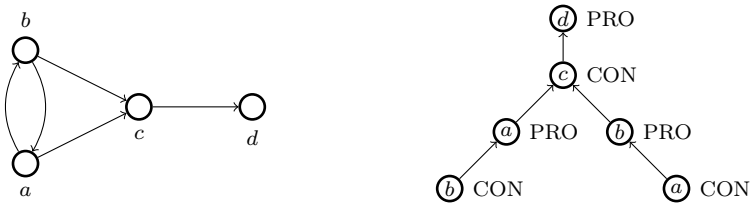


Figure 4.2: A run of the grounded game. CON wins the dispute, as he has the last word. The conclusion is that d does not belong to the grounded extension.

Closing this section, we note that in [KT99] the authors showed that different argumentation semantics for *Logic Programming (with Negation as failure)* can be computed via variations of a basic proof theory which is given in terms of derivations of trees, where each node of a tree contains an argument (or attack) against its parent node. Also similar types of dialectical proofs have been defined for Constrained Argumentation Frameworks in [DDLN10].

Now, we pass to the second type of decision procedure, which is based on labelling. The main idea is that the assignment of a specific label to an argument, has direct effects on the labels of its related arguments.

4.2 Dynamics of Labels

We have seen that there exist two major approaches for defining argument acceptability on an abstract argumentation system (apart from the approaches using numerical argument valuation): extension-based approaches and labelling-based approaches. The two are very closely related, as showed by Caminada in [Cam06].

In the study of abstract argumentation dynamics, both approaches can been useful. We have just presented some proof theories in [MC09], which study dynamics based on the extension-based approach. Now, we focus on dynamics of argument labels.

Given an abstract argumentation system $AS = \langle A, R\rangle$, we have seen previously in Chapter 2, that every complete labelling (which makes use of three labels: *in*, *out* and *undec*) satisfies the following two conditions:

- An argument is *in* if and only if all its attackers are *out* (or it has no attackers).
- An argument is *out* if and only if it has at least one *in* attacker.

We will denote by $\mathcal{L}^{AS}_{Comp}$ the set of all complete labellings of the system AS.
As we have already seen, there exist various ways to attach *in*, *out* and *undec* labels to the arguments, in order to form a complete labelling. Let us provide an example:

Example 9 *Let $AS = \langle A, R\rangle$ be an abstract argumentation system with $A = \{a, b, c, d\}$ and $R = \{(c, d), (d, c), (c, a), (c, b)\}$. This system has three complete labellings, illustrated in Figure 4.3:*

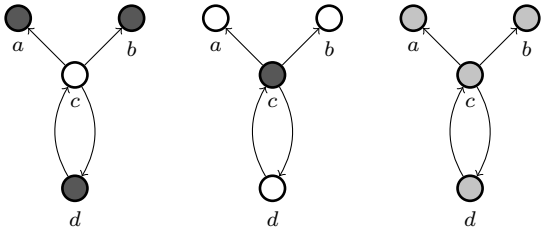


Figure 4.3: The system AS with its three complete labellings.

First, we see that there exist three different complete labellings. The choice of assigning some specific label to an argument has immediate effects on the labels of the other arguments. For example, the choice of assigning the in label to argument c forces all the other arguments' labels to be out. Similarly, the choice of assigning the undec label to some argument, forces all the other arguments' labels to be undec too.

The notion of *critical set*, introduced in [Gab09], is related to the dynamics of labels. A set of arguments X is critical, if any two complete labellings which agree on the labels they assign to X, are identical. In other words, the assignment of labels to the arguments of X enforces the values of all the other arguments' labels in the system.

Definition 25 *[Gab09] Given a system $AS = \langle A, R\rangle$, the set of arguments $X \subseteq A$ is called* **critical** *if and only if for any two complete labellings L_1, L_2, whenever L_1, L_2 agree on the arguments in X, then $L_1 = L_2$.*

Let us see the following example which is taken from [BCPR12].

Example 10 *[BCPR12] Let $AS = \langle A, R\rangle$ be a system with $A = \{a, b, c, d, e, f, g, h\}$ and $R = \{(c, a), (c, b), (c, d), (d, c), (e, f), (f, e), (g, h), (h, g)\}$. Three complete labellings of AS (among more which exist) are illustrated in Figure 4.4. It holds that $\{C, E, G\}$ (among other sets) is a critical set. Indeed, if we split the system's arguments into the following three sets: $\{A, B, C, D\}$, $\{E, F\}$ and $\{G, H\}$, then assigning one argument's label enforces all the labels of the arguments in the same set, to obtain specific values. In [BCPR12] the sets $\{A, B, C, D\}$, $\{E, F\}$ and $\{G, H\}$ are called* equivalence classes*. Therefore, assigning labels to C, E, and G, enforces the labels*

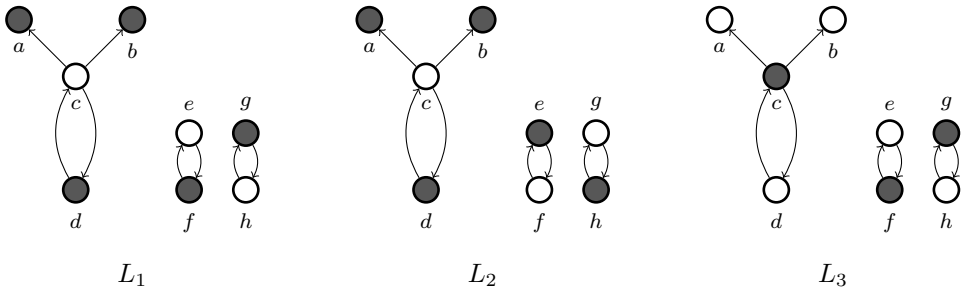


Figure 4.4: Three complete labellings of AS (L_1, L_2 and L_3). The set $\{c, e, g\}$ is critical, because by assigning labels to these arguments, the labels of all arguments in AS are enforced.

of all the arguments in the system. This fact has inspired the authors in [BCPR12] to define distances between systems, based on critical sets and equivalence classes. For example, given the three equivalence classes above, the distance from L_1 to L_2 (where 2 equivalence classes differ) could be considered greater than the distance from L_1 to L_3 (where 1 equivalence class differs).

We mentioned before that argumentation dynamics are essential in the study of change on an argumentation system. Let us now see in what sense this is true.

4.3 Dynamics and Change

The two works presented above illustrate some basic features of abstract argumentation dynamics. Either using extension-based acceptability semantics (as in the work in [MC09]), or argument labelling (as in [BCPR12]), these works show how the acceptability status of an argument influences the acceptability status of the other arguments in the system. Notice that the argumentation systems considered were static, their structure could not be modified. In the following, we want to take advantage of that knowledge, in a setting where change is allowed. By "change" we mean the *structural change* of an abstract argumentation system.

As explained in the introduction of this chapter, we separate the works on change in two types: First, we present some works which elaborate on the possible effects of specific types of change. Second, we present some works dealing with the enforcing problem. In these works change is motivated by the desire to satisfy a given goal. Of course, both types of works are closely related, but their motivations differ.

4.3.1 Effects of Change on Abstract Argumentation Systems

In this section we present a couple of works which study change on abstract argumentation systems, and they strive to identify the possible effects of specific types of change.

We start with the work in [CdSCLS10], which is based on previous work in [CdSCLS08]. The authors study the case where, given an abstract argumentation system $AS = \langle A, R\rangle$, a single argument and some attacks related to it are added.

The authors start by identifying four general types of *change operations* on an abstract argumentation system:

- Adding one interaction[3] between two existing arguments.
- Removing one existing interaction.

[3]In [CdSCLS08] the only possible interactions are binary attacks.

- Adding one argument and a set of interactions concerning it.
- Removing one argument and all its interactions.

These change operations may satisfy two types of properties which are based on:

1. The comparison of the old extensions of the system with the new ones. These properties are called *structural* and they focus on relations of set-cardinality or set-inclusion.
2. The comparison of the old status of some particular argument with its new status. The authors define the properties of *Monotony* and *Priority to Recency*.

We underline that these properties are very general, in the sense that they are not connected neither to a particular type of change (among the four types mentioned above), nor to a particular semantics.

Let AS denote an initial system and AS' the system after a change operation takes place. Also, let $\mathcal{E}(AS)$ denote the set of extensions of system AS (under given semantics). The structural properties defined in [CdSCLS10] are the following:

- *Decisive change:* $\mathcal{E}(AS')$ contains a single, non-empty extension, and this was not the case for $\mathcal{E}(AS)$.
- *Restrictive change:* The number of extensions is decreased, but it remains greater than one.
- *Questioning change:* The number of extensions is increased.
- *Destructive change:* $\mathcal{E}(AS)$ contains at least one extension (and no empty extension), and $\mathcal{E}(AS') = \emptyset$ or $\mathcal{E}(AS') = \{\{\}\}$.
- *Expansive change:* $\mathcal{E}(AS)$ and $\mathcal{E}(AS')$ have the same number of extensions and each extension of $\mathcal{E}(AS')$ strictly includes an extension of $\mathcal{E}(AS)$.
- *Conservative change:* The extensions remain the same, thus $\mathcal{E}(AS) = \mathcal{E}(AS')$.
- *Altering change:* $\mathcal{E}(AS)$ and $\mathcal{E}(AS')$ have the same number of extensions and there exists at least one extension $E_i \in \mathcal{E}(AS)$ such that $\forall E'_j \in \mathcal{E}(AS')$, $E_i \not\subseteq E_j$.

Next, the authors focus on two properties of change, based on the status of the arguments. The first property is Monotony, and it refers to the arguments which remain accepted after a change takes place. The second property is Priority to Recency, and it refers to the acceptability of newly added arguments.

Definition 26 *[CdSCLS10] The change from system AS to AS satisfies* **Monotony** *if and only if each extension of AS is included in at least one extension of AS'.*
It satisfies **Credulous Monotony** *if and only if the union of the extensions of AS is included in the union of the extensions of AS'.*
It satisfies **Skeptical Monotony** *if and only if the intersection of the extensions of AS is included in the intersection of the extensions of AS'.*
It satisfies **Partial Monotony for argument** *a if and only if when a belongs to an extension of AS, it also belongs to at least one extension of AS'.*

If we consider the addition of a single argument a in system AS, then Priority to Recency is defined as follows:

Definition 27 *[CdSCLS10] The change from AS to AS' satisfies* **Priority to Recency** *if and only if AS' has at least one extension, and the added argument a belongs to each extension of AS'.*

Then, the authors study the existing connections between, on one hand, the different types of expansions and, on the other hand, the properties of Monotony and Priority to Recency.

The work in [CdSCLS10] focuses on a specific change operation, the addition of a single argument and its interactions. Let us provide its formal definition:

Definition 28 *[CdSCLS10] Adding a single argument $z \notin A$ and a set of interactions concerning z, denoted by I_z, is a change operation defined by: $\langle A, R\rangle \bigoplus_i^a (z, I_z) = \langle A \cup \{z\}, R \cup \{I_z\}\rangle$. Here, I_z is supposed to be a non-empty set of pairs of arguments (either of the form (x, z) or (z, x) with $x \in A$.*

Finally, we provide a small example of a change and the properties it satisfies:

Example 13 (cont.) *Let us consider again the system $AS = \langle A, R\rangle$, with $A = \{a, b, c\}$ and $R = \{(a, b), (b, c), (c, b)\}$. If we add into the system the argument d and the attacks $(d, a), (a, d)$, then the modified system is $AS' = \langle A \cup \{d\}, R \cup \{(d, a), (a, d)\}\rangle$.*

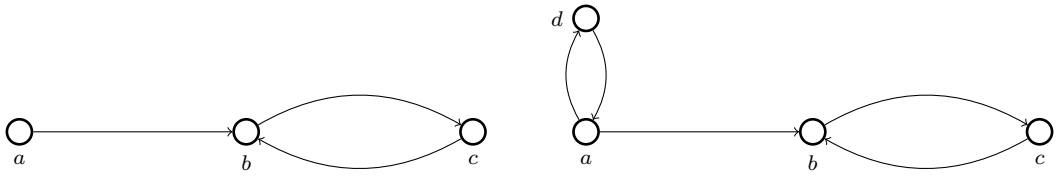


Figure 4.5: The addition of an argument d and its interactions. Initially we had $\mathcal{E}_{Gr}(AS) = \{\{a, c\}\}$. After the change we have $\mathcal{E}_{Gr}(AS') = \{\{\}\}$.

Assume that we are interested in the grounded semantics. Initially, it holds that $\mathcal{E}_{Gr}(AS) = \{\{a, c\}\}$. After the addition of argument d and its interactions, the grounded extension becomes empty: $\mathcal{E}_{Gr}(AS') = \{\{\}\}$. This change is called, according to [CdSCLS10], destructive. As a result, Monotony is not satisfied. Partial Monotony is satisfied only for arguments b and d, because they do not belong to the grounded extension, neither before, nor after the change. Finally, Priority to Recency does not hold, as the newly added argument d does not belong to the grounded extension of AS'.

We also mention the related work in [BCdSCLS11] which studies, under a similar point of view, the *deletion of an argument*, as well as the work in [BCdSCLS12] which studies the relation between adding and deleting an argument.

The work of Boella et al. in [BKvdT09a, BKvdT09b] uses a classical, three-valued argument labelling. Given a system AS, the authors separate its arguments into accepted (*in*), rejected (*out*), and undecided (*undec*). The acceptance functions $\mathcal{A}(AS)$, $\mathcal{R}(AS)$, $\mathcal{U}(AS)$ return, respectively, the sets of accepted, rejected, and undecided arguments of AS. The authors focus on grounded semantics, because it leads to a unique three-valued labelling. Then, they provide conditions under which the set of accepted arguments remains the same, after a structural change takes place on the argumentation system: an attack (or argument) addition is called *refinement*, while an attack (or argument) removal is called *abstraction*. The operations of attack refinement, attack abstraction and argument abstraction are studied.

In order to give an idea of their work in [BKvdT09a, BKvdT09b], let us state the different principles of attack abstraction, which describe under which conditions an attack removal leaves the set of accepted arguments unchanged.

Definition 29 *[BKvdT09a] An acceptance function $\mathcal{A}$ satisfies* **the $\mathcal{XY}$ abstraction principle**, *where $\mathcal{X}, \mathcal{Y} \in \{\mathcal{A}, \mathcal{R}, \mathcal{U}\}$, if for every argumentation system $AS = \langle A, R\rangle$, $\forall a \in \mathcal{X}(AS)$, $\forall b \in \mathcal{Y}(AS)$ it holds that $\mathcal{A}(AS) = \mathcal{A}(\langle A, R \setminus \{(a, b)\}\rangle)$.*

Let us see which attack abstraction principles are satisfied by the grounded semantics.

Proposition 1 *[BKvdT09a] The grounded semantics satisfies the $\mathcal{AA}$, $\mathcal{AU}$, $\mathcal{UA}$, $\mathcal{UR}$, $\mathcal{RA}$, $\mathcal{RU}$ and $\mathcal{RR}$ attack abstraction principles, and it does not satisfy the $\mathcal{AR}$ and $\mathcal{UU}$ attack abstraction principles.*

The abstraction principles $\mathcal{AA}$, $\mathcal{AU}$, and $\mathcal{UA}$ hold vacuously, since it cannot exist an attack from an accepted argument against an accepted or undecided one (so there is no actual attack removal in these cases). It also cannot exist an attack from an undecided argument against an accepted one.

Also, the abstraction principles $\mathcal{RA}$, $\mathcal{RU}$, and $\mathcal{RR}$ hold, as attacks from rejected arguments do not influence the grounded extension. This principle holds, no matter the type of the attacked argument.

Finally, the abstraction principle $\mathcal{UR}$ holds, as the attacks on a rejected argument by an undecided argument do not influence the extension. Intuitively, this means that an argument is rejected only when it is attacked by an accepted argument.

Let us see an example.

Example 13 (cont.) *Let us start with a slightly modified system: $AS = \langle A, R\rangle$, with $A = \{a, b, c, d\}$ and $R = \{(a, b), (b, c), (c, b), (d, a)\}$. In the initial system, we have: $\mathcal{A}(AS) = \{d\}$, $\mathcal{R}(AS) = \{a\}$, $\mathcal{U}(AS) = \{b, c\}$. If we remove the attack (d, a), then the grounded extension changes: $\mathcal{A}(AS') = \{d, a, c\}$. Initially, $d \in \mathcal{A}$ and $a \in \mathcal{R}$, and after the attack's removal the grounded extension is modified. Therefore, this is a counter-example, showing that under grounded semantics the $\mathcal{AR}$ attack abstraction principle does not hold. On the contrary, if we remove the attack (a, b) from the initial system, then the grounded extension does not change and $\mathcal{A}(AS'') = \mathcal{A}(AS) = \{d\}$. This is an illustration of the $\mathcal{RU}$ attack abstraction principle.*

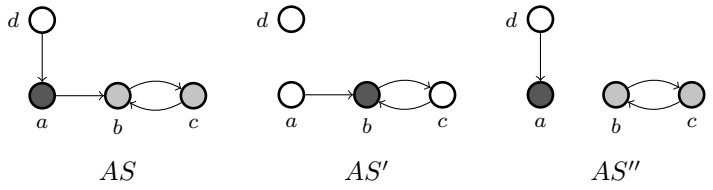


Figure 4.6: Two different abstractions (removals) of a single attack from AS, leading to the modified systems AS' and AS''.

Finally, another work which studies the effects of change on an abstract system, but under a different perspective, is found in [OW11], which introduces the notion of *strong equivalence*. While standard equivalence holds between two systems AS_1 and AS_2 if they possess the same extensions, strong equivalence holds between AS_1 and AS_2 if they are equivalent when conjoined with any other (same for both) system AS_3. This means that after the addition of some arguments and attacks (the same for both systems), they remain equivalent. The notion of *local equivalence* is more constrained, as it only considers additions of attacks between arguments already on systems AS_1 and AS_2. The authors of [OW11] provide specific conditions under which strong and local equivalence hold, for different types of Dung semantics.

4.3.2 The enforcing problem

The work in [BGP+11] is the first we present that studies the enforcing problem. It relies on the directionality principle, according to which in an abstract argumentation system $AS = \langle A, R\rangle$, the acceptability status of an argument $a \in A$ depends on the acceptability status of its attackers.

As opposed to the previous works, argumentation systems are assumed to be dynamic, but only a very specific type of change may occur: the addition of new attacks (coming from arguments outside AS) against some arguments of the system.

For a given status (label) of $a \in A$ which is desired (*in*, *out*, or *undec*, in the context of a reinstatement labelling of AS) the authors want to compute its corresponding *target sets*. These are the sets of arguments which, if they get attacked, then a can have the desired status. In order to compute target sets, the authors introduce the notion of *conditional labelling* which allows to attach to each argument the information about its possible justification statuses, depending on the changes in the system.[4] In other words, an argument's conditional labels indicate which arguments should be attacked in order for that argument to be *in*, *out*, or *undec*. This information is encoded in the form of propositional formulae.

Given an argumentation system $AS = \langle A, R \rangle$, each argument $a \in A$ is associated with three formulae: a^+, a^-, $a^?$. The meaning of formula a^+ (resp. a^-, $a^?$) is that argument a must be *in* (resp. *out*, *undec*). A generic formula associated to argument $a \in A$ is denoted a^*. The language of conditional labels is the following:

Definition 30 *[BGP+11]*

- *If $b \in A$, then b^+, b^-, $b^?$ are formulae.*
- *If $b \in A$, then b° is a formula (and it means that argument b must be attacked).*
- *$\top$ and $\bot$ are formulae ($\top$ stands for success, $\bot$ stands for failure).*
- *If a_1^* and a_2^* are formulae, then $a_1^* \wedge a_2^*$ and $a_1^* \vee a_2^*$ are formulae.*

As we said before, a conditional label of an argument indicates which arguments of the system must be attacked in order for the status of that argument to be *in*, *out*, or *undec*. Let us see what a conditional label consists of.

Definition 31 *[BGP+11] A* **conditional label** *$a^i : body_a^i$ (where a is an argument, $i \in \{+, -, ?\}$ and $body_a^i$ is a formula) is a relation between a justification status and a set of arguments to attack. The justification status is expressed by the head of the conditional label: a^i means that argument a is labelled in (resp. out, undec) when $i = +$ (resp. $i = -$, $i =?$). The set of arguments to attack is expressed by the body of the conditional label ($body_a^i$).*

From a conditional label whose body contains only atomic formulas of the type x° (connected with $\wedge$ and $\vee$), we can easily find its corresponding target sets, by first putting the body in *Disjunctive Normal Form* (DNF)[5]. Let us provide a small example:

Example 11 *The conditional label $a^+ : (b^\circ \wedge c^\circ) \vee (b^\circ \wedge d^\circ)$ conveys the following information: In order for argument a to be in, we must either attack both b and c, or both b and d. Notice that the body of the conditional label is already in DNF. So, the target sets for making argument a in are: $\{b, c\}$ and $\{c, d\}$.*

The authors then ask the following question: Starting from a given argument, how can we compute the target sets (sets of arguments to get attacked) which enforce its justification status? In order to answer this question, they follow a similar approach as Modgil and Caminada in [MC09].

Their approach for target set computation can be decomposed in three main phases:

1. Associate each argument to three (initial) labels (for *in*, *out* and *undec*).

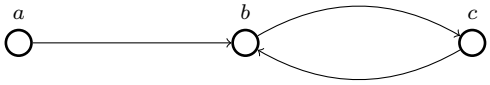
[4] In [BW10] the authors define *acceptance conditions* for arguments, with the difference that they consider *Abstract Dialectical Frameworks*, not systems with "simple" binary attack relations.

[5] A formula in DNF is a disjunction of conjunctive clauses.

2. Compute the (final) conditional labels by substitutions (as explained next).
3. Find the target sets: once the conditional labels are computed, then their equivalent DNF formulas are computed, and finally the target sets are obtained (as shown in the following example).

Let us provide an illustrative example of the procedure:

Example 12 *[BGP$^+$11] Let $AS = \langle A, R\rangle$ be an abstract argumentation system with $A = \{a, b, c\}$ and $R = \{(a, b), (b, c), (c, b)\}$.*



Let us focus on argument b. We want to compute its conditional labels, indicating when it is in, out, and undec. Let us start with the in label. The procedure starts with the initial conditional label $b^+ : a^- \wedge c^-$ because in order for b to be in, both arguments a and c must be out. The only way for a to be out is to add an attack against it (denoted a^0). As far as c is concerned, there are two ways to be out: either an attack against c is introduced (from outside the system), or we ensure that argument b is in (denoted b^+). But the latter is just a repetition of the initial goal we had, therefore b^+ will get replaced by $\top$, as we will see in what follows.

Let us see the actual computation of the conditional labelling for making b in.

$b^+ : a^- \wedge c^-$
$= a^- \wedge (b^+ \vee c^\circ)$
$= a^- \wedge (\bot \vee c^\circ)$ *(since b^+ also appears in the head, it is replaced by $\bot$)*
$= a^- \wedge c^\circ$ *($\bot \vee c^\circ$ is simplified into c°)*[6]
$= a^\circ \wedge c^\circ$

The conclusion drawn from the final conditional labelling is that there exists only one way to turn b in (because the final formula has only one disjunct): Both arguments a and c must get attacked.

If we want b to be out, we have three choices: make a in, or make c in, or finally attack b. Let us see the computation of the conditional labelling for making b out:

$b^- : a^+ \vee c^+ \vee b^\circ$
$= a^+ \vee b^- \vee b^\circ$
$= a^+ \vee \bot \vee b^\circ$ *(since b^- also appears in the head, it is replaced by $\bot$)*
$= a^+ \vee b^\circ$
$= \top \vee b^\circ$ *(since a^+ is always verified, it is replaced by $\top$)*
$= \top$ *($\top \vee b^\circ$ is simplified into $\top$)*

The conclusion is that no move is required in order for b to be labelled out (b is already out).

Finally, we compute the conditional labelling for making b undec. One of its attackers must be undec (either a or c), and every attacker must be either out or undec.

$b^? : (a^? \vee c^?) \wedge (a^- \vee a^?) \wedge (c^- \vee c^?)$
$= (\bot \vee b^?) \wedge (a^\circ \vee \bot) \wedge ((b^+ \vee c^\circ) \vee b^?)$ *(there is no way to make $a^?$)*
$= b^? \wedge a^\circ \wedge ((b^+ \vee c^\circ) \vee b^?)$
$= \top \wedge a^\circ \wedge ((\bot \vee c^\circ) \vee \top)$ *(since $b^?$ is in the head, $b^?$ (resp. b^+) is replaced by $\top$ (resp. $\bot$))*
$= \top \wedge a^\circ \wedge \top$
$= a^\circ$

The conclusion is that b can be labelled undec by attacking a.

[6] We will not provide all the simplification rules here, they can be found in [BGP$^+$11].

Concluding, the work in [BGP$^+$11] studies the enforcing problem. The authors use the notion of conditional labelling, in order to compute possible modifications of a system, achieving a particular justification status for a particular argument. The only allowed change is the addition of attacks against arguments of the system, and these attacks must come from outside the system.

Though the computation of target sets is the result of a procedure focused on a single argument, and as a result only "relevant" attackers and defenders are considered, the authors do not make specific claims on the minimality of target sets. In our work, we will revisit the notion of target set, but we will insist on minimality. Moreover, we will allow for a more general type of change. We will consider that: (i) both attack additions and attack deletions are possible (while [BGP$^+$11] only allows attack additions), and that (ii) attacks between arguments in the system may change (while [BGP$^+$11] only allows attack additions coming from outside the system).

Let us proceed to some more works studying the enforcing problem.

In [CMKMM14] the authors use *revision formulae* to describe complex goals, related to the acceptability of arguments. For example, a goal can be the following: "Either arguments a and b get accepted, or argument b gets accepted and argument c gets rejected". That goal can be represented as a propositional formula $\phi = ((a \wedge b) \vee (b \wedge \neg c))$. The work does not focus on any particular semantics, and in that sense it is quite general. An essential element of this work is that the change achieving a goal must be minimal, according to two criteria. The first (and most important) criterion is the minimality of the set of arguments whose status will be affected by the change. The second criterion is the minimality of the number of structural modifications. The only type of structural modification possible is the addition and removal of attacks (not of arguments). The revision operators which are defined, are shown to satisfy a number of rationality postulates, reflecting the AGM belief revision postulates in [AGM85].

The work in [BB10] studies the problem of revising an abstract argumentation system by adding finitely many new arguments which may interact with old ones. The authors identify several types of revisions. *Normal expansions* introduce new arguments and possibly new attack relations. The latter have to involve at least one new argument. *Strong* and *weak expansions* are normal expansions, where the direction of the new attacks is restricted. In strong extensions each new attack comes from a new argument, while in weak extensions each new attack is set against a new argument. The authors study the behaviour of a system's extensions when: (i) an expanded of the above types takes place, or (ii) the underlying semantics is changed, or (iii) both the above happen. If the semantics is changed, the revision is called *liberal*, otherwise it is called *conservative*. The contributions of [BB10] can be resumed in the following: impossibility results with respect to enforcing a desired set of arguments, possibility results for enforcing desired sets of arguments (for specific semantics), and monotonicity results for weak expansions.

Baumann in [Bau12] studies the enforcing problem, under the consideration of minimal change. In that work, change may consist in adding/removing arguments and attacks. The specificity of this work is that it introduces upper and lower bounds on the number of elementary changes (effort) needed. A numerical distance between two systems is defined, based on the number of attack relations upon which the systems disagree. The author considers admissible, complete, preferred, stable and semi-stable semantics, and the following different types of modifications are studied in turn: Normal expansions (weak and strong), arbitrary expansions and arbitrary modifications. Arbitrary expansions are normal expansions which may also add attacks among old arguments. Arbitrary modifications may also remove attacks. Finally, the notions of *minimal-E-equivalence* and *minimal change equivalence* are introduced. Two systems are called minimal-E-equivalent if the minimal effort needed to enforce a subset of arguments E is the same for both. If minimal-E-equivalence holds for every subset E of the AFs in question, the two systems are minimal change equivalent.

In [DHP14] the authors make a logical analysis of the dynamics of abstract argumentation systems. They express attack relation and argument status by means of propositional variables

and they define acceptability criteria by using formulas of propositional logic. Dynamics are studied in terms of basic operations on propositional variables.

Finally, the work in [BGK$^+$14] studies the enforcing problem under a different perspective. It develops a model of abduction in abstract argumentation, where changes on an argumentation system are hypotheses which may be used to explain an observation (e.g. the fact that a specific argument is accepted). The authors present dialogical proof theories for the main decision problems (i.e. finding hypotheses that explain the skeptical, or the credulous, support of an observation).

Chapter 5

Abstract Argumentation Dynamics[1]

In this chapter we present our contribution to the topic of abstract argumentation dynamics.

First, we define a type of dynamic argumentation system, which we call *argumentation system with modifiable attacks (ASMA)*. The intuition is that, given a set of abstract arguments A and a binary attack relation R over them, some specific attacks in R can potentially be removed from R, while other specific attacks, not initially in R, can potentially be added. Therefore, in this type of system, change consists in the addition and/or removal of some attacks. On the other hand, we assume that the ASMA's set of arguments A remains unchanged: arguments can neither be added nor removed.

Our contribution to the dynamics of abstract argumentation has two focal points:

1. The *enforcing problem* and some related issues.

2. The definition and analysis of the notion of *controversy* of a dynamic argumentation system.

The enforcing problem:
Initially, given an ASMA and an argumentative goal (e.g. the acceptance of a specific argument under given semantics), we search for the possible ways to achieve this goal, by making changes to the initial system (adding and/or deleting attacks). Basic properties of such "successful" change are studied for different semantics.

Next, we turn our attention to *minimal* successful change, inspired by the notion of target set, introduced in [BGP$^+$11]. By minimal successful change we refer to the minimal sets (with respect to the $\subseteq$ relation) of attacks whose addition/removal achieves a given goal. Minimal change is a particularly interesting type of change, as it consists in the easiest and most economical way to achieve a given goal.

Afterwards, inspired from [MC09, BGP$^+$11], we define a procedure which is used for the computation of target sets. We prove that it satisfies some important properties, mainly that it computes all the target sets for some particular types of goals.

Next, we focus on the meta-dynamics of ASMAs, in the following sense: we study how target sets may evolve, when different types of change take place on an ASMA. Our results are used to explain in what sense focusing (resp. not focusing) on target sets can be a good (resp. bad) idea when we desire to achieve a given goal.

Let us stress out that our study of the enforcing problem is valuable in two different settings. Firstly, in a single-agent setting, an agent having a dynamic argumentation system may be focusing on the status of a specific argument. That agent may be interested in identifying under which modifications of his system, that argument can become accepted, or rejected. There are many cases where this knowledge may be valuable to the agent. For example, if the agent reasons

[1] Our publications relevant to this chapter are: [KBMM12, KBM$^+$13, KPB$^+$13, KBMM14b].

abductively, and believes that he should (resp. should not) change his mind on the acceptability status of that argument, then he should be able to identify which attacks to reconsider (resp. to not reconsider). Secondly, in a multi-agent setting, where a debate is represented in the form of a single dynamic argumentation system (Gameboard), and where the debating agents' votes on its attacks can lead to their addition or removal, an agent may be interested in computing all the possible ways in which he can influence the debate in his favour.

The study of a system's controversy (stability):
Finally, in the last section of this chapter, we focus on a different aspect of dynamics, which is the controversy of a dynamic argumentation system. We shall focus on the multi-agent scenario, though our findings can also be used in single-agent scenarios. We assume a multilateral debate, which uses a central Gameboard. The Gameboard's attacks are weighted, and they are subject to change (addition/removal), according to the opinions expressed by the debating agents. Our task is to identify the characteristics of controversial debates. Roughly, controversial debates have conclusions which are "easily" put into question (e.g. a single additional agent's opinion could had reversed the debate's outcome). In order to do this, we introduce three evaluation criteria of a debate's controversy: (i) how easily its attacks can be added/removed, (ii) how easily its arguments can change acceptability status and (iii) how many arguments have undecided status. Also, we analyze how we can choose an additional expert (among a pool of experts), based on his ability to render given debate less controversial, by providing his opinion.

5.1 Abstract Argumentation System with Modifiable Attacks (ASMA)

In the definition of a specific type of abstract argumentation system we provide here, the difference to Dung's abstract argumentation system in [Dun95], is that we do not only have the standard attack relation (here denoted R), but we also have a relation R^+ which denotes the attacks which can be added to the system, and a relation R^- which denotes the attacks which can be removed from the system.

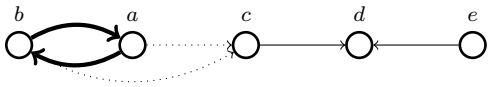
Definition 32 *An* **abstract argumentation system with modifiable attacks (ASMA)** *is a tuple $\langle A, R, R^+, R^- \rangle$, where A is a finite set of abstract arguments, $R \subseteq A \times A$ is a binary attack relation between arguments, $R^+ \subseteq A \times A$, with $R^+ \cap R = \{\}$, contains the pairs of arguments which can be added in R, and $R^- \subseteq R$ contains the pairs of arguments which can be removed from R. $R^{fix} = R \setminus R^-$ is called the set of* **fixed attacks**, *$R^{fixN} = (A \times A) \setminus (R \cup R^+)$ is called the set of* **fixed non-attacks**, *and $R^{deb} = R^+ \cup R^-$ is called the set of* **debated attacks**.

Also, for convenience, we denote $Att = R \cup R^+$ the set of attacks which are either on the system, or can be added to it.

From now on, we will focus on the attack relations more than on the arguments. We will then need the following definition.

Definition 33 *Let $\langle A, R, R^+, R^- \rangle$ be an ASMA, and let $x = (a, b) \in Att$. We refer to the argument a as the* **tail of the attack** *x, denoted $tail(x) = a$, and we refer to the argument b as the* **head of the attack** *x, denoted $head(x) = b$. Let also $y \in Att$. We will say that x* **hits** *y, denoted $hits(x, y)$, iff $head(x) = tail(y)$.*

Example 13 *Let $\langle A, R, R^+, R^- \rangle$ be an ASMA such that $A = \{a, b, c, d, e\}$, $R = \{(a, b), (b, a), (c, d), (e, d)\}$, $R^+ = \{(a, c), (b, c)\}$, $R^- = \{(c, d), (e, d)\}$. This system can be represented as follows:*



As far as the argumentation graph of the system $\langle A, R, R^+, R^- \rangle$ is concerned:

- *Fixed attacks (in $R \setminus R^-$) are represented by thick arrows.*
- *Removable attacks (in R^-) are represented by normal arrows.*
- *Addable attacks (in R^+) are represented by dotted arrows.*

For the sake of convenience, we will sometimes denote an attack $(a, b) \in Att$, simply as ab. In the above system it holds that $hits(ab, ba)$ and $hits(ba, ab)$ (as, by definition, $tail(ba) = b$, $head(ba) = a$).

In what follows we will use extension-based and labelling-based acceptability. The basic definitions of conflict-freeness, acceptable argument w.r.t. a set of arguments, and the different types of extensions, are essentially the same as in [Dun95], because given an ASMA $\langle A, R, R^+, R^- \rangle$, we just focus on its attack relation R in order to decide on the arguments' acceptability. For example, here are the definitions of conflict-free set and of acceptable argument w.r.t. a set of arguments:

Definition 34 *Let $\langle A, R, R^+, R^- \rangle$ be an ASMA, and let $C \subseteq A$. The set C is* **conflict-free** *iff $\nexists x \in R$ such that $tail(x) \in C$ and $head(x) \in C$. An argument $a \in A$ is* **acceptable w.r.t.** *C iff $\forall x \in R$: if $head(x) = a$, then $\exists y \in R$ such that $hits(y, x)$ and $tail(y) \in C$.*

In what follows, we use the following types of semantics: admissible, preferred, complete and grounded. Also, we consider both credulous and skeptical acceptability.

Definition 35 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA, and let $a \in A$. Argument a is said* **credulously accepted** *w.r.t. system SM under semantics $S \in \{Adm, Pref, Comp, St, Gr\}$, denoted $S_\exists(a, SM)$, iff a belongs to at least one extension of SM under the S semantics. Argument a is said* **skeptically accepted** *w.r.t. system SM under semantics $S \in \{Adm, Pref, Comp, St, Gr\}$, denoted $S_\forall(a, SM)$, iff a belongs to all the extensions of SM under the S semantics.*

We remind that we denote by $Sem = \{Adm, Pref, Comp\}$ the set of admissible, preferred and complete semantics. Moreover, for the sake of readability, if there is no danger of confusing which ASMA we refer to, we will simply write $\forall S \in Sem$, $S_\exists(a)$, or $S_\forall(a)$, without mentioning the specific ASMA.

5.2 Goals and (Minimal) Successful Change

In [BGP+11] the authors proposed a new kind of labelling, called *conditional labelling*. The idea is to provide the agents with a way to discover the arguments they should attack to get a particular argument accepted or rejected. Given a conditional labelling, the agents have complete knowledge about the consequences of the attacks they may raise on the acceptability of each argument, without having to recompute the labelling for each possible set of attacks they may raise.

In the context of this work, since attacks can be put into question, but arguments cannot, we regard the attack relation as the core component of an argumentation system. Therefore, we have chosen to use the *attack semantics* [VBvdT11], which is arguably more convenient for the analysis that follows. We repeat that, we will focus on cases (e.g. at some point during a debate), where it is assumed that the arguments of a system cannot change (neither new arguments can be added, nor already inserted arguments can be removed). Instead, the only change that can possibly happen is the addition of new attacks and the removal of some already inserted attacks. A central notion, related to this type of change, is the following notion of *atom*.

Definition 36 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA, let $x \in Att$, and let $d \in A$. An* **atom** *of SM is defined as follows:*

$$\begin{aligned} Atom(SM) \quad ::= \quad & \top \mid \bot \mid (x, +, \#) \mid (x, -, \#) \mid (x, 1, \#) \mid (x, 0, \#) \mid (x, ?, \#) \mid \\ & (x, 1, *) \mid (x, 0, *) \mid (x, ?, **) \mid (x, ?, *) \mid PRO(d) \mid CON(d) \end{aligned}$$

Atoms $\top$, $\bot$, $(x,+,\#)$, $(x,-,\#)$, $(x,1,\#)$, $(x,0,\#)$ *and* $(x,?,\#)$ *are called* **closed atoms**, *whereas atoms* $(x,1,*)$, $(x,0,*)$, $(x,?,**)$, $(x,?,*)$, $PRO(d)$ *and* $CON(d)$ *are called* **open atoms**.

Let us provide the meaning of the different atoms:

- Atom $(x,+,\#)$ (resp. $(x,-,\#)$) indicates the action of adding (resp. removing) the attack x from the system.
- Atom $(x,1,*)$ (resp. $(x,?,*)$, resp. $(x,0,*)$) indicates that we must find a way for attack x to become '1' (resp. '?', resp. '0').[2]
- Atom $(x,1,\#)$ (resp. $(x,?,\#)$, resp. $(x,0,\#)$) indicates that we have already found a way for attack x to become '1' (resp. '?', resp. '0'), according to the attack semantics defined in [VBvdT11].
- Atoms $PRO(d)$ and $CON(d)$ are refer to the acceptability status of d. Their exact meaning will be explained later.
- Atom $\bot$ indicates failure, whereas $\top$ indicates success.

By using the atoms $(x,+,\#)$ and $(x,-,\#)$, we define the notion of *move* on an ASMA:

Definition 37 *Let* $SM = \langle A, R, R^+, R^-\rangle$ *be an ASMA, and let* $m = \{(x,s,\#) \mid x \in Att,\ s \in \{+,-\}\}$ *be a set of atoms.* m *is called* **move on system** SM *iff* $\forall (x,+,\#) \in m$, $x \in R^+$, *and* $\forall (x,-,\#) \in m$, $x \in R^-$.
The **resulting system of playing move** m **on** $SM = \langle A, R, R^+, R^-\rangle$ *is the argumentation system* $\Delta(SM,m) = \langle A, R_m, R_m^+, R_m^-\rangle$, *where:*

1. $x \in R_m$ *iff either* $x \in R$ *and* $(x,-,\#) \notin m$, *or* $(x,+,\#) \in m$.
2. $x \in R_m^+$ *iff either* $x \in R^+$ *and* $(x,+,\#) \notin m$, *or* $(x,-,\#) \in m$.
3. $x \in R_m^-$ *iff either* $x \in R^-$ *and* $(x,-,\#) \notin m$, *or* $(x,+,\#) \in m$.

Example 13 (cont.) *Move* $m = \{(ed,-,\#),(ac,+,\#)\}$ *on system* SM *leads to the following system* $SM' = \Delta(SM,m)$*:*

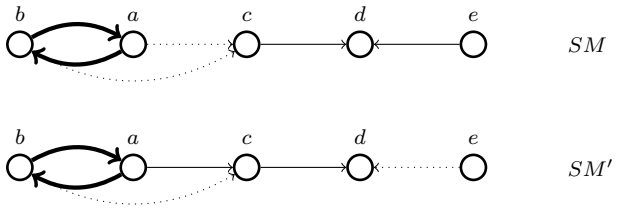


Figure 5.1: The effect of playing $m = \{(ed,-,\#),(ac,+,\#)\}$ on SM.

The choice to play a move on an system $SM = \langle A, R, R^+, R^-\rangle$, may be motivated by the desire to satisfy a specific argumentative goal. Let us formally define this notion of goal.

Definition 38 *Let* $Systems$ *be the set of all the possible ASMAs, and let* $Props$ *be a set of properties, such that* $\forall P \in Props$, P *can refer to any* $SM \in Systems$. *We define the function* f*:* $Props \times Systems \rightarrow \{true, false\}$, *such that* $\forall P \in Props, \forall SM \in Systems$, *it holds that* $f(P,SM) = true$ *iff property* P *holds, when referring to system* SM*; otherwise* $f(P,SM) = false$. *A property* P *may be chosen as a* **positive goal***: we say that goal* P *is satisfied in* SM *iff* $f(P,SM) = true$. *A negated property* $\neg P$ *may be chosen as a* **negative goal***: we say that goal* $\neg P$ *is satisfied in* SM *iff* $f(P,SM) = false$ *(that is iff* $f(\neg P,SM) = true$*).*

[2]The atom $(x,?,**)$ is similar to $(x,?,*)$, their difference is explained later.

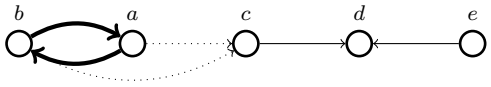
If a specific (positive or negative) goal g is not satisfied in an ASMA SM, then we search for possible moves on SM leading to a modified system where g is satisfied. Any move on SM which achieves this, is called *successful move* for g. Such a succesful move is called *target set* for g if the changes induced by it on SM are minimal.

Definition 39 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA. Let g be a goal, such that $g = P$ or $g = \neg P$, where P is a property referring to an ASMA. Finally let m be a move on M. m is called* **successful move for goal** *g iff goal g is satisfied in $\Delta(SM, m)$, that is if $f(g, \Delta(SM, m)) = true$. m is called* **target set for goal** *g iff m is minimal w.r.t. $\subseteq$ among all the successful moves for g.*

Let us now describe some specific types of goals we will focus on. Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA, let m be a move on SM, let $X \in \{\exists, \forall\}$ and finally let $S \in Sem$. We focus on the acceptance of a single argument $d \in A$ called the *issue*, and we consider two types of goals:

1. $S_X(d)$ is a positive goal. The set of successful moves for $S_X(d)$ is denoted $\mathbb{M}^S_X(SM)$, and the set of target sets $\mathbb{T}^S_X(SM)$.
2. $\neg S_X(d)$ is a negative goal. The set of successful moves for $\neg S_X(d)$ is denoted $\mathbb{M}^S_{\neg X}(SM)$, and the set of target sets $\mathbb{T}^S_{\neg X}(SM)$.[3]

Example 13 (cont.)



Let $d \in A$ be the issue. d does not belong to any admissible extension of AS. The goal $S_\exists(d)$ consisting in placing d in some admissible (or preferred, or complete) extension has three target sets: $\mathbb{T}^S_\exists = \{\{(ed, -, \#), (cd, -, \#)\}, \{(ed, -, \#), (ac, +, \#)\}, \{(ed, -, \#), (bc, +, \#)\}\}$.
Moreover, we have $\{(ed, -, \#), (bc, +, \#), (ac, +, \#)\} \in \mathbb{M}^S_\exists$, as this move is successful for $S_\exists(d)$, but it is not a target set, as it is not minimal. Now, regarding skeptical preferred semantics, there are two target sets, as $\mathbb{T}^{Pref}_\forall = \{\{(ed, -, \#), (cd, -, \#)\}, \{(ed, -, \#), (bc, +, \#), (ac, +, \#)\}\}$. Finally, as far as grounded semantics is concerned, there is a single target set, as $\mathbb{T}^{Comp}_\forall = \{\{(ed, -, \#), (cd, -, \#)\}\}$.

5.2.1 Relations among sets of Successful Moves and Target Sets

In the following we provide some properties describing the relations among different sets of successful moves and target sets. Based on these properties, we then provide a schematic representation of all these sets.

Property 3 *The following relations hold:*

1. $\mathbb{M}^{Comp}_\forall \subseteq \mathbb{M}^{Pref}_\forall \subseteq \mathbb{M}^S_\exists$
2. $\mathbb{M}^S_{\neg\exists} \subseteq \mathbb{M}^{Pref}_{\neg\forall} \subseteq \mathbb{M}^{Comp}_{\neg\forall}$

Proof 3 *(1) Let us begin with the case of the positive goals. If move $m \in \mathbb{M}^{Comp}_\forall$, then d is accepted in $SM' = \Delta(SM, m)$ under complete semantics (using skeptical acceptability), so d belongs in all the complete extensions of SM', therefore in all the preferred extensions of SM'. So, it holds that $m \in \mathbb{M}^{Pref}_\forall$. Thus, we have proved that $\mathbb{M}^{Comp}_\forall \subseteq \mathbb{M}^{Pref}_\forall$. Moreover, if $m \in \mathbb{M}^{Pref}_\forall$, then d belongs in all the preferred extensions of SM', therefore d belongs in at least one preferred extension of SM' (so, it also belongs in at least one admissible, and in at least one complete*

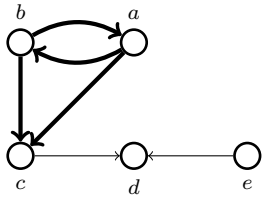
[3]When it is clear that we refer to the system SM, we shall simply write, e.g., $\mathbb{M}^S_X$ instead of $\mathbb{M}^S_X(SM)$.

extension of SM'). Thus, it holds that $m \in \mathbb{M}^S_\exists$, and we have proved that $\mathbb{M}^{Pref}_\forall \subseteq \mathbb{M}^S_\exists$. As a result, $\mathbb{M}^{Comp}_\forall \subseteq \mathbb{M}^{Pref}_\forall \subseteq \mathbb{M}^S_\exists$.

(2) Now, we consider the case of the negative goals. If move $m \in \mathbb{M}^S_{\neg\exists}$, then d does not belong in any preferred (or admissible, or complete) extension of $SM' = \Delta(SM, m)$. So, d does not belong in all the preferred extensions of SM', so it holds that $m \in \mathbb{M}^{Pref}_{\neg\forall}$. Therefore, we have proved that $\mathbb{M}^S_{\neg\exists} \subseteq \mathbb{M}^{Pref}_{\neg\forall}$. Also, if $m \in \mathbb{M}^{Pref}_{\neg\forall}$, then d does not belong in all the preferred extensions of SM', so there is a preferred extension (which is also complete) not containing d. Therefore, d does not belong in all the complete extensions of SM', so $m \in \mathbb{M}^{Comp}_{\neg\forall}$. Thus, we have proved that $\mathbb{M}^{Pref}_{\neg\forall} \subseteq \mathbb{M}^{Comp}_{\neg\forall}$. As a result, $\mathbb{M}^S_{\neg\exists} \subseteq \mathbb{M}^{Pref}_{\neg\forall} \subseteq \mathbb{M}^{Comp}_{\neg\forall}$.

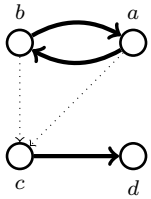
On the other hand, if we consider the corresponding sets of target sets, then neither the relations $\mathbb{T}^{Comp}_\forall \subseteq \mathbb{T}^{Pref}_\forall$, $\mathbb{T}^{Pref}_\forall \subseteq \mathbb{T}^S_\exists$, $\mathbb{T}^{Comp}_\forall \subseteq \mathbb{T}^S_\exists$, nor the relations $\mathbb{T}^S_{\neg\exists} \subseteq \mathbb{T}^{Pref}_{\neg\forall}$, $\mathbb{T}^{Pref}_{\neg\forall} \subseteq \mathbb{T}^{Comp}_{\neg\forall}$, $\mathbb{T}^S_{\neg\exists} \subseteq \mathbb{T}^{Comp}_{\neg\forall}$ hold in the general case. In order to illustrate why, we provide two counter-examples for the case of the positive goals, and two for the case of the negative goals.

Example 14 *Let $SM = \langle A, R, R^+, R^-\rangle$ be an ASMA such that $A = \{a, b, c, d, e\}$, $R = \{(a, b)$, (a, c), (b, a), (b, c), (c, d), $(e, d)\}$, $R^+ = \{\}$, $R^- = \{(c, d), (e, d)\}$, and let $d \in A$ be the issue.*



It holds that $\{(ed, -, \#), (cd, -, \#)\} \in \mathbb{T}^{Comp}_\forall$. Also, $\{(ed, -, \#)\} \in \mathbb{T}^{Pref}_\forall$ and $\{(ed, -, \#)\} \in \mathbb{T}^S_\exists$. Given that $\{(ed, -, \#)\} \subset \{(ed, -, \#), (cd, -, \#)\}$, it follows that $\{(ed, -, \#), (cd, -, \#)\} \notin \mathbb{T}^{Pref}_\forall$ and also $\{(ed, -, \#), (cd, -, \#)\} \notin \mathbb{T}^S_\exists$. So, we have proven that in the general case $\mathbb{T}^{Comp}_\forall \not\subseteq \mathbb{T}^{Pref}_\forall$ and $\mathbb{T}^{Comp}_\forall \not\subseteq \mathbb{T}^S_\exists$.

Example 15 *Let $SM = \langle A, R, R^+, R^-\rangle$ be an ASMA such that $A = \{a, b, c, d\}$, $R = \{(a, b)$, (b, a), $(c, d)\}$, $R^+ = \{(a, c), (b, c)\}$, $R^- = \{\}$, and let $d \in A$ be the issue.*

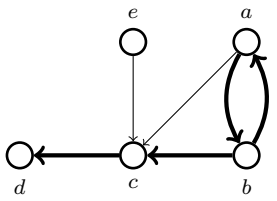


It holds that $\{(ac, +, \#), (bc, +, \#)\} \in \mathbb{T}^{Pref}_\forall$. But, $\{(ac, +, \#), (bc, +, \#)\} \notin \mathbb{T}^S_\exists$, because its subset $\{(ac, +, \#)\} \in \mathbb{T}^S_\exists$. Therefore, we have proven that in the general case $\mathbb{T}^{Pref}_\forall \not\subseteq \mathbb{T}^S_\exists$.

Example 16 *Let $SM = \langle A, R, R^+, R^-\rangle$ be an ASMA such that $A = \{a, d\}$, $R = \{(d, a)\}$, $R^+ = \{(a, d)\}$, $R^- = \{(d, a)\}$, and let $d \in A$ be the issue.*

We can verify that $\{(ad,+,\#),(da,-,\#)\} \in \mathbb{T}^{S}_{\neg\exists}$. *Also,* $\{(ad,+,\#)\} \in \mathbb{T}^{Pref}_{\neg\forall}$ *and* $\{(ad,+,\#)\} \in \mathbb{T}^{Comp}_{\neg\forall}$. *Given that* $\{(ad,+,\#)\} \subset \{(ad,+,\#),(da,-,\#)\}$, *it follows that* $\{(ad,+,\#),(da,-,\#)\} \notin \mathbb{T}^{Pref}_{\neg\forall}$ *and also* $\{(ad,+,\#),(da,-,\#)\} \notin \mathbb{T}^{Comp}_{\neg\forall}$. *Therefore, in the general case* $\mathbb{T}^{S}_{\neg\exists} \not\subseteq \mathbb{T}^{Pref}_{\neg\forall}$ *and* $\mathbb{T}^{S}_{\neg\exists} \not\subseteq \mathbb{T}^{Comp}_{\neg\forall}$.

Example 17 *Let* $SM = \langle A, R, R^{+}, R^{-}\rangle$ *be an ASMA such that* $A = \{a,b,c,d,e\}$, $R = \{(a,b),(a,c),(b,a),(b,c),(c,d),(e,c)\}$, $R^{+} = \{\}$, $R^{-} = \{(a,c),(e,c)\}$, *and let* $d \in A$ *be the issue.*



We can verify that $\{(ec,-,\#),(ac,-,\#)\} \in \mathbb{T}^{Pref}_{\neg\forall}$, *because removing both these attacks is the only way in order for argument d not to belong in all the preferred extensions. But,* $\{(ec,-,\#),(ac,-,\#)\} \notin \mathbb{T}^{Comp}_{\neg\forall}$, *because its subset* $\{(ec,-,\#)\} \in \mathbb{T}^{Comp}_{\neg\forall}$. *Therefore, in the general case* $\mathbb{T}^{Pref}_{\neg\forall} \not\subseteq \mathbb{T}^{Comp}_{\neg\forall}$.

The following property highlights a couple of links between some sets of target sets.

Property 4 *Let SM be an ASMA, and let m be a move on SM. Then the following holds:*

1. *If* $m \in \mathbb{T}^{Comp}_{\forall}$ *and* $m \in \mathbb{T}^{S}_{\exists}$, *then* $m \in \mathbb{T}^{Pref}_{\forall}$.
2. *If* $m \in \mathbb{T}^{S}_{\neg\exists}$ *and* $m \in \mathbb{T}^{Comp}_{\neg\forall}$, *then* $m \in \mathbb{T}^{Pref}_{\neg\forall}$.

Proof 4 *(1) By contradiction, let* $m \in \mathbb{T}^{Comp}_{\forall}$, $m \in \mathbb{T}^{S}_{\exists}$ *and assume that* $m \notin \mathbb{T}^{Pref}_{\forall}$. *Now,* $m \in \mathbb{T}^{Comp}_{\forall}$ *implies that* $m \in \mathbb{M}^{Comp}_{\forall}$ *(as m is minimal w.r.t.* $\subseteq$ *among the moves in* $\mathbb{M}^{Comp}_{\forall}$*). Then, from* $m \in \mathbb{M}^{Comp}_{\forall}$ *it follows that* $m \in \mathbb{M}^{Pref}_{\forall}$ *(from Property 3). Moreover, we assumed that* $m \notin \mathbb{T}^{Pref}_{\forall}$, *so there must exist another move* $m' \subset m$, *such that* $m' \in \mathbb{T}^{Pref}_{\forall}$ *(and, of course,* $m' \in \mathbb{M}^{Pref}_{\forall}$*). From* $m' \in \mathbb{M}^{Pref}_{\forall}$, *we get that* $m' \in \mathbb{M}^{S}_{\exists}$ *(from Property 3). Finally, from* $m' \in \mathbb{M}^{S}_{\exists}$ *and* $m \in \mathbb{T}^{S}_{\exists}$, *it follows that* $m \subseteq m'$. *Contradiction, since above we had* $m' \subset m$. *Therefore,* $m \in \mathbb{T}^{Pref}_{\forall}$.
(2) Similarly, we prove the second relation, regarding the negative goals. By contradiction, let $m \in \mathbb{T}^{S}_{\neg\exists}$, $m \in \mathbb{T}^{Comp}_{\neg\forall}$ *and assume that* $m \notin \mathbb{T}^{Pref}_{\neg\forall}$. *Now,* $m \in \mathbb{T}^{S}_{\neg\exists}$ *implies that* $m \in \mathbb{M}^{S}_{\neg\exists}$ *(as m is minimal w.r.t.* $\subseteq$ *among the moves in* $\mathbb{M}^{S}_{\neg\exists}$*). Then, from* $m \in \mathbb{M}^{S}_{\neg\exists}$ *it follows that* $m \in \mathbb{M}^{Pref}_{\neg\forall}$ *(from Property 3). Moreover, we assumed that* $m \notin \mathbb{T}^{Pref}_{\neg\forall}$, *so there must exist another move* $m' \subset m$, *such that* $m' \in \mathbb{T}^{Pref}_{\neg\forall}$ *(and, of course,* $m' \in \mathbb{M}^{Pref}_{\neg\forall}$*). From* $m' \in \mathbb{M}^{Pref}_{\neg\forall}$, *we get that* $m' \in \mathbb{M}^{Comp}_{\neg\forall}$ *(from Property 3). Finally, from* $m' \in \mathbb{M}^{Comp}_{\neg\forall}$, *and* $m \in \mathbb{T}^{Comp}_{\neg\forall}$, *it follows that* $m \subseteq m'$. *Contradiction, since above we had* $m' \subset m$. *Therefore,* $m \in \mathbb{T}^{Pref}_{\neg\forall}$.

Figure 5.2 graphically represents the relations between the sets of successful moves and the sets of target sets for both positive and negative goals. It illustrates the previously presented valid subset relations, as well as the subset relations which do not hold (as it has been shown through counterexamples).

Finally, the following property shows that, in the general case, the union of two target sets is not a successful move.

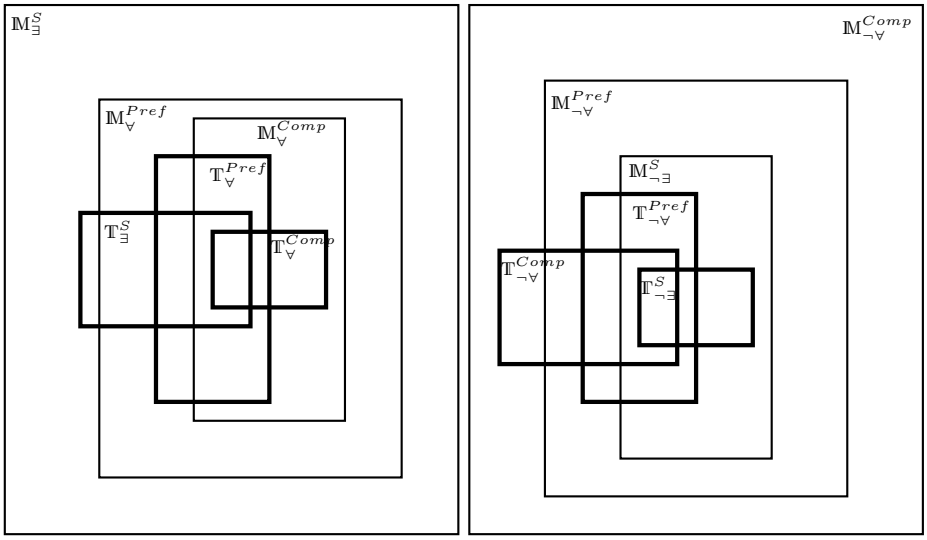


Figure 5.2: On the left: The sets of successful moves and target sets for the positive goals. On the right: The sets of successful moves and target sets for the negative goals.

Property 5 *Let SM be an ASMA, where $t_1, t_2 \in \mathbb{T}$ are two target sets for some goal g. Then, in the general case, $t_1 \cup t_2$ is not a successful move on SM.*

Proof 5 *We prove this property by offering a counter-example, illustrated in Figure 5.3. Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA, with $A = \{a, b, c, d, e, f, g, h\}$, $R = \{(b, c), (b, h), (c, d), (f, g), (f, h), (g, d), (h, c), (h, g)\}$, $R^+ = \{(a, b), (e, f)\}$ and $R^+ = \{\}$. Also, let $\neg Comp_\forall(d, AS)$ be the goal we want to achieve (drop d from the grounded extension).*

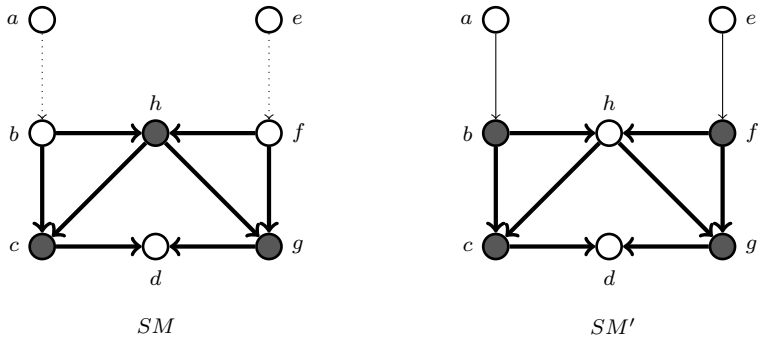


Figure 5.3: While $\{(ab, +, \#)\}$ and $\{(ef, +, \#)\}$ are both target sets for rejecting d under grounded semantics, if they are both played on SM they will not lead to the rejection of d. Arguments in the grounded extension are shown in white, arguments out of the grounded extension are shown in black.

In order to drop argument d from the grounded extension, there are two target sets: $\mathbb{T}^{Comp}_\forall = \{\{(ab, +, \#)\}, \{(ef, +, \#)\}\}$. But their union $m = \{(ab, +, \#)\} \cup \{(ef, +, \#)\}$ is not a successful move since, in the system $SM' = \Delta(SM, m)$, the argument d remains in the grounded extension, as it is defended by h.

Having presented the relations among sets of successful moves and sets of target sets, we will now focus on the content of target sets. We shall identify atoms which, under some conditions, cannot be parts of specific target sets.

5.2.2 Some properties on the content of target sets

The focus now is on the content of target sets, mainly on identifying atoms which cannot be part of specific target sets. This will give us an indication on how to compute target sets, which will be our next task.

For the following three properties we consider a slightly broader type of goal. Up until now, a positive goal was to have a single argument belong in some (all) extension(s). Here, we consider the goal of having a *set of arguments* $X \subseteq A$ belong in some (all) extension(s). The next three properties describe the possible content of such target sets and, more specifically, they indicate which atoms can never be included in them.

The first property says that if the goal is to make X become subset of some admissible (resp. of the grounded) extension, then there is no target set containing the atom $(ad, +, \#)$ with $d \in X$.

Property 6 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA and let $X \subseteq A$, with $d \in X$. Also, let $(a, d) \in R^+$. If t is a target set for the goal of making X subset of some admissible (resp. of the grounded) extension, then it holds that $(ad, +, \#) \notin t$.*

Proof 6 *By contradiction, assume that there exists a target set t for the goal of making X subset of some admissible (resp. of the grounded) extension, and that $(ad, +, \#) \in t$. Let us denote that admissible (resp. grounded) extension ext. So, in $\Delta(SM, t)$, it holds that $X \subseteq ext$. But, the move $t \setminus \{(ad, +, \#)\}$ on SM is also successful for the above goal, because in $\Delta(SM, t \setminus \{(ad, +, \#)\})$, it also holds that $X \subseteq ext$. So, t is not minimal among the successful moves for the above goal, and thus t is not a target set.*

The next property says that if the goal is to make X become subset of some admissible (resp. of the grounded) extension, then for every target set which does not contain atom $(ad, -, \#)$ with $d \in X$, that target set does not contain atom $(ba, -, \#)$.

Property 7 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA and let $X \subseteq A$, with $d \in X$. Also, let (a, d), $(b, a) \in R$. If t is a target set for the goal of making X subset of some admissible (resp. of the grounded) extension and $(ad, -, \#) \notin t$, then it holds that $(ba, -, \#) \notin t$.*

Proof 7 *By contradiction, assume that there is a target set t for the goal of making X subset of some admissible (resp. of the grounded) extension, and that both $(ad, -, \#) \notin t$ and $(ba, -, \#) \in t$ hold. Then, in $\Delta(SM, t)$, X is subset of some admissible (resp. of the grounded) extension. That extension contains d, therefore it also contains an attacker of a (that attacker is different than b). But, move $t \setminus \{(ba, -, \#)\}$ is also successful on SM, because in $\Delta(SM, t \setminus \{(ba, -, \#)\})$, X is subset of the same admissible (resp. of the same grounded) extension, as before. So, t is not minimal among the successful moves for the above goal, and thus t is not a target set.*

The next property says that if the goal is to make X subset of some admissible (resp. of the grounded) extension, then a target set which does not contain $(ad, -, \#)$ with $d \in X$, cannot contain both $(ba, +, \#)$ and $(ca, +, \#)$.

Property 8 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA and let $X \subseteq A$, with $d \in X$. Also, let (a, d), (b, a), $(c, a) \in R$. If t is a target set for the goal of making X subset of some admissible (resp. of the grounded) extension and $(ad, -, \#) \notin t$, then it we cannot have both $(ba, +, \#) \in t$ and $(ca, +, \#) \in t$.*

Proof 8 *By contradiction, assume that there exists a target set t for the goal of making X subset of some admissible (resp. of the grounded) extension, such that $(ad, -, \#) \notin t$. Also, assume that both $(ba, +, \#) \in t$ and $(ca, +, \#) \in t$ hold. Then, in $\Delta(SM, t)$, for an admissible (resp. the grounded) extension, denoted ext, it holds that $X \subseteq ext$. There are two cases, either $b \in ext$ or $b \notin ext$: (1) If $b \in ext$, then the move $t \setminus \{(ca, +, \#)\}$ is successful on SM, because in $\Delta(SM, t \setminus \{(ca, +, \#)\})$, ext*

is also an admissible (resp. the grounded) extension. (2) If $b \notin ext$, then the move $t \setminus \{(ba, +, \#)\}$ is successful on SM, because in $\Delta(SM, t \setminus \{(ba, +, \#)\})$, ext is also an admissible (resp. the grounded) extension. So, t is not minimal among the successful moves for the above goal, and thus t is not a target set.

Now we turn our attention to properties considering negative goals. In the two following properties, our focus is once again a single argument d (and not a set of arguments X, as in the three previous properties).

The first property states that if we have the goal of rejecting argument d (under grounded semantics, or under admissible semantics with credulous acceptability), then no target set contains the removal of an attack against d.

Property 9 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA, where $d \in A$ is the issue. Also, let $(a, d) \in R^-$. If $t \in \mathbb{T}^{Comp}_{\neg\forall}$ (resp. $t \in \mathbb{T}^{S}_{\neg\exists}$), then it holds that $(ad, -, \#) \notin t$.*

Proof 9 *By contradiction, assume that there exists a target set $t \in \mathbb{T}^{Comp}_{\neg\forall}$ (resp. $t \in \mathbb{T}^{S}_{\neg\exists}$), such that $(ad, -, \#) \in t$. Then, in $\Delta(SM, t)$, the grounded extension does not contain d (resp. no admissible extension contains d). But, move $t \setminus \{(ad, -, \#)\}$ is also successful on SM, because in $\Delta(SM, t \setminus \{(ad, -, \#)\})$, the grounded extension does not contain d (resp. no admissible extension contains d). So, t is not minimal among the successful moves for the above goal, and thus t is not a target set.*

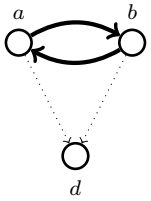
The second property states that if we have the goal of rejecting argument d under grounded semantics, then no target set contains the addition of two different attacks against d.

Property 10 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA, where $d \in A$ is the issue. Also, let $(a, d), (b, d) \in R^+$. If $t \in \mathbb{T}^{Comp}_{\neg\forall}$, then we cannot have both $(ad, +, \#) \in t$ and $(bd, +, \#) \in t$.*

Proof 10 *By contradiction, assume that there exists a target set $t \in \mathbb{T}^{Comp}_{\neg\forall}$ such that both $(ad, +, \#) \in t$ and $(bd, +, \#) \in t$ hold. Then, the grounded extension of $\Delta(SM, t)$ does not contain d and it may, or may not, contain argument a. If it contains a, then move $t \setminus \{(bd, +, \#)\}$ would have been successful on SM for this goal. Else, if it does not contain a, then move $t \setminus \{(ad, +, \#)\}$ would have been successful on SM for this goal. So, t is not minimal among the successful moves for the above goal, and thus t is not a target set.*

Notice that the last property would not hold if, instead of $\mathbb{T}^{Comp}_{\neg\forall}$, we had considered $\mathbb{T}^{S}_{\neg\exists}$, as illustrated in the following example:

Example 18 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA such that $A = \{a, b, d\}$, $R = \{(a, b), (b, a)\}$, $R^+ = \{(a, d), (b, d)\}$, $R^- = \{\}$, and let $d \in A$ be the issue.*



We can verify that $\mathbb{T}^{Comp}_{\neg\forall} = \{\{(ad, +, \#)\}, \{(bd, +, \#)\}\}$. Indeed, from the last property we already knew that no target set contains two attacks against the issue. On the other hand, $\mathbb{T}^{S}_{\neg\exists} = \{\{(ad, +, \#), (bd, +, \#)\}\}$. Therefore, we see that a target set in $\mathbb{T}^{S}_{\neg\exists}$ may indeed contain the addition of two attacks against d. Thus property 10 would not hold for $\mathbb{T}^{S}_{\neg\exists}$.

Using these properties which highlight the (im)possible content of some target sets, we will later define a rewriting procedure which takes as input an ASMA $\langle A, R, R^+, R^- \rangle$ and an issue $d \in A$, and computes all the sets of target sets for different types of argumentative goals.

5.2.3 Modifiable arguments: do they increase expressiveness?

A dynamic system could also model addable and removable arguments, as well as attacks. Since ASMAs do not do this, but they consider fixed sets of arguments, it is natural to ask if this fact restricts their expressiveness. It seems that, from a practical point of view, being able to add and remove arguments, can be in many cases a nice feature. Here we provide some initial thoughts on this issue.

In the setting of a debate, voting on arguments is very similar to voting on attacks, in the following sense: a vote is the expression of a positive or negative opinion which is not backed-up by any reasons (as opposed to the action of stating a specific attack). The following are some possible underlying reasons which, in practice, may lead an agent to express his negative opinion on an argument:

- The argument is *a priori* considered as not convincing.
- The argument is not related to the debate.
- The argument violates some rule of the debate.
- The argument is badly formed (e.g. it makes no sense).

Practically, being able to vote (on either attacks or arguments) is very convenient, because it is an efficient way to shorten some debates. For example, assume that a majority of people in a debate consider that some argument is offensive or inappropriate. Then, they may simply vote against it, and try to get it deleted, without stating any explicit attacks against it. So, introducing modifiable arguments could be a convenient addition to an ASMA.

While this is true, we argue that any ASMA can be modified to simulate argument additions and removals. Let us provide a small example.

Example 19 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA such that $A = \{a, b, c, d\}$, $R = \{(a, d), (b, d)\}$, $R^+ = \{(c, d)\}$ and $R^- = \{(b, d)\}$. The system SM is illustrated in Figure 5.4.*

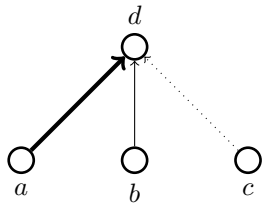


Figure 5.4: Argument d is rejected under all common semantics, and we cannot reinstate it, as the attack (a, d) is fixed.

Now, let there be an agent who wants to reinstate d. The attacks (b, d) and (c, d) are modifiable, but unfortunately for that agent, the attack (a, d) is fixed, so he can do nothing about it. The agent may express his frustration, by saying that, argument a should be considered modifiable (removable), and in that case he would try to remove it and reinstate d.

It seems that the system SM, by not allowing modifiable arguments, restricts expressiveness in this scenario. But is there anything that can be done about this? We argue that there is. The initial system could be expanded with an additional argument e, not attacked by any other argument, and by an addable attack (e, a). This is illustrated in Figure 5.5.

Therefore, practically, removing (resp. adding) argument a could be easily simulated by adding (resp. removing) the attack (e, a). Notice that if the attack (e, a) is on the system, then every common acceptability semantics will consider that argument a is rejected. Thus, in this expanded ASMA, it is possible to reinstate argument d.

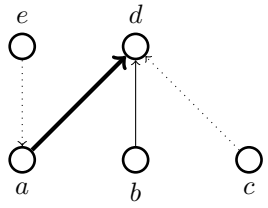


Figure 5.5: The system SM, after the addition of the argument e and the modifiable attack (e, a). Removing (resp. adding) argument a is simulated by adding (resp. removing) the attack (e, a).

5.3 A Rewriting Procedure for Target Set Computation

In this section we provide a set of rewriting rules which help us compute, for a given ASMA, all the target sets for some types of goals. In order to do this, we use the Maude[4] system [CDE+99] which is based on rewriting logic. This section is arranged as follows: we start by explaining what Maude is and why it is useful for the type of computations we want to make. Then, we analyze the core component of our program, which consists in its set of rules. Afterwards, we explain the rewriting procedure that Maude will undertake, in the context of our program. Finally, we prove some important properties of the output of this rewriting procedure.

5.3.1 The Maude system and the intuition behind our program

Maude is both a declarative programming language and a system. It is based on rewriting logic and it can model systems and the actions within those systems. Maude is a high-level, expressive language, which can model from biological systems to programming languages, including itself. A program in Maude is a logical theory, and a computation made by that program is logical deduction using the axioms of that theory.

A Maude program has some basic building blocks, which in our case are *atoms* as in Definition 36. With the help of the connectors $\wedge$ and $\vee$, atoms can be linked, in order to form conjuncts and formulas.

Definition 40
$Conjunct ::= Atom \mid (Conjunct \wedge Conjunct)$;
$Formula ::= Conjunct \mid (Formula \vee Formula)$
Let $Conjuncts$ denote the set of all possible conjuncts, and let $Formulas$ denote the set of all possible formulas. A conjunct which contains at least one open atom is called ***open conjunct****. Otherwise, it is called* ***closed conjunct****. A formula which contains at least one open conjunct is called* ***open formula****. Otherwise, it is called* ***closed formula****.*

Our Maude program, presented in Appendix B, is given as input a conjunct which describes an ASMA $\langle A, R, R^+, R^- \rangle$ and also contains either the atom $PRO(d)$ or the atom $CON(d)$, with $d \in A$. If we want to ensure the positive (resp. negative) goal of accepting (resp. rejecting) argument d under some semantics, then we start by including atom $PRO(d)$ (resp. $CON(d)$) in the initial conjunct. Maude starts from one of these two atoms and, based on a set of rewriting rules and equations, rewrites the initial formula (conjunct), thus producing new formulas, which are, in turn, rewritten. The system stops when all the computed formulas are non-rewritable. We will see that every conjunct of the final (output) formula corresponds to a move on the initial system M. The repercussions of these moves on the status of d are detailed in Property 12.

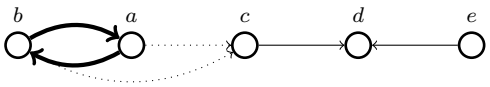
[4] http://maude.cs.uiuc.edu

5.3.2 Program structure - Rewriting Rules

First, let us present the building blocks of our program, and explain how an ASMA can be represented. The different types of entities we can define in a Maude program are called *sorts*. We have defined five types of sorts: *Attack*, *Argument*, *Sign*, *Atom* and *Conjunct*. The only argument we use in our program is the issue, which is denoted d, and is declared as a constant. Attacks are defined by using the following conventions. Firstly, the name of an attack always starts by the character '". Also, if attack $x \in R^+$ (resp. $x \in R^-$), then its name starts with '+' (resp. '−'). The structure of an argumentation system (which is passed as input to our program) can be represented by declaring the relations among its attacks (and the issue d). This is done with the help of two binary operators, *hitsArg* and *hits*. For example, the term *'-cd hitsArg d* is an atom, and it means that the attack *-cd* is set against the issue d, while the term *'+bc hits '-cd* is also an atom, and it means that $hits(+bc, -cd)$. The unary operator *isNotHit* is used in order to indicate attacks which are not hit by other attacks (thus they are certainly successful). For example, the term *isNotHit '-ed* is an atom which means there is no attack against the attack *-ed*. In addition to the above types of atoms, there are of course the atoms of Definition 36, which are triplets *(Attack, Sign, Sign)*. Lastly, the binary *and* operator corresponds to the $\wedge$ sign, and it is used in order to define a conjunct, given two smaller ones. This operator is declared as associative and commutative, which we will see that makes the use of expansion and simplification rules very easy.

This way, it is possible to represent (in the form of a single conjunct) any arbitrary ASMA, provided that for every argument in the system there exists an "attack path" towards argument d. Let us now provide an example.

Example 13 (cont.)



We are able to represent the above ASMA, in the form of the following conjunct:

```
('-cd hitsArg d) and ('-ed hitsArg d) and ('+bc hits '-cd) and ('+ac hits '-cd) and
('ba hits '+ac) and ('ba hits 'ab) and ('ab hits '+bc) and ('ab hits 'ba) and
(isNotHit '-ed)
```

We proceed to the analysis of the program's core, which is its set of rules. There exist two types of rules: **Atom expansions**, or **rewriting rules**, indicated by '=>', and **atom simplifications**, or **equations**, indicated by '='. In our program, an atom expansion replaces two atoms appearing in an open conjunct by some other atoms, whereas an atom simplification replaces two atoms found in the same conjunct by a single atom. Roughly, the firing of an expansion rule is potentially followed by a number of atom simplifications, before the next expansion rule is fired.

Let us briefly explain the intuition behind the expansion rules. Depending on whether we want to accept or reject the issue, we start from it and we navigate the attacks backwards, while adding and removing attacks, trying to enforce the status of the attacks relevant to the issue. When there exist more than one choices to achieve our goal, we explore all the possibilities (combinations of attack additions and removals). Very roughly, if at some point of the computation, the left side of an expansion rule appears, Maude may replace it with the right side of that rule. The same principle holds for equations. So, when the initial goal is $PRO(d)$, we want to see the issue d accepted. To do so, we have to consider each attack against d, one at the time, and either remove it (if it belongs to R^-), or make it '0' by making an attack which attacks it become '1'. On the other hand, when the initial goal is $CON(d)$, so we want d to be rejected, we have to make one attack against d become '1' or '?' (and add this attack, in case it was originally in R^+).

The rules can be found in Appendix B.

Rules 1-3 say that if an attack is '1', then for every attack against it, either that attack is '0'

(rule 1), or it is removed (rule 2), or (if it belongs to R^+) we introduce an atom $(x, 0, \#)$ which will lead to a simplification if we later add this attack (rule 3), thus this atom ensures that the attack can never become successful. Rules 4-5 say that if an attack is '0', then there exists an attack against it which is '1'. That attack is either already in the system (rule 4), or it is added to it (rule 5). Rules 6-12 say that if an attack is '?', then two things hold: first, there exists at least one attack set against it which is also '?' (rules 6 and 7).[5] Also, the rest of the attacks set against it are either '?', or '0', or removed (rules 8-10), or (if they belong to R^+) we introduce $(x, 0, \#)$ and $(x, ?, \#)$, which will lead to simplifications if we later add these attacks and try to make them '1' (rules 11-12). Rules 13-15 say that in the PRO case every attack against the issue is either '0' (rule 13), is removed (rule 14), or (if it belongs to R^+) we introduce an atom $(x, 0, \#)$ for the same reason as in rule 3 (rule 15). Rules 16-19, finally, say that in the CON case there exists one attack against the issue which is either '1' (rules 16 and 17) or '?' (rules 18 and 19).

Now, as far as the simplification rules (equations) are concerned: Equation 1 says that if two identical atoms appear in the same conjunct, then one of them is deleted. Equation 2 performs a simplification related to the '?' status of an attack. Equation 3 says that if an open atom and a closed atom (which are otherwise identical) appear in the same conjunct, then the open atom is deleted. Equations 4-6 say that if two atoms referring to the same attack, but indicating different status, appear in the same conjunct, then $\perp$ is introduced. Equations 7-8 say that if an attack which cannot be attacked is set to be '?' or '0', then $\perp$ is introduced. Equations 9-10 are applied in case there exist no potential attacks against d. Equation 11, finally, says that the atom $\perp$ once it appears in a conjunct, it reduces that conjunct into $\perp$.

5.3.3 The Rewriting Procedure (RP)

Now we explain how Maude's rewriting procedure works, by focusing on the specific case of our program. As we said above, the program's input is an argumentation system AS in the form of a conjunct, with the addition of either atom $PRO(d)$ or atom $CON(d)$. This initial conjunct can be seen as the root of a tree which will be gradually expanded. The rewriting procedure starts from the root, and specifically from atom $PRO(d)$ or $CON(d)$. All the applicable expansion rules are then considered, one-by-one. For every applicable expansion rule, that rule is applied, and a set of new conjuncts is computed. In every new conjunct, simplification rules are applied repeatedly, until no more simplification rules are applicable. The obtained (simplified) conjuncts are the child nodes of the root. Once such an "expansion-simplification" step is finished, all the conjuncts computed in the previous step are considered (one by one) and there follows another "expansion-simplification" step. Therefore, the tree is gradually expanded, in a Breadth-First-Search way. These expansion steps are repeated until, at some point, there are no conjuncts which can be further expanded. Finally, from every non-expandable conjunct computed (leaf of the tree), just the $(x, +, \#)$ and $(x, -, \#)$ atoms are filtered. The formal definition of Maude's rewriting procedure, in the context of our program, is given in Algorithm 1.

We shall denote the set of returned moves $\mathcal{M}_d^{PRO}$ if $initF = PRO(d)$, and $\mathcal{M}_d^{CON}$ if $initF = CON(d)$.

Now we show how the input conjunct is passed to the Maude system. We assume that we wish to accept argument d, so we introduce the atom $PRO(d)$.

Example 13 (cont.) *The input to the Maude system in this example is the following:*
> **search** `PRO(d) and ('-cd hitsArg d) and ('-ed hitsArg d) and ('+bc hits '-cd) and ('+ac hits '-cd) and ('ba hits '+ac) and ('ba hits 'ab) and ('ab hits '+bc) and ('ab hits 'ba) and (isNotHit '-ed)` =>! **C:Conjunct .**

The "search" keyword tells Maude that whenever more than one rewriting rules are applicable, it

[5] The importance of atom $(x, ?, **)$ can be seen in rules 6 and 7: if we had atom $(x, ?, *)$ (in the place of $(x, ?, **)$), then there would exist a possible rewriting which would make *all* the attacks against x become '0'. But that would be a problem, since we want at least one attack against x to become '?'. This is ensured by using atom $(x, ?, **)$.

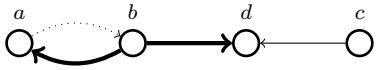
Data: An ASMA $\langle A, R, R^+, R^- \rangle$, a formula $initF = PRO(d)$ or $initF = CON(d)$, with $d \in A$, a set of expansion rules, a set of simplification rules.
Result: A set of moves $\mathcal{M}_d$.
Initialise formula $currF := initF$;
while *currF has an expandable conjunct* **do**
  Let Exp denote the set of all the expandable conjuncts of $currF$;
  foreach *conjunct* $C \in Exp$ **do**
    Initialise the set of conjuncts $repl_C := \{\}$;
    foreach *applicable rewriting rule rl on* C **do**
      if *rule rl applied on* C *gives* C' **then**
        while *a simplification can be applied on* C' **do**
          Choose such a simplification, and apply it on C' ;
        end
      end
      Add C' into the set $repl_C$;
    end
    Replace C with $C'_1 \vee C'_2 \vee \cdots \vee C'_m$ in $currF$, s.t. $\forall i \in [1 \dots m]$, $C'_i \in repl_C$;
  end
end
Initialise the set of moves $\mathcal{M}_d := \{\}$;
foreach *conjunct* C *of* $currF$ **do**
  if $C \neq \perp$ **then**
    $m := \{(x, s, \#) \mid (x, s, \#)$ appears in C, and $s \in \{+, -\}\}$; Add m into the set $\mathcal{M}_d$;
  end
end
return $\mathcal{M}_d$;

Algorithm 1: Maude's rewriting procedure, in the context of our program

must consider them all, one at a time, in a Breadth-First-Search way. This is essential in order to find all the possible rewritings, as we have already mentioned. Also, by using `=>! C:Conjunct`*, we tell Maude to continue the rewritings, until the obtained terms are non-rewritable conjuncts. Once the computation finishes, we obtain four conjuncts. The three first ones correspond to moves on the system M, while the fourth one is* $\perp$*. The corresponding moves of the three first conjuncts are:* $\mathcal{M}_d^{PRO} = \{\{(ed, -, \#), (cd, -, \#)\}, \{(ed, -, \#), (bc, +, \#)\}, \{(ed, -, \#), (ac, +, \#)\}\}$.

Let us show the details of the computation made by Maude in a smaller example:

Example 20 *Let* $SM = \langle A, R, R^+, R^- \rangle$ *be an ASMA with* $A = \{a, b, c, d\}$, $R = \{(b, a), (b, d), (c, d)\}$, $R^+ = \{(a, b)\}$ *and* $R^- = \{(c, d)\}$*. As above, let d be the argument we focus on, which we want to see accepted.*



We will denote applications of rewriting rules with arrows over which the applied expansion or simplification rule is marked. When a rewriting or simplification rule is applied on conjunct A and produces conjunct B, the atoms appearing on the rule's left side are underlined in conjunct A, while the atoms appearing on the rule's right side are in bold in conjunct B. The rewritings made by Maude, given the initial conjunct, are as follows:

`PRO(d) and ('+ab hits 'ba) and ('+ab hits 'bd) and ('ba hits '+ab) and ('bd hitsArg d) and ('-cd hitsArg d) and (isNotHit '-cd)`

$\xrightarrow{expansionRules13,14}$

PRO(d) and **('-cd,0,*)** and ('+ab hits 'ba) and ('+ab hits 'bd) and ('ba hits '+ab) and ('bd hitsArg d) and (isNotHit '-cd) $\vee$
PRO(d) and **('-cd,-,#)** and ('+ab hits 'ba) and ('+ab hits 'bd) and ('ba hits '+ab) and ('bd hitsArg d) and (isNotHit '-cd)

$\xrightarrow{simplificationRule7}$

$\bot \vee$
PRO(d) and ('-cd,-,#) and ('+ab hits 'ba) and ('+ab hits 'bd) and ('ba hits '+ab) and ('bd hitsArg d) and (isNotHit '-cd)

$\xrightarrow{expansionRule13}$

$\bot \vee$
PRO(d) and **('bd,0,*)** and ('-cd,-,#) and ('+ab hits 'ba) and ('+ab hits 'bd) and ('ba hits '+ab) and (isNotHit '-cd)

$\xrightarrow{expansionRule5}$

$\bot \vee$
PRO(d) and **('bd,0,#)** and **('+ab,+,#)** and **('+ab,1,*)** and ('-cd,-,#) and ('+ab hits 'ba) and ('ba hits '+ab) and (isNotHit '-cd)

$\xrightarrow{expansionRule1}$

$\bot \vee$
PRO(d) and ('bd,0,#) and ('+ab,+,#) and **('+ab,1,#)** and **('ba,0,*)** and ('-cd,-,#) and ('+ab hits 'ba) and (isNotHit '-cd)

$\xrightarrow{expansionRule5}$

$\bot \vee$
PRO(d) and ('bd,0,#) and ('+ab,+,#) and ('+ab,1,#) and **('ba,0,#)** and **('+ab,+,#)** and **('+ab,1,*)** and ('-cd,-,#) and (isNotHit '-cd)

$\xrightarrow{simplificationRule1}$

$\bot \vee$
PRO(d) and ('bd,0,#) and **('+ab,+,#)** and ('+ab,1,#) and **('ba,0,#)** and **('+ab,1,*)** and ('-cd,-,#) and (isNotHit '-cd)

$\xrightarrow{expansionRule3}$

$\bot \vee$
PRO(d) and ('bd,0,#) and ('+ab,+,#) and **('+ab,1,#)** and ('ba,0,#) and ('-cd,-,#) and (isNotHit '-cd)

Therefore, we have $\mathcal{M}_d^{PRO} = \{\{(ab,+,\#),(cd,-,\#)\}\}$. *Finally, notice that we have just provided only a subtree of the tree actually computed by Maude. Nonetheless, that subtree "loses" no move. For example, in the initial conjunct, once Maude considers the two possibilities for rewriting* PRO(d) and ('-cd hitsArg d) *(as shown above), it will backtrack and it will also consider the two possibilities for rewriting* PRO(d) and ('bd hitsArg d). *So, the root will actually have four child nodes, resulting from the four possible applications of expansion rules (and not only from two, as we did above). It can be verified that the complete tree computed by Maude will have no leaves which correspond to additional (different) moves.*

Properties of the RP procedure

Now, let us highlight some important properties of the RP procedure. First, we prove that RP always terminates, provided a finite number of arguments in the system we refer to.

Property 11 *Termination of RP*
Let $SM = \langle A, R, R^+, R^- \rangle$ *be an ASMA, where* $|A|$ *is finite. Then, the RP procedure always terminates.*

Proof 11 *The RP procedure starts with an initial conjunct initF containing* $PRO(d)$ *or* $CON(d)$. *For every applicable expansion rule on initF, that rule is applied on initF and a new conjunct*

is computed. These conjuncts can be seen as the children of the root node initF. By repeatedly applying expansion rules on conjuncts-nodes of the tree, new nodes are computed. In order to prove that RP always terminates, we must prove that the number of nodes of such a tree is always finite. First, from the set of rewriting rules, it follows that every conjunct computed by RP has a finite number of atoms. Moreover, there is a finite number of applicable rules on every conjunct, so the branching factor of the tree is finite. Finally, we must prove that the depth of the tree is finite. From the set of rewriting rules, it follows that a conjunct is expandable (not a leaf), if it contains an open atom and an atom of the form `(x hits y)`*, or* `(x hitsArg d)`.[6] *Notice that initF contains a finite number of* `(x hits y)` *and* `(x hitsArg d)` *atoms, because the number of arguments and attacks of the system SM is finite. Also, after the application of an expansion rule on some conjunct, the newly created conjunct contains one less* `(x hits y)` *atom, or one less* `(x hitsArg d)` *atom, than its parent-node. As a result, the depth of the tree cannot be greater than the total number of* `(x hits y)` *and* `(x hitsArg d)` *atoms in initF, which is finite. So, we have proved that the RP procedure always terminates.*

We underline that the "search" keyword ensures that, after a node-conjunct of the tree has been simplified, Maude tries, one-by-one, every applicable rewriting rule for that node-conjunct. An internal strategy is used in order to decide the order of application of these rules, but changing that order cannot affect the results of the computation.

We now analyze the output of the rewriting procedure with respect to the different argumentative goals. We shall say that:

- the procedure is *correct* for *successful moves* (resp. *target sets*) for goal g if every move it returns is successful (resp. a target set) for g,
- the procedure is *complete* for *successful moves* (resp. *target sets*) for goal g if it returns *all* the successful moves (resp. the target sets) for g.

As illustrated in Figure 5.6, correctness for target sets is not satisfied: the procedure will return, in the general case, some moves which are not target sets for any of the considered semantics. But in some cases (shown below) correctness for successful moves is ensured. On the contrary, completeness for successful moves is not satisfied: In the general case, RP does not compute all the successful moves for any semantics. However, in some cases (shown below) completeness for target sets is ensured. Of course, the most interesting lines of the following table are those for which we have "Yes" in both columns: RP returns only successful moves, and it returns all the target sets.

Property 12 *The following table illustrates for which goals the rewriting procedure is correct for successful moves and/or complete for target sets.*

Goal	Correctness for successful moves	Completeness for target sets
$S_\exists(d)$	*Yes*	*Yes*
$Pref_\forall(d)$	*No*	*No*
$Comp_\forall(d)$	*No*	*Yes*
$\neg S_\exists(d)$	*No*	*No*
$\neg Pref_\forall(d)$	*No*	*?*
$\neg Comp_\forall(d)$	*Yes*	*Yes*

Completeness for target sets, regarding the goal $\neg Pref_\forall(d)$, is left open so far. However, for the sake of readability, we draw Figure 5.6 assuming that the answer is "Yes".

[6] This means that it is quite possible for an open atom to be non-expandable. This is the case when no relevant `(x hits y)` or `(x hitsArg d)` atom is found in the same conjunct as that open atom, in order to fire an expansion rule.

Proof 12 Correctness of RP for $S_{\exists}(d)$: *In other words, we want to prove that $\mathcal{M}_d^{PRO} \subseteq \mathbb{M}_{\exists}^S$. Let $m \in \mathcal{M}_d^{PRO}$. We will prove that $m \in \mathbb{M}_{\exists}^S$. The move m corresponds to some conjunct, denoted c_m, computed by RP. From c_m we construct the set of arguments $D = \{x \mid (xy, 1, \#)$ is an atom of $c_m\}$. We will now prove that in $\Delta(SM, m) = \langle A, R_m, R_m^+, R_m^- \rangle$, it holds that D is an admissible set of arguments which defends argument d. First, let us assume that in $\Delta(SM, m)$ the set D is not conflict-free. In that case there exist two arguments $x_1, x_2 \in D$, such that $x_1x_2 \in R_m$. Now, $x_1, x_2 \in D$ implies that $\exists x_3, x_4 \in A$ such that $(x_1x_3, 1, \#)$ and $(x_2x_4, 1, \#)$ are atoms of c_m. Given that $(x_2x_4, 1, \#)$ appears in c_m, and that $x_1x_2 \in R_m$, it follows that atom $(x_1x_2, 0, \#)$ must also appear in c_m (from expansion rule 1). In turn, this means that $\exists x_5 \in A$ such that $(x_5x_1, 1, \#)$ also appears in c_m (from expansion rules 4,5). Similarly, given that $(x_1x_3, 1, \#)$ appears in c_m, it holds that $(x_5x_1, 0, \#)$ also appears in c_m. But, it is impossible for both $(x_5x_1, 1, \#)$ and $(x_5x_1, 0, \#)$ to appear in the same conjunct (as simplification rule 4 would have simplified them into $\bot$). Therefore, we have proved that D is conflict-free. Second, let us assume that in the system $\Delta(SM, m)$, the set D does not defend all its elements. In that case $\exists x_1 \in D$ and $\exists x_2 \notin D$ such that $x_2x_1 \in R_m$, and no argument of D attacks x_2. $x_1 \in D$ implies that $\exists x_0 \in A$ such that atom $(x_1x_0, 1, \#)$ appears in c_m. So, it follows that atom $(x_2x_1, 0, \#)$ also appears in c_m (from expansion rule 1), and as a result, $\exists x_3 \in A$ such that atom $(x_3x_2, 1, \#)$ also appears in c_m. By definition of the set D, notice that $x_3 \in D$. Impossible, since we assumed that no argument of D attacks x_2 in $\Delta(SM, m)$. Therefore, we have proved that D defends all its elements. Given that D is conflict-free and it defends all its elements, it follows that D is an admissible set of arguments. Finally, since for every attack $xd \in R_m$ against the issue d, it holds that atom $(xd, 0, \#)$ appears in c_m (because of expansion rule 13), it holds that argument d is defended by the set D. From this, and from the fact that D is admissible in $\Delta(SM, m)$, it follows that $D \cup \{d\}$ is admissible in $\Delta(SM, m)$. Thus, $m \in \mathbb{M}_{\exists}^S$ and we have proved that $\mathcal{M}_d^{PRO} \subseteq \mathbb{M}_{\exists}^S$.*

Correctness of RP for $\neg Comp_{\forall}(d)$: *In other words, we want to prove that $\mathcal{M}_d^{CON} \subseteq \mathbb{M}_{\neg\forall}^{Comp}$. Let $m \in \mathcal{M}_d^{CON}$. We will prove that $m \in \mathbb{M}_{\neg\forall}^{Comp}$. The move m corresponds to some conjunct, denoted c_m, computed by RP. Since the root-conjunct contains atom $CON(d)$, and given the expansion rules 16-19 for $CON(d)$, it follows that c_m contains either an atom $(xd, 1, \#)$, or an atom $(xd, ?, \#)$. In the first case, if c_m contains an atom $(xd, 1, \#)$, then as shown in the previous proof, it holds that in the system $\Delta(SM, m)$ there exists an admissible extension which contains x. Therefore, there also exists a complete extension which contains x. That complete extension does not contain argument d, as it is attacked by x. As a result, d does not belong to the grounded extension of $\Delta(SM, m)$, in other words $m \in \mathbb{M}_{\neg\forall}^{Comp}$. In the second case, if c_m contains an atom $(xd, ?, \#)$, then we show that the graph of $\Delta(SM, m)$ has a subgraph consisting of the following two elements: A cycle of attacks and a path of attacks coming from that cycle and leading to argument d. This follows from expansion rules 6 and 7, which state that when a node-conjunct which contains an atom $(xy, ?, **)$ is expanded, all its children will contain some atom $(zy, ?, **)$. The only way to continue the expansions and obtain leaf nodes which are not simplified into $\bot$, is to "navigate" the graph until a cycle of '?' attacks is formed. Additionally, from expansion rules 8-12, it follows that every attack against an argument of that cycle, or against the path connecting the cycle to d, is set to be either '?' or '0'. In this case, in $\Delta(SM, m)$ there exists no argument of the grounded extension which attacks an argument of that cycle, or of the path connecting the cycle to d.* [7] *The fact that argument d is "connected" to such a cycle, implies that d does not belong to the grounded extension of $\Delta(SM, m)$. Therefore, $m \in \mathbb{M}_{\neg\forall}^{Comp}$. Since we have shown that $m \in \mathbb{M}_{\neg\forall}^{Comp}$ always holds, we have proved that $\mathcal{M}_d^{CON} \subseteq \mathbb{M}_{\neg\forall}^{Comp}$.*

Completeness of RP for $S_{\exists}(d)$ (resp. $Comp_{\forall}(d)$): *In other words, we want to prove that $\mathbb{T}_{\exists}^S \subseteq \mathcal{M}_d^{PRO}$ (resp. $\mathbb{T}_{\forall}^{Comp} \subseteq \mathcal{M}_d^{PRO}$). Let $t \in \mathbb{T}_{\exists}^S$ (resp. $t \in \mathbb{T}_{\forall}^{Comp}$). We prove that RP constructs a tree which has a leaf node containing all the $(x, +, \#)$ and $(x, -, \#)$ atoms appearing in t, and no additional $(x, +, \#)$ or $(x, -, \#)$ atoms. Let $\{x_1d, \ldots, x_nd\}$ be the set of all the attacks against d*

[7] In other words, if we used argument labelling, we would say that attributing the *undec* label to all the arguments of that cycle and of that path, is part of a valid labelling.

in SM. Roughly, t removes a subset of these attacks, and it ensures (by making a further minimal change) that the remaining attacks do not "harm" d. Let $P = \{(x_1d, -, \#), \ldots, (x_nd, -, \#)\}$ be a set of atoms, indicating the removals of attacks in $\{x_1d, \ldots, x_nd\}$. Naturally, t contains a subset of the atoms in P, denoted $P' \subseteq P$. On the other hand, t cannot contain any atoms of the form $(xd, +, \#)$, as stated by Property 6. Moreover, from the expansion rules for $PRO(d)$ it easily follows that the tree has some node n (not a leaf, in the general case) which contains all the atoms of P', and no other $(x, +, \#)$ or $(x, -, \#)$ atoms. Let the subset of attacks against d which are not removed by t be denoted $\{x_kd, \ldots, x_ld\} \subseteq \{x_1d, \ldots, x_nd\}$. Then, according to the expansion rules for $PRO(d)$, node n also contains the atoms $(x_kd, 0, *)$, $\ldots$, $(x_ld, 0, *)$. Thus t contains the atoms in P' and some additional atoms, which result from the expansions of atoms $(x_kd, 0, *)$, $\ldots$, $(x_ld, 0, *)$. We can write $t = P' \cup Q$. Note that every atom of Q refers to an attack necessarily "connected" to some argument of $\{x_k, \ldots, x_l\}$. Let us focus on the attacks against d which are not removed. The arguments attacking d must get attacked back, in order for d to be reinstated. At this point, from Properties 7 and 8, it follows that: (1) It is impossible for any $(yx_i, -, \#)$ atom to appear in t. (2) It is impossible for two atoms $(y_1x_i, +, \#)$ and $(y_2x_i, +, \#)$ to appear in t. As a result, for every argument $x_i \in \{x_k, \ldots, x_l\}$, t can only contain 0 or 1 atoms of the type $(yx_i, +, \#)$. Now we must make sure that RP computes all these possible combinations of attack additions reinstating d. Indeed, after the expansions of all the atoms $(x_kd, 0, *)$, $\ldots$, $(x_ld, 0, *)$ appearing in n, there will appear below node n a number of nodes, and each one of them will contain 0 or 1 atoms of the type $(yx_i, +, \#)$ for every $x_i \in \{x_k, \ldots, x_l\}$. One of these nodes will obligatorily contain exactly the atoms indicating the attack additions against $\{x_k, \ldots, x_l\}$, which are indicated in t. Moreover, if a node contains atom $(yx_i, +, \#)$, then it also contains atom $(yx_i, 1, *)$. So, RP continues to search the graph backwards, considering the indirect attackers (and defenders) of d, using the expansion rules for the $(yx_i, 1, *)$ atoms. Therefore, after a finite number of expansions, the procedure will compute a node which contains exactly the $(x, +, \#)$ and $(x, -, \#)$ atoms found in t. Finally, we must make sure that no simplification rule which introduces $\bot$, can lead to the "loss of a target set". That means that if a node is simplified into $\bot$, we must prove that it would have been impossible for a target set to appear in the subtree below that node. We show that this is indeed the case. Two simplification rules can introduce $\bot$ in a tree whose root contains $PRO(d)$: The first rule is simplification rule 7, which says that if node n contains $(xy, 0, *)$, and there is no potential attacker of x in the system, then $\bot$ is introduced. Having $(xy, 0, *)$ in n means that all the target sets found in the subtree below n must lead to a modified system where there is an attack set against x. Since x has no potential attackers, this is impossible, therefore $\bot$ can be introduced without any loss of target sets. The second rule is simplification rule 4, which says that if node n contains both $(xy, 0, s)$ and $(xy, 1, s)$, then $\bot$ is introduced. Let us see why we can never "lose" a target set by this type of simplification. Let n be a node containing both $(xy, 0, s)$ and $(xy, 1, s)$. Every eventual target set found in the subtree below n leads to a modified system in which some admissible extension (resp. the grounded extension): (i) attacks the argument x (because of $(xy, 0, s)$), and (ii) contains argument x (because of $(xy, 1, s)$). This is impossible, as every admissible extension (resp. the grounded extension) is conflict-free. Therefore, no target set could be found in that subtree, thus we can introduce $\bot$ without losing any target sets.

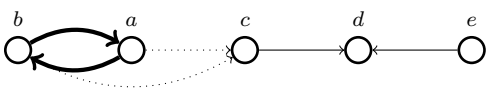
Completeness of RP for $\neg Comp_\forall(d)$: *In other words, we want to prove that $\mathbb{T}^{Comp}_{\neg\forall} \subseteq \mathcal{M}^{CON}_d$. Let $t \in \mathbb{T}^{Comp}_{\neg\forall}$. We proceed as in the previous proof of completeness, and we prove that RP constructs a tree which has a leaf node containing all the $(x, +, \#)$ and $(x, -, \#)$ atoms appearing in t, and no additional $(x, +, \#)$ or $(x, -, \#)$ atoms. Let $\{x_1d, \ldots, x_nd\}$ denote the set of attacks against argument d. Target set t may contain 0 or 1 atoms of the form $(x_id, +, \#)$. This holds because t cannot contain an atom of the form $(x_id, -, \#)$ (from Property 9), and also t cannot contain two atoms of the form $(x_id, +, \#)$ and $(x_jd, +, \#)$ (from Property 10). In accordance to the last remark, notice that the RP procedure expands the root node (which contains $CON(d)$) and creates a number of children. Some of the children contain no $(x_id, +, \#)$ atom, while the rest contain exactly one $(x_id, +, \#)$ atom. Therefore, if t has no $(x_id, +, \#)$ atom, then there exists a child node of the root which also has no $(x_id, +, \#)$ atom. Also, if t has exactly one $(x_id, +, \#)$*

atom, then there exists a child node which also has that same $(x_i d, +, \#)$ atom. Notice that from the expansion rules for $CON(d)$, if node n has the atom $(x_i d, +, \#)$, then obligatorily it also has either the atom $(x_i d, 1, s)$, or the atom $(x_i d, ?, s)$. Also, if n has no $(x_i d, +, \#)$ atom, it still obligatorily has one $(x_i d, 1, s)$ atom, or one $(x_i d, ?, s)$ atom. Now, it holds that the minimal change which drops d from the grounded extension, is the minimal change that either (i) puts an attacking argument of d in some admissible extension, or (ii) creates a cycle of attacks and a path coming from it and "hitting" d, which is not attacked by any argument of the grounded extension [8]*. Now let us consider a child of the root which contains the atom $(x_i d, 1, *)$. As we have already said before, if we take the subtree below that node, then its leaves contain all the target sets that put x in some admissible extension. Therefore, we get all the target sets which achieve condition (i). Now, consider a child of the root which contains the atom $(x_i d, ?, *)$. We prove that, if we take the subtree below that node, then its leaves contain all the target sets achieving condition (ii). As said before, expansion rules 6 and 7 lead to the creation of such a ?-cycle-path (otherwise $\perp$ is introduced at some point). But is minimality guaranteed? The answer is yes, because expansion rules 6 and 7 indicate that for an attack to be '?', it suffices to have one attack against it which is also '?'. The last thing that we must verify (as in the previous proof), is that the simplification and expansion rules, can never lead to the "loss of a target set". We will prove that simplification and expansion rules can never reduce a node's label into $\perp$, unless it was impossible for a target set to appear in the subtree below that node. Here the rewritings start from atom $CON(d)$, so apart from simplification rules 4 and 7, which have been shown not to "lose" any target sets, we must also make sure that the same holds for simplification rules 5, 6 and 8 (which can also introduce $\perp$). It is easy to show that simplification rule 8 leads to no loss of target set, as for an attack to be '?' it must be attacked by at least another one, something impossible if there exists none. Simplification rule 5 (resp. 6) roughly states that if we have an attack which is simultaneously '?' and '0' (resp. '?' and '1'), then we introduce $\perp$. We show that if a target set would appear in the subtree below the simplified node, then the same target set appears in another leaf of the tree, so it is not lost by the simplification. Let us see why this is true. If an atom $(xy, ?, s)$ and an atom $(xy, 0, s)$ (resp. $(xy, 1, s)$) appear in the same conjunct, then at some point before, expansion rule 9 had been applied, introducing an atom $(ca, 0, *)$, while an atom $(ba, ?, *)$ was "already there". From that point on, the system has rewritten these atoms navigating the attacks of the graph backwards. At some point, the atoms $(xy, ?, s)$ and $(xy, 0, s)$ (resp. $(xy, 1, s)$) were introduced in the same conjunct, so the attack xy was set to be simultaneously '?' and '0' (resp. '1'). Notice that any move which would have been computed under a node-conjunct containing $(xy, ?, s)$ and $(xy, 0, s)$ (resp. $(xy, 1, s)$) (if no simplification was taking place), would have also be computed if that node-conjunct contained $(xy, ?, s)$, but not $(xy, 0, s)$. But this possibility is indeed considered by the RP procedure, in another node of the tree. At the point (above) where expansion rule 9 is fired, thus creating a child, there is the alternative of firing expansion rule 8, thus creating another child. In that second child, instead of atoms $(ca, 0, *)$ and $(ba, ?, *)$, we have atoms $(ca, ?, *)$ and $(ba, ?, *)$. Every target set which would have been computed in the subtree below the simplified node, is certain to be computed in the subtree below that node. Therefore, no target set is lost.*

Having finished the proofs regarding the "Yes" cases of the table in Property 12, we turn our attention to the "No" cases, for which we shall provide some counter-examples. Though these desired properties do not hold, the following examples provide an intuition on why this is the case.

First, in the general case, RP is not correct for the goals $Pref_\forall(d)$ and $Comp_\forall(d)$.

Example 13 (cont.)



[8] For convenience, let us call such a subgraph, a "?-cycle-path". Also, notice that if the graph of SM contains odd-length cycles, the condition (i) may not be sufficient by itself.

We have $m = \{(ed, -, \#), (bc, +, \#)\} \in \mathcal{M}_d^{PRO}$*, but notice that* $m \notin \mathbb{M}_{\forall}^{Pref}$*, and also* $m \notin \mathbb{M}_{\forall}^{Comp}$*. Therefore, we see that neither* $\mathcal{M}_d^{PRO} \subseteq \mathbb{M}_{\forall}^{Pref}$*, nor* $\mathcal{M}_d^{PRO} \subseteq \mathbb{M}_{\forall}^{Comp}$ *hold, in the general case. So, RP is not correct for the goals* $Pref_{\forall}(d)$ *and* $Comp_{\forall}(d)$*, in the general case.*

Also, in the general case, RP is not complete for the goal $Pref_{\forall}(d)$.

Example 13 (cont.) *We remind that RP returns the set of moves* $\mathcal{M}_d^{PRO} = \{(ed, -, \#), (cd, -, \#)\}, \{(ed, -, \#), (bc, +, \#)\}, \{(ed, -, \#), (ac, +, \#)\}\}$*. We can verify that, all the target sets for the goals* $S_{\exists}(d)$ *and* $Comp_{\forall}(d)$ *are returned. On the other hand, the move* $\{(ed, -, \#), (ac, +, \#), (bc, +, \#)\}$ *is a target set for the goal* $Pref_{\forall}(d)$*, but it is not returned. Therefore, we see that* $\mathbb{T}_{\forall}^{Pref} \subseteq \mathcal{M}_d^{PRO}$ *does not hold, in the general case. So, RP is not complete for the goal* $Pref_{\forall}(d)$*.*

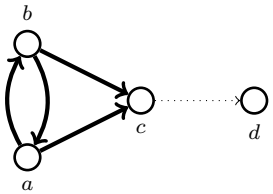
Now we focus on negative goals. In the general case, RP is not correct for the goal $\neg S_{\exists}(d)$.

Example 16 (cont.)

Here the procedure returns the set of moves $\mathcal{M}_d^{CON} = \{(ad, +, \#), \{(ad, +, \#), (da, -, \#)\}\}$*. Both these moves belong to the set* $\mathbb{M}_{\neg\forall}^{Comp}$*. But for the first move, it holds that* $\{(ad, +, \#)\} \notin \mathbb{M}_{\neg\exists}^{S}$*, because if we add the attack* (a, d)*, there will still exist some preferred (admissible and complete) extension which contains* d*, and this is the singleton* $\{d\}$*. Therefore, in the general case, it holds that* $\mathcal{M}_d^{CON} \not\subseteq \mathbb{M}_{\neg\exists}^{S}$*. So, in the general case, RP is not correct for the goal* $\neg S_{\exists}(d)$*.*

Also, in the general case, RP is not correct for the goal $\neg Pref_{\forall}(d)$.

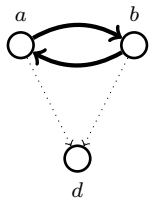
Example 21 *Let* $SM = \langle A, R, R^+, R^- \rangle$ *be an ASMA with* $A = \{a, b, c, d\}$*,* $R = \{(a, b), (a, c), (b, a), (b, c)\}$*,* $R^+ = \{(c, d)\}$*,* $R^- = \{\}$*, and let* $d \in A$ *be the issue.*



Notice that $\{(cd, +, \#)\} \in \mathcal{M}_d^{CON}$*, but also* $\{(cd, +, \#)\} \notin \mathbb{M}_{\neg\forall}^{Pref}$*, as even if we add the attack* cd*, argument* d *will still appear in all the preferred extensions. Therefore, in the general case,* $\mathcal{M}_d^{CON} \not\subseteq \mathbb{M}_{\neg\forall}^{Pref}$*. So, in the general case, RP is not correct for the goal* $\neg Pref_{\forall}(d)$*.*

Finally, in the general case, RP is not complete for the goal $\neg S_{\exists}(d)$.

Example 18 (cont.)



Here the procedure returns the set of moves $\mathcal{M}_d^{CON} = \{\{(ad, +, \#)\}, \{(bd, +, \#)\}\}$. *Therefore, the move* $\{(ad, +, \#), (bd, +, \#)\}$ *which is the only target set of* $\mathbb{T}^S_{\neg\exists}$, *is not returned by the procedure. Therefore, in the general case, the relation* $\mathbb{T}^S_{\neg\exists} \subseteq \mathcal{M}_d^{CON}$ *does not hold. So, in the general case, RP is not complete for the goal* $\neg S_\exists(d)$.

To conclude, we emphasize that *RP* is both correct for successful moves, and complete for target sets, for the goals $S_\exists(d)$ and $\neg Comp_\forall(d)$. It is also complete for target sets for the goal $Comp_\forall(d)$.

Finally, Figure 5.6 graphically represents the links between the sets of successful moves, the sets of target sets for both positive and negative goals, as well as the sets of moves $\mathcal{M}^{PRO}$ and $\mathcal{M}^{CON}$ which are returned by the *RP* procedure.

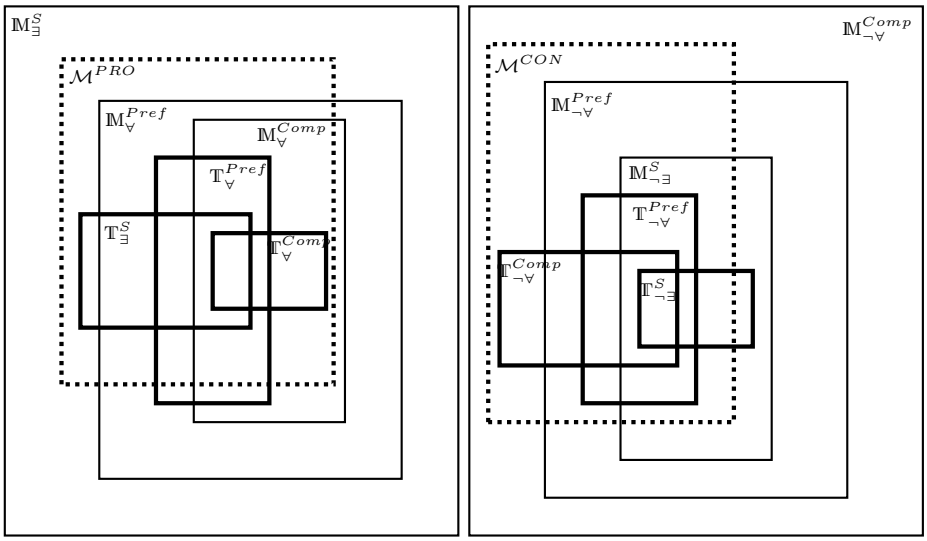


Figure 5.6: On the left: the sets of successful moves and target sets for the positive goals, and $\mathcal{M}^{PRO}$ (in the dotted rectangle). On the right: the sets of successful moves and target sets for the negative goals, and $\mathcal{M}^{CON}$ (in the dotted rectangle).

At this point, we have concluded our analysis of the basic properties of target sets. Having in mind that debating users may want to adopt specific goals during a debate, we also defined a rewriting procedure with which users can identify the easiest ways (target sets) to achieve their goals.

But in practice, things do not always go according to plan. Therefore, in the following, we analyze what happens in two different cases: firstly, when "desired" changes take place (changes which are in a target set for our goal); and secondly, when "undesired" changes take place (changes which are *not* in any target set for our goal). This analysis will further highlight the practical usefulness of target sets.

5.4 Target Set Evolution

Given an ASMA $SM = \langle A, R, R^+, R^- \rangle$ and an argumentative goal g, we want to analyze the possible repercussions that a move m on SM may have, on the set of target sets for g. If an agent wishes to satisfy g, then it seems natural to play moves which focus on target sets for g. On the other hand, playing moves which do not focus on target sets for g seems counter-intuitive. Here we analyze the evolution of target sets in these two cases, by providing some interesting properties and a number of illustrative examples.

5.4.1 Playing outside target sets

We first turn our attention into what happens if we play moves which are not part of any target set for a given goal g. Intuitively, we will not get closer to achieving g. We will see that playing such an move could even lead us farther away from g, in the sense that we will need to perform more moves later to satisfy it.

We begin by providing three small examples:

Example 22 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA with $A = \{a, b, d\}$, $R = \{(a, d)\}$, $R^+ = \{(a, b)\}$, $R^- = \{(a, d)\}$. Also, let $g = Comp_\forall(d, AS)$ be the goal we want to achieve (put d in the grounded extension).*

Figure 5.7: In this example, playing outside $\mathbb{T}(SM)$ leads to no change in the target sets.

The system is illustrated in Figure 5.7. Since the goal is to put d in the grounded extension, it holds that $\mathbb{T}(SM) = \{\{(ad, -, \#)\}\}$. So, $(ab, +, \#)$ does not belong to any target set of SM for g. If we play move $m = (ab, +, \#)$, we obtain a modified system $SM' = \Delta(SM, m)$ which has exactly the same set of target sets $\mathbb{T}(SM') = \{\{(ad, -, \#)\}\}$. So, in this example no target set is modified, and $\mathbb{T}(SM) = \mathbb{T}(SM')$.

Example 23 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA with $A = \{a, b, c, d\}$, $R = \{(b, a), (a, d)\}$, $R^+ = \{(c, a)\}$, $R^- = \{(b, a), (a, d)\}$. Also, let $g = \neg Comp_\forall(d, AS)$ be the goal we want to achieve (drop d from the grounded extension). The system is illustrated in Figure 5.8. It holds that $\mathbb{T}(SM) = \{\{(ba, -, \#)\}\}$. We see that $(ca, +, \#)$ does not belong to any target set for g. Now, if we play move $m = (ca, +, \#)$, we obtain a modified system $SM' = \Delta(SM, m)$ which has a set of target sets $\mathbb{T}(SM') = \{\{(ba, -, \#), (ca, -, \#)\}\}$. We see that there is a new target set which is bigger than an old target set.*

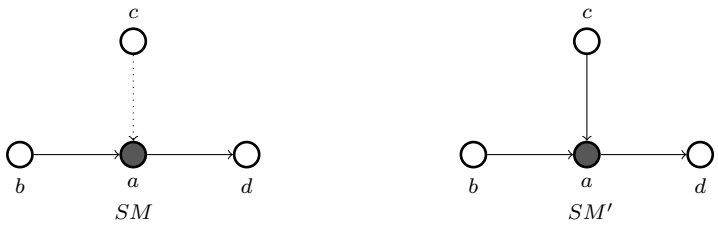


Figure 5.8: In this example, playing outside $\mathbb{T}(SM)$ leads to bigger target sets.

Example 24 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA with $A = \{a, b, c, d, e\}$, $R = \{(b, a), (a, d)\}$, $R^+ = \{(c, a), (e, c)\}$, $R^- = \{(b, a), (a, d)\}$. Also, let $g = \neg Comp_\forall(d, AS)$ be the goal we want to achieve (drop d from the grounded extension).*

The system is illustrated in Figure 5.9. As in the previous example, for the goal g, it holds that $\mathbb{T}(SM) = \{\{(ba, -, \#)\}\}$. We see that $(ca, +, \#)$ does not belong to any target set. Now, if we play move $m = (ca, +, \#)$, we obtain a modified system $SM' = \Delta(SM, m)$ which has a set of target sets $\mathbb{T}(SM') = \{\{(ba, -, \#), (ca, -, \#)\}, \{(ba, -, \#), (ec, +, \#)\}\}$. In this example, we see that the number of target sets increases (from one it becomes two), though they are all bigger than the old target set.

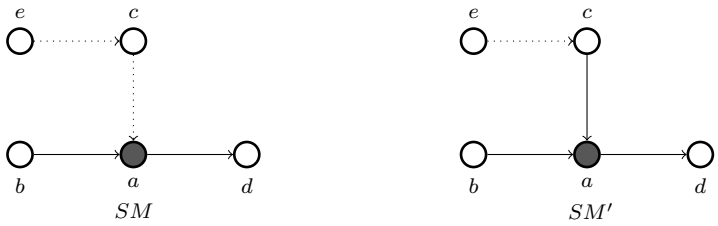


Figure 5.9: In this example, playing outside $\mathbb{T}(SM)$ leads to bigger target sets. Furthermore, it increases their number.

From these examples, our intuition that playing outside target sets leads us farther away from the goal is strengthened. The following properties will describe exactly what happens when a move outside of target sets is played on a system. But first we introduce some needed notation. Let m denote a set of atoms concerning the attacks of an ASMA $SM = \langle A, R, R^+, R^- \rangle$.[9] For the sign $x = +$ (resp. $x = -$), we denote its inverse sign $\overline{x} = -$ (resp. $\overline{x} = +$). Also, $\overline{m}$ denotes the set of atoms in m with inversed signs, so $(ab, x, \#) \in \overline{m}$ iff $(ab, \overline{x}, \#) \in m$. Moreover, function $simp(m)$ "simplifies" the set m, as follows: $(ab, x, \#) \in simp(m)$ iff $(ab, x, \#) \in m$ and $(ab, \overline{x}, \#) \notin m$. For example, if $m = \{(ab, +, \#), (cd, +, \#), (cd, -, \#)\}$, then $simp(m) = \{(ab, +, \#)\}$.

In the following, SM denotes an ASMA and $\mathbb{T}(SM)$ denotes the set of target sets for a specific goal g (we shall always consider the same goal). Also, m is a move on GB such that m *does not contain any atom of any target set of* SM.[10] Therefore, after playing m, the goal remains unsatisfied in the resulting system $SM' = \Delta(SM, m)$,[11] while the set of target sets changes and becomes $\mathbb{T}(SM')$.

The first property states that if we play such a move on system SM, then for every new target set in $\mathbb{T}(SM')$, there is at least one old target set in $\mathbb{T}(SM)$ which is smaller (or equal). In a sense, there are no "completely new" target sets (unrelated to the old ones).

Property 13 *If m is a move on SM, such that $SM' = \Delta(SM, m)$, and no atom of m belongs in a target set of SM, then $\forall t' \in \mathbb{T}(SM')\ \exists t \in \mathbb{T}(SM)$ such that $t \subseteq t'$.*

Proof 13 *For every $t' \in \mathbb{T}(SM')$, it holds that move $simp(t' \cup m)$ on SM is successful. To see why, we start with SM and we make the changes in m. We end up with SM' and the goal remains unsatisfied, as m is not successful. Then, we make the changes in t' and the goal becomes satisfied, since t' is a target set of SM'. In case that $(ab, x, \#) \in t'$ and $(ab, \overline{x}, \#) \in m$, during the process just described, we both add and remove the attack (a, b), so we could have simply avoided this (hence the simplification $simp(t' \cup m)$). Therefore, move $simp(t' \cup m)$ on SM is, indeed, successful. From this, it follows that $\exists t \subseteq simp(t' \cup m)$ such that t is a target set on SM. By definition of the simplification function, we have $simp(t' \cup m) \subseteq t' \cup m$. Therefore, $t \subseteq t' \cup m$. We also know that $t \cap m = \{\}$, because m contains no atom of a target set of SM. As a result $t \subseteq t'$.*

The next property states that for every new target set, there is no old target set which is strictly bigger.

Property 14 *If m is an move on SM, such that $SM' = \Delta(SM, m)$, and no atom of m belongs in a target set of SM, then $\forall t' \in \mathbb{T}(SM')\ \nexists t \in \mathbb{T}(SM)$ such that $t' \subset t$.*

Proof 14 *By contradiction, let us assume that $\exists t' \in \mathbb{T}(SM')$ such that $\exists t \in \mathbb{T}(SM)$ with $t' \subset t$. From Property 13 we have that $\exists t^* \subseteq t'$ such that t^* is a target set of SM. From $t^* \subseteq t'$ and*

[9] m is not necessarily a move on system SM, as for example, m may contain both $(ab, +, \#)$ and $(ab, -, \#)$.

[10] We shall simply say that m is not in a target set.

[11] Because if m was successful, then a subset of m would be a target set. Impossible, since m contains no atoms of any target set.

from our assumption that $t' \subset t$, it follows that $t^* \subset t$. Therefore, t is not a target set on SM. Contradiction.

Also, for every old target set, there is at least one new target set which is bigger (or equal). In a sense, no old target set "disappears".

Property 15 *If m is a move on SM, such that $SM' = \Delta(SM, m)$, and no atom of m belongs in a target set of SM, then $\forall t \in \mathbb{T}(SM)\ \exists t' \in \mathbb{T}(SM')$ such that $t \subseteq t'$.*

Proof 15 *For every $t \in \mathbb{T}(SM)$ it holds that move $t \cup \overline{m}$ on SM' is successful. To see why, we start with SM' and we make the changes in $\overline{m}$. We end up with SM (where the goal is not satisfied). Then, we make the changes in t and the goal becomes satisfied, since $t \in \mathbb{T}(SM)$. Notice that no atom can be both in m and t (nor in $\overline{m}$ and t), because move m contains no atom of a target set on SM. This is why we did not need to write the simplification $simp(t \cup \overline{m})$ (in place of $t \cup \overline{m}$). Now, since move $t \cup \overline{m}$ on SM' is successful, it follows that $\exists t' \subseteq t \cup \overline{m}$ such that t' is a target set on SM'. Therefore, we can write $t' = t^* \cup \overline{m}^*$, with $t^* \subseteq t$ and $\overline{m}^* \subseteq \overline{m}$.*

Now, we note that there is a successful move on SM which contains all the atoms of t^* and a subset of the atoms of m. Let us see why this is true. We begin with SM and we make the changes in m (the goal remains not satisfied). We obtain SM'. Then, we make the changes in $t' = t^* \cup \overline{m}^*$, and the goal becomes satisfied, since $t' \in \mathbb{T}(SM')$. So, we see that $simp(m \cup t^* \cup \overline{m}^*)$ is a successful move on SM. Notice that we are using the simplification function simp *because all the atoms in $\overline{m}^*$ will be simplified with their "corresponding" atoms in m. Therefore, there exists a target set on SM (denoted t^B) which is a subset of $simp(m \cup t^* \cup \overline{m}^*)$. So, $t^B \in \mathbb{T}(SM)$ and $t^B \subseteq simp(m \cup t^* \cup \overline{m}^*)$. Notice that $simp(m \cup t^* \cup \overline{m}^*) \subseteq m \cup t^*$ (because all the atoms of $\overline{m}^*$ are simplified with their inverse atoms of m). Therefore, $t^B \subseteq m \cup t^*$. We know that t^B does not contain any atoms of m (because we initially said that no target set of SM contains atoms of m), so it holds that $t^B \subseteq t^*$. Moreover, we remind that $t^* \subseteq t$, so we have $t^B \subseteq t^* \subseteq t$. Since both t^B and t are target sets of SM, none is a subset of the other, so $t^B = t^* = t$. Finally, in the equality $t' = t^* \cup \overline{m}^*$, we can replace t^* with t, and obtain $t' = t \cup \overline{m}^*$. Thus, we have proven that $t \subseteq t'$.*

Finally, the cardinality of the new set of target sets is greater (or equal) to the cardinality of the old set of target sets.

Property 16 *If m is a move on SM, such that $SM' = \Delta(SM, m)$, and no atom of m belongs in a target set of SM, then $|\mathbb{T}(SM')| \geq |\mathbb{T}(SM)|$.*

Proof 16 *Property 15 states that $\forall t \in \mathbb{T}(SM)\ \exists t' \in \mathbb{T}(SM')$ such that $t \subseteq t'$. First, we will prove that there is such a target set t', for which an additional "constraint" is satisfied, namely it holds that $t' = t \cup \overline{m}^*$, with $\overline{m}^* \subseteq \overline{m}$. Let us see why this is the case. The move $t \cup \overline{m}$ is successful on SM'. Therefore, there is a target set t' on SM' such that $t' \subseteq t \cup \overline{m}$. We want to prove that $t' = t \cup \overline{m}^*$, with $\overline{m}^* \subseteq \overline{m}$, in other words that t' contains all the atoms of t (and not only a subset of them). Let us assume this is not true, therefore t' contains only a subset of the atoms of t, in other words $t' = t^* \cup \overline{m}^*$, with $t^* \subset t$ and $\overline{m}^* \subseteq \overline{m}$. We will prove this to be impossible. Notice that $simp(m \cup t^* \cup \overline{m}^*)$ is successful on SM. Therefore, there is a target set t^B on SM, such that $t^B \subseteq simp(m \cup t^* \cup \overline{m}^*) \subseteq m \cup t^*$. But, a target set on SM cannot contain atoms of m, so $t^B \subseteq t^*$. Now, from $t^B \subseteq t^*$ and $t^* \subset t$, we get $t^B \subset t$. This is impossible, since both are target sets of SM. Therefore, we have proven that t' is of the form $t' = t \cup \overline{m}^*$, with $\overline{m}^* \subseteq \overline{m}$. Now, in order to prove the inequality $|\mathbb{T}(SM')| \geq |\mathbb{T}(SM)|$, it suffices to show that there exists no pair of target sets (t_1,t_2) of SM which "correspond" to the same target set (t') of SM', under the previous constraint. Again, let us assume the opposite and try to obtain a contradiction. So, we assume that $\exists t_1, t_2, t'$ with $t_1, t_2 \in \mathbb{T}(SM)$, $t_1 \neq t_2$, $t' \in \mathbb{T}(SM')$ such that $t' = t_1 \cup \overline{m}_1^*$ and $t' = t_2 \cup \overline{m}_2^*$, with $\overline{m}_1^*, \overline{m}_2^* \subseteq \overline{m}$. In this case, from $t' = t_1 \cup \overline{m}_1^*$ and $t' = t_2 \cup \overline{m}_2^*$, we get $t_1 \cup \overline{m}_1^* = t_2 \cup \overline{m}_2^*$. Notice that no atom of $\overline{m}_1^*$ (the same holds for $\overline{m}_2^*$) can also belong to t_1 or t_2. This is true because $\overline{m}_1^*, \overline{m}_2^* \subseteq \overline{m}$, and we have initially assumed that no atom of m belongs to a target set of SM (the same holds for the sets $\overline{m}$, $\overline{m}_1^*$, and $\overline{m}_2^*$). So, it follows that $t_1 = t_2$,*

and this contradicts with $t_1 \neq t_2$. Thus, there exists no pair of target sets (t_1,t_2) of SM which "correspond" to the same target set (t') of SM', under the previous constraint.

Let us draw some conclusions from the above properties. After playing a move whose atoms are not found in any target set, our intuition that the goal becomes "harder" to satisfy (or at least not "easier") is confirmed, essentially by Property 13 which states that for every new target set, there was an old one, which was smaller (or equal). In a sense, after the move m, for every way to achieve the goal, there was a shorter way available before. Note that even in cases where the number of target sets increases (possible, as shown in Example 5.9), it still gets harder to satisfy the goal, for the reason stated above.

5.4.2 Playing in target sets

The previous results indicate that playing outside of target sets for a goal takes us farther away from its satisfaction. But what does happen if we choose to play a move m in a target set? If $m \subset t \in \mathbb{T}(SM)$, then according to the definition of target set, $t \setminus m$ will become a target set of the modified system $\Delta(SM, m)$. In this case the target set t will "shrink". But what can happen to the other target sets?

Let us first present some illustrative examples:

Example 25 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an ASMA with $A = \{a, b, c, d, e, f\}$, $R = \{(a, d), (e, d)\}$, $R^+ = \{(b, a), (c, a), (f, e)\}$, $R^- = \{\}$. Also, let $g = Comp_\forall(d, SM)$ be the goal we want to achieve (put d in the grounded extension).*

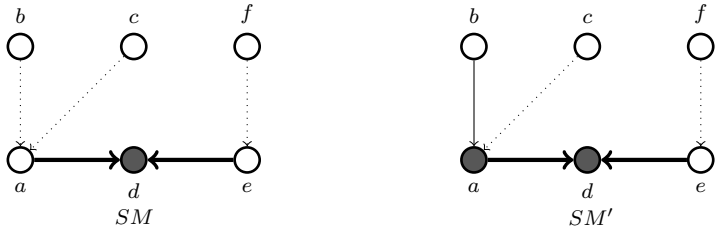


Figure 5.10: By playing in $\mathbb{T}(SM)$, adding (b, a), the target sets get smaller. Furthermore, their number decreases.

The system is illustrated in Figure 5.10. The set of target sets for the goal g is $\mathbb{T}(SM) = \{\{(ba, +, \#), (fe, +, \#)\}, \{(ca, +, \#), (fe, +, \#)\}\}$. So, the atom $(ba, +, \#)$ belongs in the first minimal target set. If we play move $m = \{(ba, +, \#)\}$ on SM, we obtain a modified argumentation system SM' which has a new set of minimal target sets $\mathbb{T}(SM') = \{\{(fe, +, \#)\}\}$. In this case, notice that the cardinality of the set of target sets has decreased.

The system is illustrated in Figure 5.11. If instead of adding the attack (b, a), we add the attack (f, e), we obtain a modified argumentation system SM'' such that $\mathbb{T}(SM'') = \{\{(ba, +, \#)\}, \{(ca, +, \#)\}\}$. So, in this case the cardinality of the set of target sets remains the same, and all the new target sets get smaller.

Example 26 *Let $SM = \langle A, R, R^+, R^- \rangle$ be an argumentation system with $A = \{a, b, c, d, e\}$, $R = \{(a, d), (b, a), (b, d)\}$, $R^+ = \{(c, b), (e, c)\}$, $R^- = \{(a, d), (b, d)\}$. Also, let $g = Comp_\forall(d, SM)$ be the goal we want to achieve (put d in the grounded extension).*

The system is illustrated in Figure 5.12. The set of target sets is $\mathbb{T}(SM) = \{\{(bd, -, \#)\}, \{(cb, +, \#), (ad, -, \#)\}\}$. The atom $(cb, +, \#)$ belongs to the second target set. If we play move $m = \{(cb, +, \#)\}$, then we obtain a modified argumentation system SM' such that $\mathbb{T}(SM') = \{\{(bd, -, \#), (cb, -, \#)\}, \{(bd, -, \#), (ec, +, \#)\}, \{(ad, -, \#)\}\}$. We notice that the cardinality

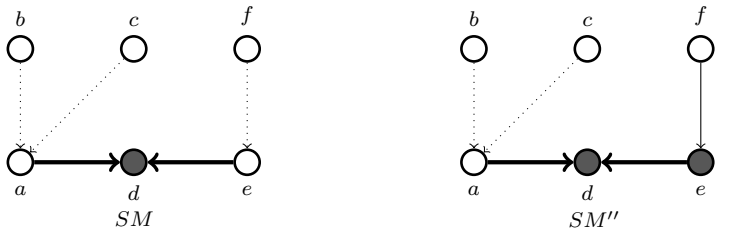


Figure 5.11: By playing in $\mathbb{T}(SM)$, adding (f, e), the target sets get smaller. Their number remains unchanged.

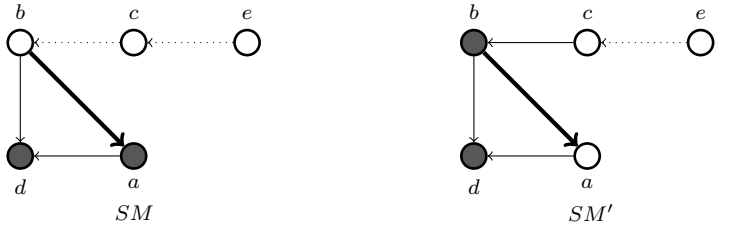


Figure 5.12: By playing in $\mathbb{T}(SM)$, adding (c, b), the following happens: One target set gets smaller (the one we played in), while others get bigger. Furthermore, the number of target sets increases.

of the set of target sets increases. Also, the target set $\{(cb, +, \#), (ad, -, \#)\}$ *of course shrinks, whereas the target set* $\{(bd, -, \#)\}$ *gets "replaced" by two bigger target sets:* $\{(bd, -, \#), (cb, -, \#)\}$ *and* $\{(bd, -, \#), (ec, +, \#)\}$.

Let us conclude on the possible results of playing a move in a target set: playing such a move certainly shrinks that specific target set. However, the repercussions on the other target sets of SM is uncertain: some may shrink too, but others may remain unchanged, or even grow. Moreover, as we have seen in the examples, the cardinality of $\mathbb{T}(SM)$ could decrease, remain unchanged, or increase in $\Delta(SM, m)$.

We want to stress out the fact that at least one "path" towards the satisfaction of the goal becomes shorter (one target set shrinks), and this is an important fact which was not verified when the move was made outside of target sets.

At this point we have concluded our analysis of the notion of target set, which has been defined on the basis of the notion of ASMA, a dynamic abstract argumentation system with modifiable attacks. Given that a dynamic system can be subject to change, two questions that naturally come up are the following: if such a system represents a debate which has finished, then how certain can we be about the conclusions we draw from it? And how can we address this uncertainty? These will be the focal points of the next section.

5.5 Controversy of Argumentation Systems

At the beginning of this chapter we have defined ASMAs, in order to reason about structural modifications of abstract argumentation systems. Now, we turn our attention to debates among experts. The structural modifications considered by ASMAs are the additions and removals of attacks. Though this type of modifications may take place in debates (as presented in Chapter 3), a shortcoming of using ASMAs in debates is that they simply partition attacks into fixed and

debated. But how do we decide which attacks are fixed and which attacks are debated? Furthermore, we may want to consider an additional level of detail in our analysis: separate attacks which are strongly supported by the agents (though not fixed) from attacks on which the agents' opinions are split. In order to address these issues, we fall back to the use of WASs and the notion of Gameboard in Definition 23.

We shall consider a procedure which takes place in *three* phases. The first phase consists in the expression of the agents' opinions, and their aggregation on a single GB. Recall that we do not commit to any specific protocol here. Then comes an evaluation phase which allows to determine how controversial the GB is. If required, there follows a third phase where an additional expert is chosen and asked to provide his opinion, with the goal of making the GB less controversial.

- **Debate phase**: The agents express their opinions. They may introduce new arguments and attacks, and they may vote on already introduced attacks. The expressed opinions are aggregated on a single GB.
- **Evaluation phase**: The result of the debate is evaluated, with respect to the stability of its elements (which will be explained shortly).
- **Stabilization phase**: In this final phase we search for the best way to stabilize the debate. We assume that we cannot indefinitely continue debating on its controversial points, so instead we promote a quick and efficient way to stabilize the debate. More specifically, we assume that there is a pool of agents, with known expertise, but whose opinions on the points of the debate are unknown. We propose a method to choose among the agents, the one who is deemed the most capable of stabilizing the debate by offering his opinion.

5.5.1 Debate Phase: The experts express their opinions

Without committing to any specific debate protocol, we remind the basic elements of this phase, as they have been previously presented:

1. Every agent $i \in N$ has expertise in topics $exp(i) \subseteq T$.
2. Agents introduce arguments on the common Gameboard.
3. Every argument a concerns topics $top(a) \subseteq T$.
4. Agents vote on attacks between arguments.

As presented in Chapter 3, the Gameboard is a WAS $\langle A_{GB}, R_{GB}, Eval\rangle$, where $Eval$ is the set of evaluation vectors of all the attacks in R_{GB}. The evaluation vector of attack (a, b) is $\vec{v}(a, b) = \langle w(a, b), mw(a, b)\rangle$ and it aggregates all the votes cast on (a, b).

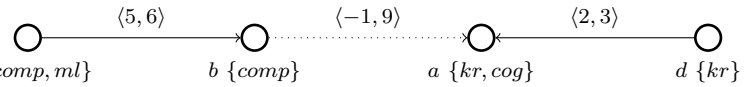
In the following, in order to lighten the notation of a Gameboard $\langle A_{GB}, R_{GB}, Eval\rangle$, we drop the subscripts from A_{GB} and R_{GB}, and we use the name $\mathbb{W}$ (instead of the usual name GB).

Until the end of this section, we will use labelling-based acceptability. To keep things as simple as possible, we will use the labelling which maximizes the number of *undec* labels. This labelling is unique and it corresponds to the grounded extension, as shown in [Cam06].

We denote by $lab^{\mathbb{W}}(a)$ the label of argument a, with respect to the counterpart system of $\mathbb{W}$. Also, we denote by $\mathbb{L}^{\mathbb{W}}_{in}$ (resp. $\mathbb{L}^{\mathbb{W}}_{out}$, $\mathbb{L}^{\mathbb{W}}_{undec}$) the set of *in* (resp. *out*, *undec*) labels of the counterpart system of $\mathbb{W}$.

Let us reconsider a previously given example.

Example 7 (cont.) *After the votes of the three PC members, let the obtained Gameboard* $\mathbb{W}$ *be as follows:*



The counterpart AS of $\mathbb{W}$ *is illustrated below.* $\mathbb{L}^{\mathbb{W}}_{in} = \{c, d\}$, $\mathbb{L}^{\mathbb{W}}_{out} = \{a, b\}$ *and* $\mathbb{L}^{\mathbb{W}}_{undec} = \{\}$.

5.5.2 Evaluation phase: measuring the controversy

Once the agents have expressed their opinions, the question is whether the GB is controversial, and if so, to what extent. In real argumentation debates, we usually want the GB to be as uncontroversial as possible, in order to avoid discussions and discontentments about the outcome of the procedure. Such disagreements could appear if the majority is not clearly defined, or if the final outcome can be interpreted in several ways. Several criteria can be used to measure how controversial a GB is. In the following, we propose three such criteria.

The first criterion is the *stability of the attacks*. An attack is called *stable* if it is difficult to be questioned. Naturally, a choice has to be made, in order to define what exactly "difficult" means. The choice made in this work (of course not unique) is that, if there exists *no agent* (not even an expert in all topics of T) who can change a given attack's sign by voting, then this specific attack is considered *stable*.

The second criterion is the *persistence of the arguments' acceptability status*. It is linked to the previous criterion, as an argument is called *persistent* if its acceptability status (its label) does not depend on the insertion and deletion of unstable attacks. From the definition of stable attack we have previously given, it follows that an argument's status is persistent if there is no expert able to change its status, by voting on attacks.

Finally, the third criterion is the *decidability of the arguments' acceptability status*. An argument can be either undecided (attributed the *undec* label), or decided (attributed either the *in* or the *out* label). Of course, the more *undec* arguments there are, the more controversial the debate is, as its conclusions are unclear.

Attack Stability

Let us begin with the first criterion, the stability of the attacks. We consider a natural qualitative scale and we introduce three types of attacks. The *beyond any doubt* attacks ($R^{\mathbb{W}}_{bd}$) are the ones on which sufficiently many agents agree (or, as a particular case, no agent has stated them). The *strong* attacks ($R^{\mathbb{W}}_{str}$) are defined as the attacks which are not beyond any doubt, but a single expert cannot change the sign of their weights. Finally, the *weak* attacks ($R^{\mathbb{W}}_{wk}$) are neither beyond any doubt nor strong, therefore a single agent may be able to change their sign. Thus, weak attacks are the most controversial attacks of a system.

Definition 41 *Let* $\mathbb{W} = \langle A, R, Eval\rangle$ *be a GB.* $\forall (a, b) \in R$, *let* $\vec{v}(a, b) = \langle w(a, b),\ mw(a, b)\rangle$ *be the evaluation vector of* (a, b). *The set of attacks* R *is partitioned into the three following sets:*

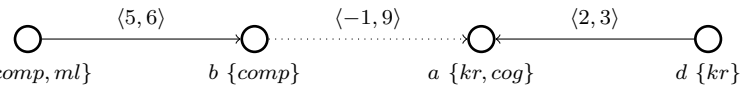
- *An attack* (a, b) *is beyond any doubt if either* $\vec{v}(a, b) = \langle 0, 0\rangle$, *or if its weight is "very significant". The latter is true when two conditions hold: The number of voters on* (a, b) *is greater than a threshold* $\delta \in \mathbb{N}$, *and* $\frac{|w(a,b)|}{mw(a,b)}$ *is greater than a threshold* $\epsilon \in]0, 1[$. *Thus, the set of* **beyond any doubt attacks** *is defined as:*
  $R^{\mathbb{W}}_{bd} = \{(a, b) \in R \mid$ *either* $\vec{v}(a, b) = \langle 0, 0\rangle$, *or* $\frac{mw(a,b)}{|top(a,b)|} > \delta$ *and* $\frac{|w(a,b)|}{mw(a,b)} > \epsilon\}$
- *The set of* **strong attacks** *is defined as:*
  $R^{\mathbb{W}}_{str} = R^{\mathbb{W}}_{strP} \cup R^{\mathbb{W}}_{strN}$, *where:*
  $R^{\mathbb{W}}_{strP} = \{(a, b) \in R \mid (a, b) \notin R^{\mathbb{W}}_{bd},\ w(a, b) > 0,\ w(a, b) - |top(a, b)| > 0\}$
  $R^{\mathbb{W}}_{strN} = \{(a, b) \in R \mid (a, b) \notin R^{\mathbb{W}}_{bd},\ w(a, b) \leq 0,\ |w(a, b)| - |top(a, b)| \geq 0\}$

- *The set of* **weak attacks** *is defined as:*
  $R^{\mathbb{W}}_{wk} = R^{\mathbb{W}}_{wkP} \cup R^{\mathbb{W}}_{wkN}$, *where:*
  $R^{\mathbb{W}}_{wkP} = \{(a,b) \in R \mid (a,b) \notin R^{\mathbb{W}}_{bd},\ w(a,b) > 0,\ w(a,b) - |top(a,b)| \leq 0\}$
  $R^{\mathbb{W}}_{wkN} = \{(a,b) \in R \mid (a,b) \notin R^{\mathbb{W}}_{bd},\ w(a,b) \leq 0,\ |w(a,b)| - |top(a,b)| < 0\}$

The notation strP means "strong attack with positive weight" and strN means "strong attack with negative weight".[12]

Note that the debate's designer can set the values of δ and ϵ as wished, in order to increase or decrease the number of beyond any doubt attacks. Intuitively, δ is a threshold used to assess if there is a sufficient number of voters on any given attack, while ϵ is a threshold used to assess how unanimous the votes have been on any given attack.

Example 7 (cont.)



Let $\delta = 4$ and $\epsilon = 0.5$. We have $R^{\mathbb{W}}_{str} = \{(c,b)\}$, $R^{\mathbb{W}}_{wk} = \{(b,a),(d,a)\}$, and $R^{\mathbb{W}}_{bd} = R \setminus (R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk})$. The attack (c,b) is not beyond any doubt, because $\delta = 4$ and only 3 agents have voted on it. Also, (c,b) is not weak, because no single agent can change its sign by voting on it. As a result, (c,b) is a strong attack. Finally, the attacks (b,a) and (d,a) are not beyond any doubt, and a single expert in all topics could change their signs by voting. Therefore, (b,a) and (d,a) are weak attacks.

Before continuing, let us draw a correspondence between the above types of attacks, and the types of attacks of an ASMA $\langle A, R, R^+, R^- \rangle$. As we will see, the latter capture less information.

- The beyond any doubt attacks correspond to the union of fixed attacks R^{fix}, and fixed non-attacks R^{fixN} of an ASMA.
- The weak attacks correspond to the debated attacks R^{deb} of an ASMA.
- The strong attacks do not have an exact equivalent in an ASMA, as they are neither *totally* fixed, nor *totally* debated.

The beyond any doubt and strong attacks, which are difficult to change, shall be called *stable* attacks.

Definition 42 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a WAS. The set of* **stable attacks** *of $\mathbb{W}$ is $R^{\mathbb{W}}_{stab} = R^{\mathbb{W}}_{bd} \cup R^{\mathbb{W}}_{str}$. The set of* **unstable attacks** *of $\mathbb{W}$ is $R^{\mathbb{W}}_{\overline{stab}} = R^{\mathbb{W}}_{wk}$.*

Argument Persistence

Let us now focus on the second criterion regarding debate stability. Roughly, a *persistent argument* is an argument whose status (label) remains unchanged, regardless of future changes in the weak attacks. Thus, a persistent argument is an argument whose label cannot be changed by the vote of a single expert (even if he is expert in all topics of T). To compute the set of persistent arguments, we need to consider all the possible changes over the set of unstable (weak) attacks.

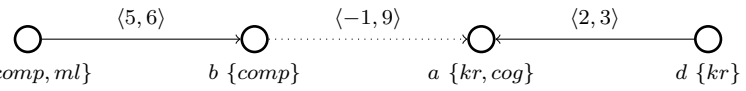
Definition 43 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, and let $R^{\mathbb{W}}_{\overline{stab}}$ be its set of unstable attacks. For all $R^{\mathbb{W}}_{\overline{stab},i} \subseteq R^{\mathbb{W}}_{\overline{stab}}$ let $\mathbb{W}^{alt}_i = \langle A, R, Eval^{alt}_i \rangle$ be an* **alternative GB** *such that $\forall r \in R^{\mathbb{W}}_{\overline{stab},i}$ with $w(r) \neq 0$, we have $w^{alt}_i(r) = -w(r)$; $\forall r \in R^{\mathbb{W}}_{\overline{stab},i}$ with $w(r) = 0$, we have $w^{alt}_i(r) = +1$; and $\forall r \in (R \setminus R^{\mathbb{W}}_{\overline{stab},i})$ we have $w^{alt}_i(r) = w(r)$. We denote by $\mathrm{Alt}(\mathbb{W})$ the* **set of alternative GBs** *of $\mathbb{W}$. It holds that $|\mathrm{Alt}(\mathbb{W})| = 2^{|R^{\mathbb{W}}_{\overline{stab}}|}$.*

[12] Not to be confused with R^+ and R^- which are the addable and removable attacks of an ASMA.

We can now define the sets of persistent and non-persistent arguments.

Definition 44 *Let* $\mathbb{W} = \langle A, R, Eval \rangle$ *be a GB and let* $\text{Alt}(\mathbb{W})$ *be its set of alternative GBs. The* **set of persistent arguments** *of* $\mathbb{W}$ *is defined as* $A^{\mathbb{W}}_{pers} = \{a \in A \mid \forall \mathbb{W}^{alt}_i \in \text{Alt}(\mathbb{W}) \; \mathbb{L}^{\mathbb{W}}(a) = \mathbb{L}^{\mathbb{W}^{alt}_i}(a)\}$ *The* **set of non-persistent arguments** *of* $\mathbb{W}$ *is defined as* $A^{\mathbb{W}}_{\overline{pers}} = A \setminus A^{\mathbb{W}}_{pers}$.

Example 7 (cont.)



There are four alternative GBs of $\mathbb{W}$*, due to the weak attacks* (b, a) *and* (d, a)*. In all four alternative GBs, the labels of arguments b (resp. c, d) are out (resp. in, in), while the label of argument a differs across the four alternative GBs, as it depends on the inclusion of the attack* (d, a)*. Therefore,* $A^{\mathbb{W}}_{pers} = \{b, c, d\}$ *and* $A^{\mathbb{W}}_{\overline{pers}} = \{a\}$.

Argument Decidability

Finally, the last criterion of a debate's controversy is the decidability of the arguments' status. In practice, we usually prefer to be able to either accept or reject any given argument, while undecidability hinders the drawing of conclusions. Therefore, we prefer to have more arguments with *in* and *out* labels, than arguments with *undec* labels. Until the end of this chapter, for simplicity reasons, we will consider a single type of labelling, the one which maximizes the number of *undec* labels. In that labelling, according to [Cam06], the set of arguments labelled *in* (here denoted $\mathbb{L}^{\mathbb{W}}_{in}$) corresponds to the unique grounded extension.

Definition 45 *Let* $\mathbb{W} = \langle A, R, Eval \rangle$ *be a GB. The* **set of decided arguments** *of* $\mathbb{W}$ *is defined as* $A^{\mathbb{W}}_{dec} = \mathbb{L}^{\mathbb{W}}_{in} \cup \mathbb{L}^{\mathbb{W}}_{out}$*. The* **set of undecided arguments** *of* $\mathbb{W}$ *is denoted* $A^{\mathbb{W}}_{undec} = \mathbb{L}^{\mathbb{W}}_{undec}$.

The three criteria of a GB's controversy which we have introduced, allow us to determine to what extent the GB, and therefore the debate, is controversial. In some cases, if the controversy is considered important, it could be very useful to know how to pick an additional expert from a pool of experts, and ask for his opinion, in a way that is likely to stabilize the debate.

5.5.3 Phase 3: Asking the opinion of an additional expert

As explained above, the result of a debate process represented as a GB, can be more or less controversial. Next we try to answer the following question: If there is a pool of additional agents, who have not already participated in the debate, then how can we choose one among them and ask his opinion, if our task is to render the debate less controversial? Two key assumptions are made on these additional experts: Their topics of expertise are known, but their opinions on the attacks on the GB are not known. The main difficulty at this phase lies exactly at this point, as the choice of agent to ask depends on his expertise, but the decision-maker does not know his *opinion*.

Let i be an expert who has not taken part to the discussion so far, and is asked for his opinion on $\mathbb{W}$. Given that i's topics of expertise are known, we can calculate the effects of his possible moves on $\mathbb{W}$, though we do not know *a priori* his opinion on the attacks. We will not ask i's opinion on beyond any doubt attacks ($R^{\mathbb{W}}_{bd}$), as we are certain about them, but only on strong and weak attacks ($R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk}$). Of course, after the expert's vote, some weak attacks may become strong, but the opposite may also happen. So, by asking the opinion of expert i, we face $2^{|R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk}|}$ possible GBs, without being able to know which one we will end up with.

Definition 46 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, and let i be an expert. We denote by $\text{Poss}_i(\mathbb{W}) = \{\mathbb{W}_{i,1}, \ldots \mathbb{W}_{i,n}\}$ the set of* **possible** *GBs we may end up with, if i gives his opinion on every weak and strong attack of $\mathbb{W}$.*

Note that $|\text{Poss}_i(\mathbb{W})| \leq 2^{|R^{\mathbb{W}}_{str} \cup R^{\mathbb{W}}_{wk}|}$, because agent i's opinion on every weak or strong attack can be either positive or negative. The difficulty now lies in the comparison of the available experts, in order to choose the one who can make the GB as uncontroversial as possible, but without us knowing their opinions. In particular, we observe that it may not be a good heuristic to select the expert with the highest number of topics of expertise, because these topics may not be the most relevant ones. More surprisingly, we also observe that it may not be appropriate to always prefer an expert who declares a strict superset of topics over another expert, because the additional impact provided by the extra topics may actually jeopardize an attack which was considered "strong" before. So, a careful study is required.

Dominance relations over experts

In order to choose an expert who can turn the debate uncontroversial, we focus on the notions of attack stability, argument persistence, and argument decidability. First, we study attack stability and we define a dominance relation over experts depending on their ability to "reinforce" and "weaken" some attacks. Then, we focus on argument persistence, and we define a dominance relation over experts depending on their ability to turn the arguments persistent (and non-persistent). Finally, as far as argument decidability is concerned, we define two dominance relations over experts depending on their ability to change the values of the arguments' labels. It is important at this point to observe that the difficulty we face here is that, when comparing experts, we do not compare two GBs, but two sets of possible GBs (those that can be obtained when questioning the experts). This leads to various natural definitions of (strict, easily adapted to weak) dominance:

- *i necessarily dominates j* if any GB that can be reached by i is "better" than any GB that can be reached by j.
- *i possibly dominates j* if there exists a GB that can be reached by i which is "better" than a GB that can be reached by j.
- *i optimistically dominates j* if the best GB that can be reached by i is "better" than the best GB that can be reached by j.
- *i pessimistically dominates j* if the worst GB that can be reached by i is "better" than the worst GB that can be reached by j.

Observe that while the necessary dominance guarantees that the GB obtained by one agent will be better, the optimistic and pessimistic dominance do not. However, they provide good reasons to prefer an expert over another one. But what do we mean exactly by "better" and what is compared precisely? In what follows, we elaborate on this, and we provide some properties focusing on optimistic and pessimistic dominance.

Dominance based on attack stability

We start by focusing on the capability on the experts to increase (resp. decrease) the weights of some weak (resp. strong) attacks, and turn them into strong (resp. weak) attacks.

Definition 47 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, let i be an expert, and let $\mathbb{W}_i \in \text{Poss}_i(\mathbb{W})$ be a possible GBs among those which i can reach. The set of attacks $R_i \subseteq R^{\mathbb{W}}_{wk}$* **are reinforced** *in $\mathbb{W}_i$ iff $\forall r \in R_i$ it holds that $r \in R^{\mathbb{W}_i}_{str}$. The set of attacks $R'_i \subseteq R^{\mathbb{W}}_{str}$* **are weakened** *in $\mathbb{W}_i$ iff $\forall r \in R'_i$ it holds that $r \in R^{\mathbb{W}_i}_{wk}$.*

We can now define the notion of (optimistic and pessimistic) *reinforce-dominance*: An expert i optimistically reinforce-dominates an expert j on a *GB* $\mathbb{W}$ iff j can reinforce only a subset of the attacks that i can reinforce, and i pessimistically reinforce-dominates j on $\mathbb{W}$ iff i can weaken only a subset of the attacks that j can weaken.

Definition 48 *Let $\mathbb{W} = \langle A, R, Eval\rangle$ be a GB, and let i, j be two experts. We say that i* **optimistically reinforce-dominates** *j on $\mathbb{W}$ iff: given that $\mathbb{W}_i \in \text{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}_j \in \text{Poss}_j(\mathbb{W})$) is a possible GB which contains the $\subseteq$-maximal set of stable attacks, denoted R_i (resp. R_j), it holds that $R_j \subseteq R_i$. We say that i* **pessimistically reinforce-dominates** *j on $\mathbb{W}$ iff: given that $\mathbb{W}'_i \in \text{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}'_j \in \text{Poss}_j(\mathbb{W})$) is a possible GB which contains the $\subseteq$-maximal set of unstable attacks, denoted R'_i (resp. R'_j), it holds that $R'_i \subseteq R'_j$. We say that i* **reinforce-dominates** *j on $\mathbb{W}$ iff i optimistically and pessimistically reinforce-dominates j.*

Let us now define some dominance relations based on the persistence of arguments.

Dominance based on argument persistence

Definition 49 *Let $\mathbb{W} = \langle A, R, Eval\rangle$ be a GB, let i be an expert, and let $\mathbb{W}_i \in \text{Poss}_i(\mathbb{W})$. The set of arguments $A_i \subseteq A^{\mathbb{W}}_{\overline{pers}}$* **are turned persistent** *in $\mathbb{W}_i$ iff $\forall a \in A_i$ it holds that $a \in A^{\mathbb{W}_i}_{pers}$. The set of arguments $A'_i \subseteq A^{\mathbb{W}}_{pers}$* **are turned non-persistent** *in $\mathbb{W}_i$ iff $\forall a \in A'_i$ it holds that $a \in A^{\mathbb{W}_i}_{\overline{pers}}$.*

We can now define the notion of (optimistic and pessimistic) *persist-dominance*: An expert i optimistically persist-dominates an expert j on a *GB* $\mathbb{W}$ iff j can turn persistent only a subset of the arguments that i can turn persistent, and i pessimistically persist-dominates j on $\mathbb{W}$ iff i can turn non-persistent only a subset of the arguments that j can turn non-persistent.

Definition 50 *Let $\mathbb{W}$ be a GB, and let i, j be two experts. We say that i* **optimistically persist-dominates** *j on $\mathbb{W}$ iff given that $\mathbb{W}_i \in \text{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}_j \in \text{Poss}_j(\mathbb{W})$) is a possible GB which contains the $\subseteq$-maximal set of persistent arguments, denoted A_i (resp. A_j), we have $A_j \subseteq A_i$. We say that i* **pessimistically persist-dominates** *j on $\mathbb{W}$ iff given that $\mathbb{W}'_i \in \text{Poss}_i(\mathbb{W})$ (resp. $\mathbb{W}'_j \in \text{Poss}_j(\mathbb{W})$) is a possible GB which contains the $\subseteq$-maximal set of non-persistent arguments, denoted A'_i (resp. A'_j), we have $A'_i \subseteq A'_j$. We say that i* **persist-dominates** *j on $\mathbb{W}$ iff i optimistically and pessimistically persist-dominates j.*

Let us now provide some properties regarding attack stability and argument persistence. We start with a quite intuitive property, stating that, as more attacks get reinforced (while no attacks get weakened) the set of persistent arguments will monotonically increase. Inversely, as more attacks get weakened (while no attacks get reinforced) the set of persistent arguments will monotonically decrease.

Property 17 *Let $\mathbb{W}$ be a GB. Then:*

(1) If a subset of weak attacks $R_1 \subseteq R^{\mathbb{W}}_{wk}$ is reinforced, while the weights of the other attacks do not change, then the set of persistent arguments will monotonically increase.

(2) If a subset of strong attacks $R_2 \subseteq R^{\mathbb{W}}_{str}$ is weakened, while the weights of the other attacks do not change, then the set of persistent arguments will monotonically decrease.

Proof 17 *(1) Let $R_1 \subseteq R^{\mathbb{W}}_{wk}$ be a subset of weak attacks which are reinforced, whereas the weights of the other attacks remain unchanged. Let $\mathbb{W}'$ denote the GB obtained after the attacks in R_1 are reinforced. Also, let $a \in A^{\mathbb{W}}_{pers}$ be a persistent argument of the initial system $\mathbb{W}$. So, it holds that $\forall \mathbb{W}^{alt} \in \text{Alt}(\mathbb{W})$, $lab^{\mathbb{W}}(a) = lab^{\mathbb{W}^{alt}}(a)$[13]. As $R^{\mathbb{W}'}_{wk} \subseteq R^{\mathbb{W}}_{wk}$, it holds that $\text{Alt}(\mathbb{W}') \subseteq \text{Alt}(W)$. Thus $\forall \mathbb{W}'^{alt} \in \text{Alt}(\mathbb{W}')$, we have $lab^{\mathbb{W}'}(a) = lab^{\mathbb{W}'^{alt}}(a)$. Therefore, the argument a remains persistent in $\mathbb{W}'$, so $a \in A^{\mathbb{W}'}_{pers}$. As a result, the set of persistent arguments will monotonically increase.*

[13]For the sake of simplicity, and without loss of generality, we do not mention here the agent modifying $\mathbb{W}$.

(2) Let $R_2 \subseteq R_{str}^{\mathbb{W}}$ be a subset of strong attacks which are weakened, whereas the weights of the other attacks remain unchanged. Let $\mathbb{W}'$ denote the GB obtained after the attacks in R_2 are weakened. Also, let $a \in A_{\overline{pers}}^{\mathbb{W}}$ be a non-persistent argument of the initial system $\mathbb{W}$. So, it holds that $\exists \mathbb{W}^{alt} \in \text{Alt}(\mathbb{W})$, such that $lab^{\mathbb{W}}(a) \neq lab^{\mathbb{W}^{alt}}(a)$ As $R_{wk}^{\mathbb{W}} \subseteq R_{wk}^{\mathbb{W}'}$, it holds that $\text{Alt}(\mathbb{W}) \subseteq \text{Alt}(\mathbb{W}')$. Thus $\exists \mathbb{W}'^{alt} \in \text{Alt}(\mathbb{W}')$, such that $lab^{\mathbb{W}'}(a) \neq lab^{\mathbb{W}'^{alt}}(a)$. Therefore, the argument a remains non-persistent in $\mathbb{W}'$, so $a \in A_{\overline{pers}}^{\mathbb{W}'}$. As a result, the set of non-persistent arguments will monotonically increase, which means that the set of persistent arguments will monotonically decrease.

From the previous property, it easily stems that, if an expert is able to turn persistent (resp. non-persistent) an argument a, and he is able to turn persistent (resp. non-persistent) an argument b, then he is able to turn persistent (resp. non-persistent) both aguments a and b, at the same time.

Property 18 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, let i be an expert, and let $a, b \in A$ be two arguments. Then:*

(1) If i is able to turn a persistent, and i is able to turn b persistent, then i is able to turn both arguments a, b persistent, in the same system $\mathbb{W}_i \in Poss_i(\mathbb{W})$.

(2) If i is able to turn a non-persistent, and i is able to turn b non-persistent, then i is able to turn both arguments a, b non-persistent, in the same system $\mathbb{W}_i \in Poss_i(\mathbb{W})$.

Proof 18 *(1) Let i be able to turn a persistent, by reinforcing the attacks of $R_1 \subseteq R_{wk}^{\mathbb{W}}$. Let also i be able to turn b persistent, by reinforcing the attacks of $R_2 \subseteq R_{wk}^{\mathbb{W}}$. Then, from property 17, it follows that i is able to turn persistent a and b (in the same system $\mathbb{W}_i \in Poss_i(\mathbb{W})$), by reinforcing the attacks of $R_1 \cup R_2$.*

(2) Let i be able to turn a non-persistent, by weakening the attacks of $R_1 \subseteq R_{str}^{\mathbb{W}}$. Let also i be able to turn b non-persistent, by weakening the attacks of $R_2 \subseteq R_{str}^{\mathbb{W}}$. Then, from property 17, it follows that i is able to turn non-persistent a and b (in the same system $\mathbb{W}_i \in Poss_i(\mathbb{W})$), by weakening the attacks of $R_1 \cup R_2$.

Finally, let us provide a property which shows the link between the reinforce-dominance, and the persist-dominance relations.

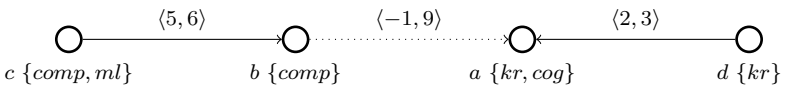
Property 19 *If an expert i reinforce-dominates an expert j on a GB $\mathbb{W}$, then i persist-dominates j on $\mathbb{W}$. The converse does not hold, in the general case.*

Proof 19 *($\rightarrow$) Expert i reinforce-dominates expert j on $\mathbb{W} = \langle A, R, Eval \rangle$. (1) Assume that j can turn $a \in A$ persistent. To do so, j has to reinforce a set of attacks R_j. As i reinforce-dominates j, i can also reinforce R_j, thus i can also turn a persistent. (2) Assume that i can turn $a \in A$ non-persistent. To do so, i has to weaken a set of attacks R_i. As i reinforce-dominates j, j can also weaken R_i, thus j can also turn a non-persistent. From (1) and (2), we obtain that i persist-dominates j.*

($\leftarrow$) Consider the following example: Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, with $A = \{a, b, c\}$, $R_{wkP}^{\mathbb{W}} = \{(b, a), (c, a)\}$ and $R_{bd}^{\mathbb{W}} = R \setminus R_{wkP}^{\mathbb{W}}$. It holds that $lab^{\mathbb{W}}(a) = out$, and $a \in A_{\overline{pers}}^{\mathbb{W}}$, because if the weights of both weak attacks become negative, then the label of a becomes in. Assume that i can reinforce (b, a) while j can reinforce (c, a). Then, i persist-dominates j, but i does not reinforce-dominate j. So, in the general case, the converse does not hold.

We now provide an example illustrating how we can choose a single expert, from a pool of experts, who can make the debate as uncontroversial as possible, with respect to the criteria of attack stability and argument persistence.

Example 7 (cont.)



The PC chair is worried that the authors of the paper will not be convinced by the current decision, as two of the three attacks are weak, and the argument proposing the acceptance of the paper (a) is non-persistent. So, the question is which expert to choose in order to make the decision uncontroversial. Here are some available experts (strong attacks are in bold), together with the consequences of their (potential) votes.

Expert	(c,b): ⟨ **5,6** ⟩ (strong)		(b,a): $\langle -1,9\rangle$ (weak)		(d,a): $\langle 2,3\rangle$ (weak)	
	$s=+1$	$s=-1$	$s=+1$	$s=-1$	$s=+1$	$s=-1$
1: $\{comp, ml\}$	⟨ **8,9** ⟩	$\langle 2,9\rangle$	$\langle 0,12\rangle$	$\langle -2,12\rangle$	$\langle 2,3\rangle$	$\langle 2,3\rangle$
2: $\{comp, kr\}$	⟨ **7,9** ⟩	$\langle 3,9\rangle$	$\langle 1,12\rangle$	⟨ **-3,12** ⟩	⟨ **4,6** ⟩	$\langle 0,6\rangle$
3: $\{comp, cog\}$	⟨ **7,9** ⟩	$\langle 3,9\rangle$	$\langle 1,12\rangle$	⟨ **-3,12** ⟩	$\langle 3,6\rangle$	$\langle 1,6\rangle$
4: $\{ml, kr\}$	⟨ **6,9** ⟩	⟨ **4,9** ⟩	$\langle 0,12\rangle$	$\langle -2,12\rangle$	⟨ **4,6** ⟩	$\langle 0,6\rangle$
5: $\{ml, cog\}$	⟨ **6,9** ⟩	⟨ **4,9** ⟩	$\langle 0,12\rangle$	$\langle -2,12\rangle$	$\langle 3,6\rangle$	$\langle 1,6\rangle$
6: $\{cog, kr\}$	⟨ **5,6** ⟩	⟨ **5,6** ⟩	$\langle 1,12\rangle$	⟨ **-3,12** ⟩	⟨ **5,6** ⟩	$\langle -1,6\rangle$

First, the PC chair observes that expert 1 is necessarily reinforce-dominated by experts 4, 5, and 6. So certainly 1 must not be chosen. No other expert is necessarily reinforce (strictly) dominated in this example. For instance, expert 3 is not necessarily reinforce-dominated by expert 6, because if 3 votes negatively on (b,a) while 6 votes positively on this attack, the GB reached by expert 6 is not strictly better, in the sense of Pareto. Next, expert 6 reinforce-dominates all the other experts, as he can reinforce both (b,a) and (d,a), and he cannot weaken (c,b). No expert reinforce-dominates expert 6, for instance, expert 2 can weaken (c,b), and expert 4 cannot reinforce (b,a). Interestingly, expert 2 optimistically reinforce-dominates expert 4, but is pessimistically reinforce-dominated by the same expert. Finally, both expert 4 and expert 6 persist-dominate all the other experts (as they can turn a persistent, and they cannot turn b non-persistent).

We will not commit to a specific way of choosing an expert (as several choices exist), but let us propose an informal way to do this, based on the criteria of attack stability and argument persistence: choose the expert who reinforce-dominates most of the others (e.g. expert 6 in Example 7); if the choice is not clear, then choose the expert who persist-dominates most of the others; if the choice is not clear, then focus on either the optimistic or the pessimistic dominance relations, depending on context.

Dominance based on argument decidability

Now we analyze the third criterion of a debate's controversy, its sets of decided and undecided arguments.

The ability of experts to influence an argument's label depends on their ability to "change" some attacks of the system. In the following, we say that an expert can "change" an attack if and only if, by voting on that attack, he is able to change the sign of its weight from positive to non-positive (or vice-versa). In other words, an expert can "change" an attack if and only if, by voting on it, he is able to add/remove that attack from the counterpart system of the GB.

We start our analysis by introducing the change-dominance relation, which takes into account the attacks that the experts can change. This dominance relation has no optimistic or pessimistic versions.

Definition 51 *Let $\mathbb{W} = \langle A, R, Eval\rangle$ be a GB, and let i, j be two experts. We say that i* **change-dominates** *j on $\mathbb{W}$ if and only if $R_j \subseteq R_i$, where R_j (resp. R_i) is the set of attacks that j (resp. i) can change (add or remove from the counterpart of $\mathbb{W}$).*

By changing the attacks of a system, an agent can turn some arguments decided, and turn some others undecided.

Definition 52 *Let $\mathbb{W} = \langle A, R, Eval\rangle$ be a GB, let i be an expert, and let $\mathbb{W}_i \in \text{Poss}_i(\mathbb{W})$. We say that the set of arguments $A^{\mathbb{W}_i}_{dec}$* **are turned decided** *in $\mathbb{W}_i$, and that the set of arguments $A^{\mathbb{W}_i}_{undec}$* **are turned undecided** *in $\mathbb{W}_i$.*[14]

Property 18 states that if expert i can turn persistent argument a, and he can turn persistent argument b, then he can turn persistent both arguments a and b, in some system $\mathbb{W}_i \in Poss_i(\mathbb{W})$. If instead of turning arguments persistent, the agent wants to turn arguments decided, the above property does not hold, in the general case, as the following example illustrates.

Example 27 *Let $\mathbb{W} = \langle A, R, Eval\rangle$ be a GB with $A = \{a, b, c, d, e, f, g, h\}$. We abstract from the content of the attacks' evaluation vectors, but we assume that we have: $R^{\mathbb{W}}_{wkN} = \{(a, b), (b, a)\}$, $R^{\mathbb{W}}_{wkP} = \{(e, g), (f, h)\}$, $R^{\mathbb{W}}_{strP} = \{(a, c), (c, e), (g, e), (b, d), (d, f), (h, f)\}$, while all the other attacks are beyond any doubt (no agent has voted on them). The weak attacks with negative (resp. positive) weights are illustrated in dotted (resp. normal) arrows, and the strong attacks are illustrated in thick arrows. The system is illustrated in Figure 5.13. We can see that $A^{\mathbb{W}}_{undec} = \{e, g, f, h\}$.*

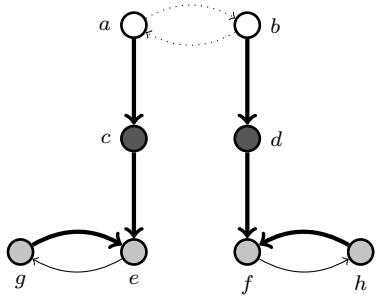


Figure 5.13: Some arguments in the above system can be turned decided. But the opposite can happen as well.

Now, let i be an agent who can change the attacks (a, b) and (b, a), and turn their weights positive. Which arguments can be turned decided, and which ones can be turned undecided by agent i?

1. *If i changes the attack (a, b), then he turns decided all the arguments, except from e and g.*
2. *If i changes the attack (b, a), then he turns decided all the arguments, except from f and h.*
3. *If i changes both attacks (a, b) and (b, a), then something very different happens: he turns undecided all the arguments of the system.*

Therefore, if an agent i can turn decided an argument (e.g. argument e), and also another argument (e.g. argument f), then, in the general case, there is no $\mathbb{W}_i \in Poss_i(\mathbb{W})$ where both arguments are turned decided.

Let us now define two dominance relations over experts, based on their ability to turn arguments decided (and undecided). The first relation is the decide(1)-dominance relation, which compares the systems that agents can bring about. The second relation is the decide(2)-dominance relation, which compares the sets of arguments that agents can turn decided (and undecided), but not obligatorily at the same time.

[14] We say that an argument is *turned decided* or *turned undecided* regardless of its initial label.

We begin with the decide(1)-dominance relation. Roughly, we say that expert i decide(1)-dominates expert j on $\mathbb{W}$ if and only if: for any system that j can bring about, i can bring about a system with a superset of decided arguments (optimistic dominance), while for any system that i can bring about, j can can bring about a system with a superset of undecided arguments (pessimistic dominance).

Definition 53 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, and let i, j be two experts. We say that i* **optimistically decide(1)-dominates** *j on $\mathbb{W}$ iff: $\forall \mathbb{W}_j \in Poss_j(\mathbb{W})$ it holds that $\exists \mathbb{W}_i \in Poss_i(\mathbb{W})$ with $A_{dec}^{\mathbb{W}_j} \subseteq A_{dec}^{\mathbb{W}_i}$. We say that i* **pessimistically decide(1)-dominates** *j on $\mathbb{W}$ iff: $\forall \mathbb{W}_i \in Poss_i(\mathbb{W})$ it holds that $\exists \mathbb{W}_j \in Poss_j(\mathbb{W})$ with $A_{undec}^{\mathbb{W}_i} \subseteq A_{undec}^{\mathbb{W}_j}$. We say that i* **decide(1)-dominates** *j on $\mathbb{W}$ iff i optimistically and pessimistically decide(1)-dominates j on $\mathbb{W}$.*

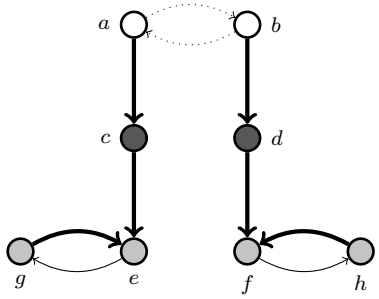
We now define the decide(2)-dominance relation. This dominance relation does not compare the systems that agents can bring about, but instead it compares the sets of arguments that agents can turn decided (and undecided), not obligatorily at the same time.

Definition 54 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB, and let i, j be two experts. We say that i* **optimistically decide(2)-dominates** *j on $\mathbb{W}$ iff: $A_i = \{a \in A \mid i$ can turn a decided $\}$, $A_j = \{a \in A \mid j$ can turn a decided $\}$, and it holds that $A_j \subseteq A_i$. We say that i* **pessimistically decide(2)-dominates** *j on $\mathbb{W}$ iff: $A'_i = \{a \in A \mid i$ can turn a undecided $\}$, $A'_j = \{a \in A \mid j$ can turn a undecided $\}$, and it holds that $A'_i \subseteq A'_j$. We say that i* **decide(2)-dominates** *j on $\mathbb{W}$ iff i optimistically and pessimistically decide(2)-dominates j on $\mathbb{W}$.*

Let us illustrate these two dominance relations with the help of an example:

Example 5.5.3 (cont.) *Let i, j and k be three experts such that:*

- *Expert i can change (a, b) and (b, a).*
- *Expert j can change (e, g) and (f, h).*
- *Expert k can change (a, b), (b, a), (e, g) and (f, h).*



We begin with expert i. As said before, i can turn g, e decided (by changing (b, a)), and he can also turn f, h decided (by changing (a, b)), but he cannot turn all four arguments g, e, f, h decided, at the same time. On the other hand, if expert j changes both attacks (e, g) and (f, h), then he can turn all arguments decided. Notice that expert j cannot bring about any new undecided arguments. Finally, expert k has more choices, as he can change four attacks. Two extreme choices he has are the following: (i) to change (a, b) and (b, a), thus turning all arguments undecided, and (ii) to change (e, g) and (f, h), thus turning all arguments decided.

As far as the dominance is concerned, let us start by the change-dominance relation. We note that expert k change-dominates both i and j, while no other expert change-dominance another one.

Now we pass to the decide-dominance relations. We first compare expert j, with experts i and k. We remind that j can potentially turn all the arguments decided, and he can turn undecided

(at most) the set of arguments $\{e, g, f, h\}$ (if he changes no attacks). It holds that j decide(1)-dominates i and k, because j can bring about a system "better" than the "best" systems i and k can bring about, while j cannot bring about a "worse" system than the "worst" systems i and k can bring about. Moreover, j decide(2)-dominates i and k, because no expert can turn decided more arguments than j (he can turn decided all arguments in A), while experts i and k can turn undecided all the arguments that j can turn undecided (e, g, f, h) and even more. For the above reasons, no expert decide(1)-dominates (neither decide(2)-dominates) j.

Finally, we compare experts i and k. It holds that k decide(1)-dominates (and decide(2)-dominates) i. Interestingly, i does not decide(1)-dominate k, because i cannot turn all arguments decided, at the same time (something that k can do, by changing (e, g) and (f, h)). But, i decide(2)-dominates k, because i and k can turn decided (and undecided) exactly the same sets of arguments.

Let us provide a property showing a link between the change-dominance and the decide(1)-dominance relations.

Property 20 *If expert i change-dominates expert j on a GB $\mathbb{W}$, then i optimistically decide(1)-dominates j on $\mathbb{W}$. On the contrary, if expert i decide(1)-dominates expert j, then i does not change-dominate j, in the general case.*

Proof 20 *($\rightarrow$) Let expert i change-dominate expert j on a GB $\mathbb{W}$. Then, for every system $\mathbb{W}_j \in Poss_j(\mathbb{W})$ that j can bring about, since i change-dominates j, i can also bring about a system by changing exactly the same attacks. Therefore, i optimistically decide(1)-dominates j on $\mathbb{W}$. ($\leftarrow$) As we have seen in 5.5.3, j decide(1)-dominates k, but j does not change-dominate k. So, if expert i decide(1)-dominates expert j, then i does not change-dominate j, in the general case.*

Finally, let us provide a property showing a link between the decide(1)-dominance and the decide(2)-dominance relations.

Property 21 *If expert i decide(1)-dominates expert j on a GB $\mathbb{W}$, then i decide(2)-dominates j on $\mathbb{W}$. The converse does not hold, in the general case.*

Proof 21 *($\rightarrow$) Let expert i decide(1)-dominate expert j. Assume that the systems experts i and j can bring about are, respectively, $Poss_i(\mathbb{W}) = \{\mathbb{W}_{i1}, \ldots, \mathbb{W}_{im}\}$ and $Poss_j(\mathbb{W}) = \{\mathbb{W}_{j1}, \ldots, \mathbb{W}_{jn}\}$. First, for every system $\mathbb{W}_{jx} \in Poss_j(\mathbb{W})$ that j can bring about, there is a system $\mathbb{W}_{ix} \in Poss_i(\mathbb{W})$ that i can bring about, such that $A_{dec}^{\mathbb{W}_{jx}} \subseteq A_{dec}^{\mathbb{W}_{ix}}$. As a result, $\bigcup_{x=1}^{n} A_{dec}^{\mathbb{W}_{jx}} \subseteq \bigcup_{x=1}^{n} A_{dec}^{\mathbb{W}_{ix}}$. Similarly, for every system $\mathbb{W}_{iy} \in Poss_i(\mathbb{W})$ that i can bring about, there is a system $\mathbb{W}_{jy} \in Poss_j(\mathbb{W})$ that j can bring about, such that $A_{undec}^{\mathbb{W}_{iy}} \subseteq A_{undec}^{\mathbb{W}_{jy}}$. As a result, $\bigcup_{y=1}^{n} A_{undec}^{\mathbb{W}_{iy}} \subseteq \bigcup_{y=1}^{n} A_{undec}^{\mathbb{W}_{jy}}$. Thus, i decide(2)-dominates j.*

($\leftarrow$) In example 5.5.3, expert i decide(2)-dominates expert k, but i does not decide(1)-dominate k. Therefore, the converse does not hold, in the general case.

We hold that the three proposed criteria for measuring debate controversy (attack stability, argument persistence, and argument decidability) are intuitive and they can be useful for the analysis of real debates.

Also, we have proposed a number of dominance relations over experts, which can help us choose the expert who is more likely to stabilize a debate. We would like to leave the freedom to the debate administrators (or users) to decide which criterion (or which combination of criteria) is the more suitable for the specific debate setting they have in mind. For example, in some cases optimistic dominance relations could be considered more suitable than pessimistic ones, while in other cases the opposite could happen. Also, it could be argued, for example, that in some cases argument persistence is more (or less) important than argument decidability. Moreover, it

is entirely possible to simultaneously focus on two (or even on all three) controversy criteria, and to use at the same time more than one dominance relation in order to choose an expert.

Finally, a practical issue that should be taken into account is the following: are we equally interested in all the arguments and attacks (as it has been the case in the dominance relations we have defined), or are we more interested in specific arguments and specific attacks? In the latter case, variations of the dominance relations could be defined.

At this point we have finished the analysis of the dynamics of abstract argumentation systems. Later, in Chapter 7, we will come back to the notion of (minimal) change, as we will focus on the actual procedure of debating, and we will define and analyze several debate protocols and agent strategies.

Part III

Protocols for Argumentative Dialogues

Chapter 6

Background on Argumentative Dialogues

A dialogue's general objective is the communication between its participants. The authors in [MP09] provide an informative overview of the nature and basic elements of dialogues. An important contribution to the study of dialogues has been the categorization of dialogues proposed in [WK95], which is based on: the information the participants have at the beginning of a dialogue (of relevance to the topic of discussion), their individual goals in the dialogue, and their shared goals in the dialogue. In that work several types of dialogues have been identified.

In **Information-Seeking Dialogues** one participant seeks the answer to some question(s) from another participant, who is believed by the first to know the answer(s). In **Inquiry Dialogues** the participants collaborate to answer some question or questions whose answers are not known to any participant. In **Negotiation Dialogues**, the participants bargain over the division of some scarce resource. If a negotiation dialogue terminates with an agreement, then the resource has been divided in a manner acceptable to all participants. Participants of **Deliberation Dialogues** collaborate to decide what action or course of action should be adopted in some situation. Here, participants share a responsibility to decide the course of action, or, at least, they share a willingness to discuss whether they have such a shared responsibility. As with negotiation dialogues, if a deliberation dialogue terminates with an agreement, then the participants have decided on a mutually-acceptable course of action. In **Eristic Dialogues**, participants quarrel verbally as a substitute for physical fighting. Finally, **Persuasion Dialogues** involve one participant seeking to persuade another to accept a proposition he or she does not currently endorse. In our work we will focus on this type of dialogues.

In the context of multi-agent systems, the need to standardize dialogues has led to the definition of languages for agent communications, such as DARPA's Knowledge Query and Manipulation Language (KQML) [FFMM94] and the Foundation for Intelligent Physical Agents' (now IEEE FIPA) Agent Communications Language (FIPA ACL) [FIP08]. These languages have been inspired by the typology of dialogues in [WK95].

Every dialogue language has three basic components: syntax, semantics and pragmatics.

- **Syntax:** The syntax of a language concerns the surface form of words and phrases, and how these may be combined. Accordingly, defining the syntax of a dialogue language usually involves the specification of the possible utterances which agents can make and the rules which specify the allowed order of utterances. A rigorous definition of the syntax allows to formally study the properties of dialogue languages and protocols.
- **Semantics:** Dialogues are a means of communication between agents. Therefore, apart from

being formal constructs obeying to a specific syntax, they must also convey some meaning which is understood and shared by the dialogue's participants and designers. In that way, dialogues can be used in open, distributed agent systems, as parts of real applications.

- **Pragmatics:** In [MP09] the study of language pragmatics is viewed as dealing with those aspects of linguistic meaning not covered by considerations of truth and falsity. These are basically the desires and intentions of speakers which are usually communicated by means of speech acts. Examples of speech acts are utterances in which a speaker proposes that some action be undertaken, or promises to undertake it, or commands another to perform it. In [Aus75, Sea69] spoken utterances were classified by their intended and actual effects on the world (for example the internal mental states of their hearers), and pre-conditions for the realization of these effects were developed. We note that, in our work, we will not study dialogue language pragmatics.

The use of argumentation can be valuable in any type of dialogue, and it is especially important in persuasion dialogues, because rational agents get convinced to adopt a position, only if they have good reasons to do it. But, the addition of argumentation in the context of a dialogue increases its complexity. Some important issues which are raised are the following: arguments must be formally defined, the dialogue's protocol must address issues related to the exchange of arguments, disagreement over arguments and their relations is possible among the participants, and there must exist a way of drawing conclusions, based on all the arguments which have been exchanged during the debate.

In this chapter, we provide an overview of the literature on argumentative dialogues, focusing on persuasion dialogues.

In Section 6.1, we focus on the works of Prakken in [Pra05, Pra06] which analyze the basic elements of argumentative dialogues. Both [Pra05, Pra06] focus on persuasion dialogues and more specifically on disputes, where two entities, a proponent and an opponent, have contradicting opinions over a given issue, and each one tries to convince the other. A communication language is used where different locutions have different meanings. Though we present some possibilities in the literature, we will later focus on argument games where the agents are only allowed to send arguments and state relations over arguments. Commitments are usually public statements which force an agent to act in a way that respects his previous utterances. We will insist on the notion of protocol which is a set of rules designating, at every step of the dialogue, which agents are allowed to play moves and what types of moves are allowed. Key aspects of a protocol are the notions of *turn-taking*, *current winner*, and *allowed moves*. Next, we turn our attention to the behaviour of the agents who take part in an argumentative dialogue. When following a specific dialogue protocol, at some points, they may be given the possibility to chose their next utterance among a set of possible ones. In those cases, the agents need a way to decide their utterance: they need a strategy. We will provide some basic definitions related to agent strategies, in the context of Game Theory, as presented in [RL09].

In Section 6.2, we are interested in the more specific setting of multilateral argumentative dialogue games. Though many existing works, as [Pra05, Pra06] do not exclude the participation of multiple entities, not many works specifically study multilateral settings, and the new issues they raise. The work in [BM11] is one of the first to define a specific setting for multilateral debates and to answer specific questions about them. Our work has been inspired by the work in [BM11].

In Section 6.3, we provide a typology of multilateral argumentative dialogue games. We begin with a basic typology presented in [MP09], we analyze the choices made in the protocol of [BM11], and then we enrich this typology by proposing specific choices which can be made for the different protocol elements (and the agents' strategies).

Finally, since we are interested in argumentative dialogues which can be deemed valuable in practice, in Section 6.4 we present a set of criteria, from the literature, which can be used to evaluate argumentative dialogues. More specifically, we present some criteria to evaluate the

quality of argumentative dialogues, proposed in [MPW02], and some criteria to evaluate persuasion dialogues, proposed in [AdSC13]. Also, we highlight the basic criteria for evaluating the complexity of such dialogues, proposed in [TG10].

6.1 Basic Elements of Argumentative Dialogues

In [Pra05] and [Pra06], Prakken provides an informative overview of the main components of argumentative dialogues. In the title of [Pra05] ("Coherence and flexibility in dialogue games for argumentation"), the term *coherence* indicates that an argumentative dialogue must be focused on its goal, while the term *flexibility* indicates that, at the same time, the participants of such a dialogue must be given some freedom to choose their actions.

The author proposes a general framework for argumentative dialogue which has the form of a game. The main assumptions of [Pra05] and [Pra06] are the following:

- A dialogue focuses on some topic[1] tp, which is a proposition of a language L_{tp}.
- There are two parties arguing about tp: a proponent (P) who wants tp to be accepted, and an opponent (O) who wants tp to be rejected.
- There is a set of outcome rules ($\mathcal{O}$) which define, at any point of the dialogue, the winner and loser parties, given: (i) the dialogue itself so far, (ii) the main claim of the dialogue, and (iii) the background knowledge upon which all the participants agree.

While the analysis provided in [Pra05, Pra06] is useful in different dialogue types (e.g. in deliberation dialogues), it is clear from the above points that the author focuses on persuasion dialogues, where two parties with conflicting opinions over a specific topic, try to persuade each other about the correctness of their opinion.

6.1.1 Persuasion dialogues

In [WK95] persuasion dialogues are defined as dialogues where the goal is to resolve a conflict of points of view between at least two participants by verbal means. A point of view with respect to a proposition can be positive, negative or of critical doubt. A participant's individual aim is to persuade the other participant(s) to adopt his personal point of view. According to [WK95], a conflict of points of view is resolved if all parties share the same point of view on the proposition that is the topic of the conflict. Walton and Krabbe distinguish *disputes* as a subtype of persuasion dialogues where two parties disagree about a single proposition, such that at the start of the dialogue one party has a positive and the other party a negative point of view towards that proposition. Therefore, it is evident that persuasion dialogues have their own types of dialogue goals (both global and individual), roles of the participants, and possible outcomes.

Usually, in persuasion dialogues, in order for the parties to achieve their goal of convincing the others, they are able to construct and send arguments. In some settings, the agents' beliefs may be dynamic. Then, an agent may start the dialogue as a proponent of a proposition, but then he may be convinced by the other party's arguments, and thus finish the dialogue as an opponent. In these cases persuasion is evident, because the opinions of the agents are subject to change. On the other hand, there are settings where the agents' beliefs are static. Even in these settings, such argumentative dialogues are usually considered to be persuasion dialogues, in the sense that the winner of the debate is the one who has "persuaded" a rational third party (an external observer) of the debate that his opinion was the correct one. In this respect, in [Pra06] persuasion dialogues are separated into *pure persuasion dialogues* and *conflict resolution dialogues*. In pure persuasion dialogues the winning party must force the other one to adopt his point of view on a proposition,

[1] Notice that, in our work, the term topic has a different meaning than in [Pra05]. At this point, we would have used the word "issue" instead.

while in conflict resolution dialogues this is not obligatory (e.g. in a trial, the accused can be found guilty, even if he does not concede that he is).

In the legal domain, persuasion dialogues can take place in the processes of *arbitration* and *mediation* which differ conceptually, as explained in [BLZ04]. Arbitration is essentially conflict (dispute) resolution, as described above. It is a process in which a neutral third party (the arbitrator) renders a decision after a hearing at which all parties have the opportunity to be heard. On the other hand, mediation differs, in the sense that the third party's role is to help and not to judge. Mediation is a private, informal dispute resolution process in which a neutral third party (the mediator) helps disputing parties to reach an agreement. The mediator has no power to impose a decision on the parties. In mediation, negotiation is usually vital in order to reach an agreement.

In [Pra06] a list of persuasion systems is provided, with a brief analysis of each one's characteristics. Finally, *argument games* are the type of persuasion systems which mainly interest us in this work. The basic feature of argument games, is that the proponent and the opponent move *arguments and nothing else*. There are no additional propositional attitudes such as claiming, for example. The basic structure of argument games is the following: The proponent begins with an argument for a claim he wants to defend. At each other move, the opponent replies with counterarguments that undermine the acceptance of the claim, while the proponent replies with counterarguments that support the acceptance of the claim. Of course, argument acceptance depends on the chosen semantics.

6.1.2 Dialogue system for argumentation

Let us see how Prakken defines a dialogue system for argumentation in [Pra05].

Definition 55 ([Pra05]) *A* **dialogue system proper** *is a triple $\mathcal{D} = (L_c, Pr, C)$ where L_c (the communication language) is a set of locutions, Pr is a protocol for L_c, and C is a set of effect rules of locutions in L_c, specifying the effects of the locutions on the participants' commitments.*

Definition 56 ([Pra05]) *A* **dialogue system for argumentation** *is a pair $(\mathcal{L}, \mathcal{D})$ where $\mathcal{L}$ is a logic for defeasible argumentation and $\mathcal{D}$ is a dialogue system proper.*

A logic for defeasible argumentation $\mathcal{L}$ is needed because the main feature of an argumentative dialogue is, naturally, the exchange of arguments. Therefore we need to be able to construct arguments, to define attack relations between them, and finally to compute the sets of accepted (and rejected) arguments, once the dialogue has finished. On the other hand, the dialogue system proper $\mathcal{D}$ defines the rest of the elements of the dialogue.

Communication language

A communication language L_c contains a set of locutions *Loc*. A locution $l \in Loc$ is of the form $p(c)$, where p is an element of a given set P of performatives and c either is a member or subset of a language L_{tp}, or an argument. Different performatives correspond to different types of locutions. For example, we may have the following performatives:

- $claim(\phi)$: Claiming that ϕ holds.
- $why(\phi)$: Asking for an argument supporting ϕ.
- $argue(A)$: Sending an argument A.
- $concede(\phi)$: Accepting that ϕ holds.
- $retract(\phi)$: Retracting ϕ, which had previously been claimed.

Some types of locutions can be replies to specific previous locutions in a dialogue. For example, $why(\phi)$ can be a reply to a previous $claim(\phi)$ locution, thus putting into question the validity of ϕ. Also, if argument A has conclusion ϕ, then the locution $argue(A)$ can be a reply to $why(\phi)$, explaining why ϕ holds.

In order to define a dialogue protocol Pr, [Pra05] first defines the basic notion of *move*. A move m in a dialogue has four basic elements:

- $id(m)$: the identifier of the move (a number),
- $pl(m)$: the player of the move (proponent or opponent),
- $s(m)$: the speech act performed in the move,
- $t(m)$: the target of the move (a previous move in the dialogue).

A series of moves from the agents consists in a *dialogue*. We note that in [Pra05], it is prohibited to have several moves played at the same time, by different players. This is a constraint we also consider in our work. Now, we can see what a dialogue protocol consists of.

Dialogue protocol

A crucial element of a dialogue is the protocol Pr which dictates how the dialogue may proceed, at every point. Roughly, a protocol is a function which, at every point of a dialogue (timestep), returns the set of possible moves for the agents. As said before, a protocol must balance two opposing elements: the dialogue's coherence and flexibility. Therefore, a protocol should only allow moves which are, in some way, relevant to the issue, but at the same time it should provide some freedom to the participants in order to decide their next move from a set of possible moves.

A crucial element of a protocol, sometimes considered apart, is the *turn-taking* function. The turn-taking function takes as input the dialogue up to some point, and it returns the player(s) who are to play next. In the general case, the turn-taking function can designate, at some point, more than one agents. Also, in some cases, a player may be allowed to play, at some point, multiple moves, the one after the other. In that case, this series of moves by the same agent, is called a *turn* of the dialogue.

Let us provide an example of a dialogue protocol proposed in [Pra05]. It has the form of five rules which must be respected, whenever the agents are to choose their next move. The rules are presented in a non-formal way:

1. **Rule 1:** A move is legal only if moved by the player-to-move (designated by the turn-taking function).
2. **Rule 2:** A replying move must be a proper reply to its target, according to L_c.
3. **Rule 3:** One cannot reply to one's own moves.
4. **Rule 4:** If a player backtracks, the new move must be different from the first one ("backtracking" here means any alternative reply to the same target in a later turn).
5. **Rule 5:** Surrenders (declaring acceptance of a proposition) may not be revoked (taken back).

Moreover, in [Pra05] the author distinguishes four categories of dialogue rules. (1) *Basic protocol rules* should be respected in all discussions. (2) *Context-dependent protocol rules* hold only in specific contexts of application. (3) *Conventions* formulate behaviour that participants should ideally have to promote coherence of the dialogue. (4) *Player strategies and heuristics* are meant to promote each player's individual goal.

Commitments

Finally, the third element of a dialogue system proper $\mathcal{D} = (L_c, Pr, C)$ is the commitment function C. It assigns, at every point of a dialogue, to every agent, all the propositions to which the agent has committed, until that point. A dynamic structure, called commitment store, contains all the commitments of an agent. Commitments are public statements the agents have already made during the dialogue, so they are expected to not contradict them later, unless they choose to revoke them. In that case, they should retract the corresponding propositions from their commitment stores.

Commitments are crucial in social interactions, for many reasons: First, they provide a simple way for agents to know, *a priori*, what to expect from other agents in the future. Usually, an agent expects from another agent to act according to the commitments he has already made. Second, commitments provide the means to an agent to evaluate his trust on other agents. Usually, the more an agent contradicts or revokes his commitments, the less he will be trusted by other agents.

A commitment store can be used in order to identify direct contradictions (e.g. an agent has committed to both ϕ and $\neg\phi$). A concrete example could be an agent committing that the attack (a, b) holds, and also committing that it does not hold. A commitment store can be also used to identify indirect contradictions (e.g. from the commitments of an agent, both ϕ and $\neg\phi$ can be inferred). The dialogue's designer can choose how these cases should be treated, for example if such a contradiction is found, then the agent in question could be asked to act immediately and remove the contradiction. Another important use of commitment stores is to avoid repetitions of the same move, by the same agent, thus helping to ensure dialogue termination.

Dialogue termination and winners

According to [Pra05], the most usual termination criterion of a dialogue is the following: the dialogue terminates if and only if the opponent concedes the proponent's main claim, or the proponent retracts his main claim. In another, more "mathematical" approach, a dialogue terminates if and only if no legal continuation is possible. Realistic dialogues will not usually terminate by retraction or concession of the main claim (as usually no party admits loss), but by external agreement or decision to terminate it.

In practice, it is important to know, at every point of the debate, which would be the winner, in case the dialogue was to stop at that exact point. Therefore, so called *"anytime" outcome definitions*, are needed. An anytime outcome definition can help us define the notions of "current winner", "relevant move", as well as the turn-taking.

There are two main approaches which can be used to compute the current winner of a dialogue.

- Take all the premises of the arguments in the dialogue (which are not challenged or retracted). If the main claim can be justified based on these premises, then the proponent is the current winner, otherwise the opponent is the current winner.
- Construct a tree (or, more generally, a graph) from the agents' locutions. Given the tree's (or graph's) structure, the acceptability of the main claim can be decided, in some way.

An interesting question is the following: Under which conditions do the two approaches compute the same winner? If the second approach always computes the same winner as the first one, then the dialogue is said to be *sound*. Reversely, if the first approach always computes the same winner as the second one, then the dialogue is said to be *fair*.

In [Pra05] a dialogue is represented by a *dialogue tree* which indicates, for every move in the dialogue, all its replies in the debate. Roughly, the root of the dialogue tree contains the main claim of the debate. In order to decide on the acceptability of the main claim, we can label the tree's nodes as proposed in [Pra05]: A node is labelled *in* if it has no children, or if all its children are labelled *out*, while a node is labelled *out* if at least one of its children is labelled *in*. Finally, if the root is labelled *in*, then the proponent wins; while if the root is labelled *out*, the opponent wins. Notice that, since we label a tree, every node is either *in* or *out* and there is no need for a third label (*undec*) as in [Cam06].

Let us see a simple example of an argument game, where all utterances are assertions of arguments as replies (attacks) to previous arguments.

Example 28 *The following arguments a, b, c, d have been put forward in a debate. Each argument is a tuple consisting of a set of premises and a conclusion, in propositional logic.*
$a = \langle\{NiceCar,\ NiceCar \rightarrow BuyCar\},\ BuyCar\rangle$
$b = \langle\{\neg SafeCar,\ \neg SafeCar \rightarrow \neg BuyCar\},\ \neg BuyCar\rangle$
$c = \langle\{ExpensiveCar,\ ExpensiveCar \rightarrow \neg BuyCar\},\ \neg BuyCar\rangle$
$d = \langle\{ReportX,\ ReportX \rightarrow SafeCar\},\ SafeCar\rangle$

Argument a is the main claim of the debate and it was asserted by PRO (who wants the claim to be accepted). Arguments b and c were asserted by OPP (who wants the claim to be rejected), as replies to a. Finally, argument d was asserted by PRO, as reply to b. The following dialogue tree illustrates all the utterances (arguments, in this case) and their relations (which utterances are replies to which others).

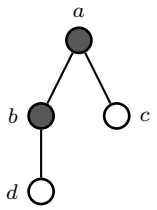


Figure 6.1: A dialogue tree. *in* arguments (c, d) are in white, *out* arguments (a, b) are in black.

Following the labelling method proposed in [Pra05], arguments c, d are in, while arguments a, b are out. Using an anytime outcome definition we would say that, since argument a (the main claim) is out, OPP is the current winner of the debate.

Relevant moves

In order to keep a dialogue focused on the acceptability of its main claim, it is essential to make sure that every move is in some way "relevant" to it. In [Pra05] the author introduces the notions of strongly relevant and weakly relevant move.

Roughly, a move in a dialogue d is *strongly relevant* iff it changes the dialogical status of d's initial move. On the other hand, a move in d is *weakly relevant* iff it either "reinforces" the winning position of the speaker or if it "weakens" the winning position of the hearer. Therefore, in protocols allowing only strongly relevant moves, every played move changes the current winner. On the other hand, in dialogues allowing weakly relevant moves, multiple moves may be played by an agent (or group of agents) before the current winner changes.

Example 28 (cont.) *Let us continue the previous example, by considering two additional arguments:*
$e = \langle\{Offer,\ Offer \implies \neg ExpensiveCar\},\ \neg ExpensiveCar\rangle$
$f = \langle\{BoringCar,\ BoringCar \implies \neg BuyCar\},\ \neg BuyCar\rangle$

In Figure 6.2 we can see what happens if argument e is uttered:

Argument e can be uttered as a reply to argument c. In that case, the label of argument a changes and becomes in. Therefore uttering argument e is a strongly relevant move at this point of the debate, and it makes PRO the current winner.

In Figure 6.3 we can see what happens if argument f is uttered, instead of argument e.

Argument f can be uttered (instead of e) as a reply to argument a. Notice that the label of argument a remains out, so uttering f is not a strongly relevant move according to [Pra05]. On

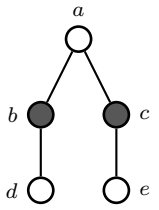


Figure 6.2: Argument e was uttered last, and it turned the main claim's label into in.

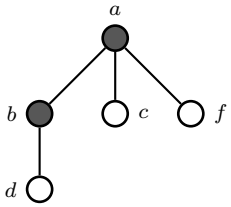


Figure 6.3: Argument f was uttered last, "reinforcing" the status of the main claim.

the other hand, it is a weakly relevant move, in the sense that it reinforces the position of OPP: winning has now become more difficult for PRO, as they must find counter-arguments to both c and f.

6.1.3 Agent Strategies

Agents who play dialogue games desire to increase their chances of winning or, more generally, to maximize their expected utility. In [RL09] the authors study how game theory can be used for that purpose, in the context of argumentative dialogue games. They underline that, in general, game theory can be used to achieve two goals: 1) undertake precise analysis of interaction in particular strategic settings, with a view to predicting the outcome; 2) design rules of the game in such a way that self-interested agents behave in some desirable manner (e.g. tell the truth); this is called mechanism design. They argue that both goals are important in argumentative dialogue games. First, the agents want to know the best way to play those games. Second, the game's analyst or designer wants to know the game's possible outcomes, given rational play by the agents. Finally, the designer may want to create games where the agents are better off stating their true opinions, than lying. For these reasons, it is important to study how game theory can be used for the definition of agent strategies in argumentative dialogue games.

In [RL09] a strategy for agent i (of some agent type $\theta(i)$), denoted $s_i(\theta_i)$, is defined as "a plan that describes what actions the agent will take for every decision that the agent might be called upon to make, for each possible piece of information that the agent may have at each time it is called to act". Therefore, in a debate, an agent who follows a specific strategy knows what to do at every moment, under any conditions.

Let us remind some basic notions of game theory [OR94]:

- Let I denote a set of self-interested agents (who want to maximize their own utility).
- Let $\theta_i \in \Theta_i$ denote the type of agent i which is drawn from some set of possible types Θ_i. The type of an agent represents his private information (e.g. preferences and beliefs).
- An agent's preferences are over outcomes $o \in \mathcal{O}$, where $\mathcal{O}$ is the set of all possible outcomes.

- An agent's preferences can be expressed by a utility function $u_i(o, \theta_i)$ which depends both on the outcome, $o \in \mathcal{O}$, and on the agent's type, θ_i. Agent i prefers outcome o_1 to o_2 when $u_i(o_1, \theta_1) > u_i(o_2, \theta_2)$.

Let Σ_i denote the set of all possible strategies for agent i, and thus $s_i(\theta_i) \in \Sigma_i$. When it is clear from the context, θ_i will be dropped to simplify notation. Let the strategy profile $s = (s_1(\theta_1), \ldots, s_I(\theta_I))$ denote the outcome that results when each agent i is playing strategy $s_i(\theta_i)$. As a notational convenience we define $s_{-i}(\theta_{-i}) = (s_1(\theta_1), \ldots, s_{i-1}(\theta_{i-1}), s_{i+1}(\theta_{i+1}), \ldots, s_I(\theta_I))$ to be the strategy profile including all agent strategies, except the one of agent θ_i.

In game theory, *solution concepts* determine the outcomes that will arise if all agents are rational and strategic. Two of the most well known and used are the solution concepts of *dominant strategy* and of *Nash equilibrium*.

A strategy s_i^* of agent i is dominant if, by using that strategy, the agent is certain to obtain a greater utility, than what he would have obtained by using any other strategy. This holds regardless of the other players' choice of strategies. Formally:

Definition 57 *[OR94] A strategy s_i^* of agent i is a* **dominant strategy** *if $\forall s_{-i}$, $\forall s_i'$, it holds that $u_i(s_i^*, s_{-i}, \theta_i) \geq u_i(s_i', s_{-i}, \theta_i)$.*[2]

Dominant strategies are a strong solution concept, because they require that there exists a best way to play, regardless of the choices of the other agents. A weaker solution concept is the pure strategy Nash equilibrium. A strategy profile $s^* = (s_1^*, \ldots, s_I^*)$ is a pure strategy Nash equilibrium if no agent has incentive to change his strategy, given that no other agent changes his strategy. Formally:

Definition 58 *[OR94] A strategy profile $s^* = (s_1^*, \ldots, s_I^*)$ is a* **pure strategy Nash equilibrium** *if $\forall i, \forall s_i'$ it holds that $u_i(s_i^*, s_{-i}^*, \theta_i) \geq u_i(s_i', s_{-i}^*, \theta_i)$.*

6.2 Multilateral Argumentative Dialogues

There are several works on argumentative dialogues which allow, in principle, the participation of numerous agents. For example, in [Pra05] and [Pra06] the two opponents, PRO and OPP, could be either two agents, or two groups of agents.

Nonetheless, very few works study the specific challenges of multilateral argumentative dialogues. Such dialogues give rise to new types of questions. What types of protocols can coordinate multilateral dialogues? How can turn-taking be defined in multilateral debates, and how does it affect the debate? How can a group's power be measured in a debate? How can a group's agents coordinate in order to maximize their chances of winning? The work in [BM11] defines a framework for multilateral debates, and it was one of the first works to study specific issues of these debates.

Let us highlight the similarities and differences of [BM11], compared to the more general framework of argument games in [Pra05].

We start with its main **similarities** to [Pra05]:

- The agents are partitioned in two groups, PRO and CON, which have contradicting opinions on the acceptability of a given issue. The dialogue that takes place between them is a persuasion dialogue (debate). PRO (resp. CON) agents want the issue to end up being accepted (resp. rejected) in the debate.

[2]For notational convenience, we denote $u_i(s_i^*, s_{-i}, \theta_i)$ the utility of agent i, who is of type θ_i, when he uses strategy s_i^*, while all the other agents use the strategies of the profile s_{-i}.

- The agents are able to state arguments and attacks, which are put on a central structure (argumentation system), called *Gameboard* (GB). During the debate, as new utterances are made by the agents, the GB is progressively expanded.
- The agents have commitment stores, containing their previous utterances.

Its main **differences** to [Pra05] are the following:

- The main claim of the debate is always a single (abstract) argument, called *issue* of the debate, not a proposition of a language L_t (as in [Pra05]).
- Each agent possesses a private abstract argumentation system, which obligatorily contains the issue. An agent belongs to PRO (resp. CON) if the issue is found (resp. is not found) in the grounded extension of his private system.
- The agents may have different opinions on the validity of attack relations. Such disagreements are resolved through voting. During the debate, the agents are able to vote on attacks, either positively or negatively, and this may result to their addition or removal from the common Gameboard.
- The Gameboard is a graph (not a tree) in the general case.

Based on the above general framework, in [BM11] the authors define a specific debate protocol, as follows:

The protocol proceeds in timesteps, with PRO and CON alternating turns. Let A^t_{GB} and R^t_{GB} be respectively the argumentation system, the set of arguments and the set of attack relations on the GB after timestep t. Within the PRO and CON groups no coordination takes place: the agents may for instance play asynchronously and the authority simply picks the first permitted and relevant move before returning the token to the other side. *Permitted* moves are positive assertions of attacks xRy (with $y \in A^t_{GB}$), or contradictions of (already introduced) attacks (with $(x, y) \in R^t_{GB}$). A move is *relevant* at timestep t for a PRO agent (resp. CON agent) if it puts the issue back in (resp. drops the issue from) the grounded extension of the GB, denoted $E_{Gr}(GB^t)$. Moreover, the protocol prevents the repetition of the same move from the same agent. For this reason, each agent has a commitment store which contains the attack relations and the *non*-attack relations he has added on the GB until timestep t. Finally, the agents are assumed to be truthful, so if an agent has (resp. does not have) the attack (x, y) in his private system, then he can only vote for (resp. against) it during the debate. The proposed protocol is as follows:

1. The agents report their individual view on the issue to the central authority, which then assigns (privately) each agent to PRO or CON.
2. In the first timestep of the debate, the issue is on the GB, and the turn is given to CON.
3. Until a group of agents cannot move, we have:
   (a) agents independently propose moves to the central authority;
   (b) the central authority picks the first (or at random) relevant move from the group of agents whose turn is active, updates the GB, and passes the turn to the other group. The update of the GB consists in either adding an attack on the GB (and its attacking argument, if it was not already there), or in removing an attack from the GB. [3]

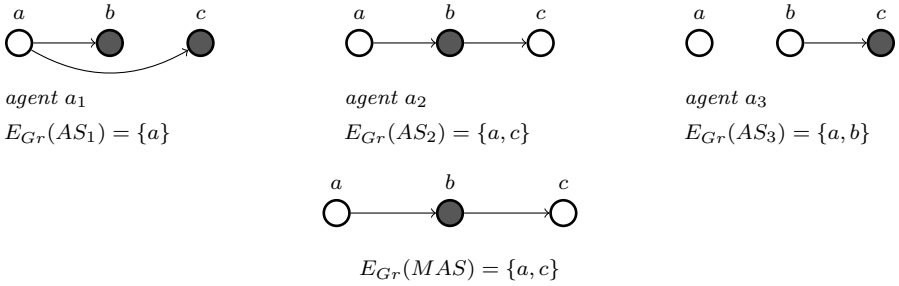
Finally, if group X is active, but it has no relevant moves to propose, then that group is the debate's *loser*, whereas the other group is the debate's *winner*.

[3] The authors consider an additional type of move which "reinforces" an already introduced attack, but we will not refer to this type of move here.

Before providing an example of an actual debate, let us see how the notion of Merged Argumentation System is instantiated. We remind that the Merged Argumentation System is a system which consists in a "rational" aggregation of all the agents' private systems. In [BM11], given the agents' private systems $AS_1 = \langle A_1, R_1 \rangle, \ldots, AS_n = \langle A_n, R_n \rangle$, the Merged Argumentation System is defined by $MAS = \langle \cup_{i=1}^{n} A_i, R_M \rangle$, where $(a, b) \in R_M$ iff the attack relation (a, b) is found in a strict majority of the private systems (in case of tie, the attack is not on the MAS).

Let us see an example of a debate following this protocol:

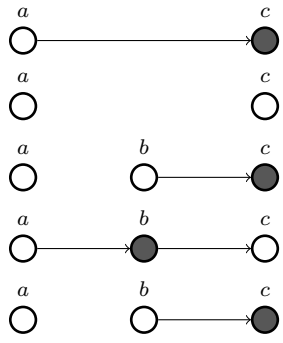
Example 29 *[BM11] Let a_1, a_2, a_3 be three agents and let AS_1, AS_2, AS_3 be their corresponding private systems, illustrated below. The grounded extension of system AS_i is denoted $E_{Gr}(AS_i)$.*



Let the issue of the dialogue be argument c. We have $CON = \{a_1, a_3\}$, $PRO = \{a_2\}$. The Merged Argumentation System (MAS), as defined in [BM11], is illustrated below the three private systems. It contains the attacks (a, b) and (b, c) as both of them are supported by a strict majority of agents (two over three). No other attack is supported by a strict majority of agents. It holds that $\mathcal{E}_{Gr}(MAS) = \{a, c\}$, so the issue c is accepted in the merged system.

Also, let RP_i^t denote the commitment store of agent i at timestep t, containing the attacks on which i has provided his opinion. At the beginning, we have $RP_1^0 = RP_2^0 = RP_3^0 = \emptyset$, $GB^0 = \langle \{c\}, \emptyset \rangle$ and $E(GB^0) = \{c\}$.

A sequence of moves allowed by the protocol is the following:



- $t = 1$ - a_1 *plays for CON:* $RP_1^1 = \{(a, c)\}$
- $t = 2$ - a_2 *plays for PRO:* $RP_2^2 = \{(a, c)\}$
- $t = 3$ - a_3 *plays for CON:* $RP_3^3 = \{(b, c)\}$
- $t = 4$ - a_2 *plays for PRO:* $RP_2^4 = \{(a, b), (a, c)\}$
- $t = 5$ - a_3 *plays for CON:* $RP_3^5 = \{(b, c), (a, b)\}$
- $t = 6$ - *PRO cannot play a relevant move.*

In brief, initially a_1 asserts argument a and attack (a, c) and a_2 responds by removing the attack. Then, a_3 asserts argument b and the attack (b, c). a_2 can still defend his position by adding (a, b) which is a relevant move, making PRO the current winner. But, finally, a_3 is able to remove the attack (a, b) and drop c from the grounded extension. In timestep $t = 6$, the PRO group does not have any relevant moves available (as agents cannot repeat moves), so it cannot change the acceptability status of the issue. The debate has finished, and it holds that $E(GB^6) = \{a, b\}$. Therefore, CON wins the debate, and PRO loses it.

The authors of [BM11], after formally defining the previous debate protocol, they study questions such as: Does the MAS's outcome coincide (sometimes, always, never) with the debate's

outcome? Does the opinion of the biggest group coincide (sometimes, always, never) with the debate's (or the MAS's) outcome? From the previous example, we can already tell that the MAS's outcome, in the general case, does not coincide neither with the biggest group's opinion, nor with the debate's outcome.

Also, the authors define the, so called, *Global Arguments-Control Graph*, which captures the control that PRO and CON have over assertions and removals of attacks. Then, they use it in order to answer questions about the determinism of the debate's outcome, such as: is group X a possible (or certain) winner of the debate?

6.3 A Typology of Protocols for Multilateral Dialogue Games

In this section, we analyze the basic elements of a protocol for a dialogue game, and then we propose a typology of protocols, by exploring several potential choices for these elements.

Roughly, a protocol for a dialogue game designates, at every step of the game, all its possible continuations. For example, it indicates which participant(s) may play next, as well as all their allowed utterances. It also indicates how commitments are updated after every utterance.

In [MP09] the authors present a generic framework for the specification of the key components of protocols for dialogue games. The authors assume that the topics of discussion between the agents (the inner layer of communications) can be represented in some logical language. As far as the wrapper (outer) layer of communications is concerned, a dialogue game specification comprises the following elements:

1. **Commencement Rules:** These rules define the circumstances under which the dialogue commences.

2. **Locutions:** These rules indicate the permitted utterances (e.g. assert propositions, challenge propositions, assert arguments and attacks).

3. **Rules for Combination of Locutions:** These rules define the dialogical contexts under which particular locutions are permitted or not, or obligatory or not. For instance, it may not be permitted for a participant to assert both proposition p and $\neg p$ in the same dialogue (unless he retracts the former).

4. **Commitments:** These rules define under which conditions and how exactly the agents' commitment stores are updated.

5. **Rules for Combination of Commitments:** These rules define how commitments are manipulated in case of utterances leading, for example, to conflicting commitments.

6. **Rules for Speaker Order:** These rules define the order in which speakers may make utterances.

7. **Termination Rules:** These rules define the circumstances under which the dialogue ends.

For every one of the above elements (e.g. rules for speaker order), there is a great variety of choices. Also, a choice regarding a specific element may influence the available choices regarding some other elements (e.g. choosing a specific speaker order may influence the available choices of termination rules).

Let us see the choices made in the protocol of [BM11], as far as the above elements are concerned.

1. **Commencement Rules:** At the beginning of the debate, the GB contains only one argument, the issue, and the CON group gets the token. There are no other commencement rules.

2. **Locutions:** Assertions of arguments, assertions of binary attacks, as well as votes (positive or negative) on already introduced attacks.
3. **Rules for Combination of Locutions:** There are some prohibited types of utterances: 1) repeating an utterance, 2) casting both a positive and a negative vote on the same attack. We also remind that agents are assumed to be truthful.
4. **Commitments:** The agents commit on every utterance they make during the debate. They cannot retract them.
5. **Rules for Combination of Commitments:** There are not any.
6. **Rules for Speaker Order:** They have been described in Section 6.2. PRO and CON groups alternate turns.
7. **Termination Rules:** The debate ends when the active group has no relevant moves to propose.

Now that we have presented the basic elements of debate protocols, we continue by proposing a debate protocol typology. We restrain our scope in protocols for multilateral dialogue games whose framework is similar to [BM11]: we consider the same types of locutions and commitments as those in [BM11]. In this typology we explore the available choices regarding the locutions (permitted utterances) and the rules for speaker order. Additionally, we explore different ways in which the agents may be regrouped (crucial in order to define the debate's winners; it can also affect the rules for speaker order). We finally explore different types of agent strategies. Agent strategies do not form part of debate protocols, but they are related to them, in the sense that they dictate how the agents choose their moves when the protocol gives them the token. Our typology is not comprehensive, as that would be an enormous task outside the scope of this work. Nonetheless, it offers a wide range of possibilities, and it has been used for the definition of the protocols in Chapter 7. We believe that this typology can be of practical interest to a protocol designer, even if its constituents are not necessarily novel.

6.3.1 Agent Groups

In multi-party debate protocols, the participants can be separated in different groups. The motivation behind this separation is usually related to the goals of the agents in the debate. For example, if a subset of agents have the same goal (e.g. to make an argument accepted), then it may be an interesting choice (from the designer's or from the agents' viewpoint) to put them together, in one group. Separation in groups is crucial for the definition of important protocol elements, e.g. turn-taking and current winners. Usually, the agent groups partition the set of agents, as every agent belongs in one group. The separation of the agents in groups depends on the following:

- **Agents' focus**: The agents taking part in a debate focus on achieving a specific goal. In an argument game, typically, the agents' goals concern the acceptability of one (or more) argument(s).
  - **Single issue scenario**: This is the most usual focal point of debating agents in existing works on argumentative debates. All agents are interested in the acceptability status of a specific argument, usually called *issue of the debate*. On the other hand, they are indifferent on the status of the other arguments in the debate.
  - **Multiple issue scenario**: This is a more complex case, as the fact that the agents focus on multiple arguments could make difficult their separation in groups.
- **Types of groups**: In this work, we identify two closely related types of groups: categories and clusters.

- **Categories**: The agents are partitioned in categories when there exists some exogenously given information, indicating the possible groups (categories) in which the agents can be put. Based on that information, the agents can be partitioned in k predefined categories $K_1, \ldots, K_k$. Categories can be defined whether the agents use extension-based argument acceptability, or numerical argument valuation. (i) Assume that the agents argue on whether the (single) issue is accepted under grounded semantics or not. Then, a priori, we are able to define two meaningful categories, the first one containing all the agents believing the issue is accepted, and the second one containing the rest of the agents. (ii) Let the agents argue on the numerical evaluation of the issue. The regrouping of the agents in categories can be done with the help of $k-1$ threshold values (exogenous information), which partition the range of an argument's possible values into k intervals. In the extreme case where $k = 2$, we can see the debate as opposing two groups, PRO and CON, such that: $CON = \{a_i \in N \mid v_i(d) < thr\}$ and $PRO = \{a_i \in N \mid v_i(d) \geq thr\}$. For example, let the issue of a debate be argument d, and let the evaluations of the issue by four agents a_1, a_2, a_3 and a_4, be respectively: $v_1(d) = -0.9, v_2(d) = -0.1, v_3(d) = +0.1, v_4(d) = +0.4$. By setting $thr = 0$, for the above four agents, we would have two categories of agents with the following compositions: $CON = \{a_1, a_2\}$ and $PRO = \{a_3, a_4\}$. The debate protocols which partition the agents in categories are called **category-based protocols**.
- **Clusters**: In this case there is no exogenously given information on how to partition the agents. Instead, in order to partition the agents in groups (clusters) we must examine their private systems and compare them. In this work, we shall define clusters when the agents use numerical argument evaluation.[4] One possibility is to cluster the agents in k groups $(C_1, \ldots, C_k)$ on the basis of the similarity of their evaluations of the issue (in the sense of minimizing the maximum difference between two agents of a group). For example, if in a debate some agents with similar issue evaluations are satisfied with having the issue's evaluation fall close to theirs, then this could motivate them to form a cluster. For the four previous agents, and for $k = 3$, we would have the following three clusters $C_1 = \{a_1\}, C_2 = \{a_2, a_3\}, C_3 = \{a_4\}$; while, for $k = 2$, we would have the following two clusters $C_1 = \{a_1\}, C_2 = \{a_2, a_3, a_4\}$. An interesting extreme case is that of n clusters, where each group is a singleton. In such a case (as we will see), every agent may be guided by the prospect of matching the issue's collective valuation with his private valuation. The debate protocols which partition the agents in clusters are called **cluster-based protocols**.

- **Group Dynamicity**: There are two possibilities, as far as the groups' dynamicity is concerned:
  - **Fixed groups**: The groups are usually fixed, when the agents' private systems are not subject to change. In these cases, their opinions on the arguments' acceptability cannot change. Thus, the agents have no reason to leave their group.
  - **Dynamic groups**: On the other hand, if the agents are able to dynamically update their beliefs during the debate (e.g. add or remove specific arguments and attacks from their systems), this can make them change their opinions on the arguments' acceptability. In these cases, the definition of dynamic groups is probably a valuable feature.

6.3.2 Rules for speaker order (turn-taking)

Another central element of debate protocols is the turn-taking, the order in which the agents are allowed to play their "move" in the debate, whatever a move may consist in. The choice of turn-taking can, in some cases, influence the debate's result in a significant way. Let us see the main types of turn-taking, ordered in increasing complexity. We shall assume that different agents

[4] This does not mean that it is impossible to define clusters when agents use extension-based acceptability, on the contrary, it may be an interesting subject of future research.

cannot simultaneously play their moves. Otherwise, even more complicated types of turn-taking could be defined.

1. **"Unregulated" turn-taking**: Turn-taking can never be completely unregulated, so what we mean here is the use of simple rules, introducing an element of chance.
   - **First-to-send**: Here we assume that the central authority asks for the agents' moves, and that the first one sent by an agent is chosen and played on the Gameboard.
   - **Random**: We can identify two small variations: Either (i) a time interval is given to all agents, who send their moves, and when the time expires, the central authority randomly chooses one among the proposed moves, or (ii) the central authority randomly chooses an agent and asks for his move.
2. **Fixed turn-taking**: By fixed turn-taking, we mean that the turn-taking follows a specific pattern. Using a fixed turn-taking, the central authority can try to ensure some fairness, which may be absent in "unregulated" turn-taking. For example, imagine an agent who is usually the quickest to provide his move, and a central authority which uses the first-to-send rule. This is certainly unfair to the other agents, who will have difficulties in participating to the debate as actively as the quick agent. A major choice, if we use a fixed turn-taking, is to decide whether it will be based on agent groups or on individual agents.
   - **Based on individual agents**: In this case, the playing order of the agents will be the same, throughout the debate. A simple choice, which we use later, is round-robin turn-taking.
   - **Based on agent groups**: In this case, the playing order of the groups is first defined. So, once a group has played its move(s), it passes the token to the next group. As above, we may have a round-robin turn-taking, but based on agent groups. Of course, we still need a way to attribute the token into the agents of the group which has the token. Naturally, there are numerous choices available.
3. **State-based turn-taking**: While fixed turn-taking is usually more fair than unregulated turn-taking, it suffers from the following problem: at times, the token can be passed to agents who will play moves irrelevant to the issue, thus not advancing the debate towards a conclusion. If the central authority wishes to accelerate the debate, it could choose itself the agent who gets the token, based on information about who is able to play a *relevant* move, advancing the debate towards a conclusion. In state-based turn-taking, both the actual state of the GB, as well as information on the agents (e.g. the group they belong to) is used by the central authority in order to decide who will get the token.

6.3.3 Locutions (permitted utterances)

Assuming that the agents have been partitioned in groups and that turn-taking has been defined, the next big question is how to define, at every step of the debate, the subset of legal locutions (moves) among all the moves that the agents could theoretically play. We propose the following three-step filtering of moves, especially useful in multilateral debates:

1. **Permitted moves**: This is the first level of filtering of the agents' moves. Moves violating basic conventions of the debate must be filtered out at this point. For example, in [BM11] there are two basic conventions: (i) The agent playing the move must be truthful (he cannot state an attack which he does not have in his system). (ii) No move can be repeated by the same agent. So, in [BM11] these two rules allow us to identify the permitted moves for every agent, at any given moment of the debate.
2. **Relevant moves**: As mentioned before, the central authority may want to speed up the debate, by forcing the agents to play moves which directly influence the issue's acceptability.

Therefore, a second filtering of the moves can be defined, in order to only allow relevant moves. In the context of our work, the exact definition of relevant move depends on the type of protocol considered. Roughly: (i) In category based protocols, a relevant move for group X, either makes group X the current winner, or it "helps" to make X the current winner. (ii) In cluster based protocols, a relevant move for a group X, turns the valuation of the issue on the GB closer to the representative value of cluster X.

3. **Most impacting moves**: If the two first levels of filtering have provided several relevant moves, then the central authority (in order not to just choose one at random) may select the most impacting move(s). In the case of numerical argument valuation, the most impacting moves may be those which provoke the biggest change of the issue's valuation (on the GB). The reason of preferring the most impacting moves is that, by letting major changes happen early, we hope that the outcome of the debate will be quickly stabilized. The criterion of most impacting move is therefore used as a tie-breaker, in presence of multiple relevant moves. A protocol's designer can propose different definitions of move impact.

6.3.4 Agent Strategies

Finally, we propose a typology for agent strategies in argumentative dialogue games. We should point out that agent strategies are not technically part of a game's protocol. The protocol, at every step of the game, designates all the possible moves for the agent (or agents) having the token. As we have seen, these are usually relevant moves which advance the debate towards a conclusion. So the question is: what happens when, at some point, there are many possible moves at the disposal of the agent having the token? The protocol may offer him the freedom to choose a move among them, so the agent would better be able to make a good choice. This is where a strategy is useful. It consists of a set of rules that the agent uses whenever he must choose his move among several possible ones.

In order to define a strategy, an agent needs to have some information on the game, and to be able to process it. We now identify different types of strategies, based on the type of information they require.

1. **Random strategies**: Given a set of relevant moves, an agent without an elaborated strategy, may simply choose one among them, at random. Random strategies may be useful when it is difficult to gather sufficient information on how the game could proceed, or to think of specific ideas increasing the probability of winning, or when most available moves seem equally good. In those cases, random strategies provide an easy solution, and by using them, an agent can (at least) take part in the debate. Moreover, random strategies can provide a useful benchmark for more elaborated strategies, in order to decide if these are worth the additional computational cost.

2. **(k)-history-based strategies**: These strategies, noted simply $h(k)$-strategies, select moves based on the last k moves uttered in the debate. For instance, the following rule may be part of a $h(1)$-strategy:

   > "If someone just attacked argument a, I will try to defend it."

   We argue that history-based strategies are often present (and useful) in real-life debates. In this type of debates, the participants rely more on the good timing and the emotional impact of their moves, as they desire to win the attention and sympathy of the audience. On the other hand, they tend to not have a holistic view of the debate. For example, they introduce counter-arguments in order to show how knowledgeable they are, and how good their reflexes are to defend their positions. Also, they usually strive to have the last word (to introduce the last argument) even if there still exist important arguments put forward by other agents which have not been counterattacked.

3. **(k)-state-based strategies**: These strategies, noted $s(k)$-strategies, select moves based on the last k states of the Gameboard. For instance:

   "If a is currently accepted, then I will utter the attack (d, a)."

   In formal argumentative debates, state-based strategies are arguably more useful than history-based strategies. In such debates, the fact that there are strict rules defining argument acceptability, forces the agents to choose their moves by taking into consideration every single move played until that point. This is of course possible thanks to the central structure (Gameboard) which aggregates all played moves, from the beginning of the debate.

4. **Strategies based on opponent modelling:** These strategies use a model of the opponent, and based on it, they try to predict his behaviour in the debate. This information is used by an agent in order to choose his moves. Examples of such works are found in [RPRS08, HSM$^+$13, RTO13]. For example, an agent may: 1) avoid arguments which he believes the opponent may use against him later, 2) prefer arguments against which he believes the opponent has no satisfactory replies.

5. **Strategies based on attitudes:** In these strategies an agent has a specific profile which defines his attitude in a debate (the general way he chooses his moves), as in [AM02]. For example, a cautious agent may avoid sending arguments, unless absolutely necessary; an impatient agent may send as many arguments as he can, as soon as he can; a sensitive agent may try to reply every time that one of his arguments has been attacked. Why would agents use this kind of strategies? First, they may believe that by adopting a specific attitude they maximize their expected utility. They may also have some hidden goals, for example to send signals about their profile to other agents. Finally, attitudes of human agents are usually affected by their character, so they are often difficult to change.

Note that an agent can adopt a mixed type of strategy. For example, an agent may use the debate's history and opponent modelling, or he may use opponent modelling together with a specific attitude. The work in [KMM05] considers that agent strategies can be based on personal attitudes (or agent types), but they must also conform to certain social norms, and use some specific tactics.

We now focus on strategies using information on the actual debate, namely state-based strategies and history-based strategies. Let us compare them on the notion of information basis.

We say that strategy s has a richer information basis than strategy s' (noted $s \rhd s'$) when s is able to use more information than s' in order to select a move. Equivalently, we say that s' has a poorer information basis than s.

First, we note that random strategies have the poorest information basis among all strategies, as the agents employing them do not make use of any information, except from the information needed in order to compute the relevant moves (which is the minimum information needed, in general). Trivially, in round t, for $k, k' \in \mathbb{N}$ such that $k' < k < t$, a $s(k)$-strategy (resp. a $h(k)$-strategy) has a richer information basis than a $s(k')$-strategy (resp. a $h(k')$-strategy). Also, in round t (and given that the initial Gameboard of the debate GB^0 is known), both $s(t)$-strategies and $h(t)$-strategies are fully expressive, in the sense that they can capture the whole evolution of the Gameboard, from GB^0 into GB^t; but it also holds that $h(t) \rhd s(t)$, because $h(t)$-strategies can tap into the additional information of which agent played which move (and when), something that $s(t)$-strategies cannot do. On the contrary, in round t, for $k < t$, state-based and history-based strategies are incomparable: For instance, a $s(1)$-strategy based on the last state of the Gameboard may capture intuitively more information than a $h(1)$-strategy based on the last move, but it misses the information of which was the last move uttered. Finally, notice that a $s(k)$-strategy, based on the last k states of the Gameboard

$$[GB^t, GB^{t-1}, GB^{t-2}, \ldots, GB^{t-(k-1)}]$$

has a poorer information basis than a strategy based on the single state $GB^{t-(k-1)}$ and on the last $k-1$ moves played.

6.4 Evaluating Argumentative Dialogues

After having analyzed the basic elements of argumentative dialogues, especially focusing on the central notion of dialogue protocol, we now turn our attention to an important issue when such dialogues are to be implemented. In [MPW02] the authors propose a number of criteria (desiderata) in order to evaluate the *quality* of argumentative dialogues. These desiderata can be used during the design and assessment of dialogue game protocols. The authors initially assume that agents engaged in dialogues are "autonomous, willing and free participants, able to enter and withdraw from dialogues as and when they see fit". Also, "within each dialogue, they remain autonomous, and are not compelled to accept or reject any proposition". The thirteen proposed desiderata are the following:

1. **Stated dialogue purpose:** The dialogue purposes need to be stated, so that all participants are aware of them in advance of entering the dialogue. Successful resolution of a dialogue will occur when its stated purposes are achieved.
2. **Diversity of individual purposes:** Individual purposes of agents may conflict or coincide, but they must be consistent with the overall purpose of the dialogue.
3. **Inclusiveness:** The dialogue must be open to properly qualified agents who are willing to participate.
4. **Transparency:** Participants to a dialogue should know the rules and structure of the dialectical system prior to commencement of the dialogue.
5. **Fairness:** The dialogue should not give a priori any advantage to particular participants. Nonetheless, there may be agents with different roles (e.g. seller and buyer agents).
6. **Clarity of argumentation theory:** How clear is the process of argument construction? How clear are the dialogue rules?
7. **Separation of syntax and semantics:** Ensuring that the protocol syntax is defined separately from its semantics enables the verification of conformity with protocol syntax, even if the protocol semantics cannot be completely verified.
8. **Rule-consistency:** The locutions and rules of a dialogue system should together be internally consistent; that is, they should not lead to deadlocks (where no participant may utter a legal locution), nor infinite cycles of repeated locutions.
9. **Encouragement of resolution:** Resolution of each dialogue (normal termination) should be facilitated, and not precluded, by the locutions and rules of a dialectical system.
10. **Discouragement of disruption:** Normally, the rules of a dialectical system should discourage or preclude disruptive behaviour, such as uttering the same locution repeatedly.
11. **Enablement of self-transformation:** A dialectical system should permit participants to undergo self-transformation in the course of a dialogue; e.g., participants to a negotiation should be able to change their preferences or their valuations of utility as a result of information they receive from others in the dialogue, or express degrees of belief in propositions.
12. **System simplicity:** The locutions and rules of a dialectical system should be as simple as possible,
13. **Computational simplicity:** The outcome of the dialogue should have small computational demands, both from its users and from the system itself.

Also, in [TG10] the authors propose three criteria which can be used to evaluate, not the quality, but the *complexity* of argumentative dialogues. These criteria are the following:

- **Protocol:**
  The authors differentiate protocols on the basis of their general complexity: (1) The simplest types are *direct argument mechanisms*, in which the dialogue is a one-shot game. All agents must choose their utterances, and they all play them at the same time. (2) A little more complex are the *synchronous argument mechanisms*, where the agents are allowed to play at the same time. But then, this step can be repeated, until some termination condition is met. (3) The most complex are the *dialectical argumentation mechanisms*, where agents take turns during the debate.

- **Awareness:**
  The second key element of the complexity of argumentative dialogues is the degree of awareness agents have, which refers to their amount of knowledge on the other agents' beliefs and preferences. Naturally, the more an agent knows about the others, the more he may be able to choose efficiently his utterances. The cases, from simplest to more complex are the following: (1) No awareness: an agent has no information at all on the others' systems. In this case, when the agent chooses his move in the debate, he is based only on his own system as well as on the common system (and the utterances made) up to that point. (2) Full awareness: The agent knows exactly all the agents' systems. Of course, between no-awareness and full awareness, there is a range of cases. The works in [RPRS08, HSM$^+$13, RTO13] fall between the two extreme cases. In these works, an agent has partial information on the other agents' potential replies, and he uses it in order to compute the best move at any given point of the debate.

- **Agent types (preferences):**
  In [TG10] every agent is equipped with a utility function which evaluates the debate's outcome. As far as the agents' preferences over the outcome are concerned, from the simplest to the more complex case, we have: (1) Focus on a single proposition (focal element). The simplest case is the binary accept/reject scenario. (2) Focus on multiple propositions (multiple focal elements). In this case, the utility of an agent is maximal when all his focal elements have the desired status. (3) Focus on multiple propositions (counting indicator function). In this case, the more his focal elements have the desired status, the more the agent's utility increases.

6.4.1 Evaluating persuasion dialogues

In [AdSC13] the authors' scope is narrower, as they focus on persuasion dialogues. They propose six postulates which can be used to evaluate their quality.

1. **Finiteness:**
   A dialogue system DS respects finiteness *if and only if* every possible generated dialogue consists of a finite set of moves, and every move has a finite size. Finiteness issues can be tricky when the agents are able to update their beliefs dynamically during the debate, or to retract arguments they have asserted, as this may lead to move repetitions. Usually, given a finite number of debating agents, finiteness can be guaranteed if some restrictions are put on move repetitions.

2. **Non-determinism:**
   A dialogue system DS respects non-determinism *if and only if* it is able to generate at least two dialogues with different outcomes. From one perspective, non-determinism may be considered as a wished property: it ensures that agent strategies may have some utility in the dialogue, as no agent is, a priori, deemed to lose. From another perspective, if we assume that there exists a "correct" opinion on an issue, then we cannot be certain that the outcome of a debate will agree with that opinion.

3. **Consistency:**
   Roughly, a dialogue system DS respects consistency *if and only if* for every possible dialogue, there are no contradictory conclusions obtained.

4. **Natural-attacks-allowance:**
   A dialogue system DS respects natural-attacks-allowance *if and only if* all attacks set during the dialogue are either rebuts (inconsistent pair of conclusions) or assumption-attacks (inconsistent pair conclusion - premise). This postulate is needed in order to avoid cases where the participants set attacks which are not logically founded.

5. **Dissimulation:**
   In a dialogue the agents may hide arguments and information in order to reach their objectives. A dialogue system should thus allow dissimulation of information. More formally, a dialogue system DS respects dissimulation *if and only if* it can generate some dialogues where the winner would have been different if an agent had uttered (and thus not concealed) a specific argument.

6. **Non-triviality:**
   Arguments are used to justify some assertions by providing new evidences. Tautological arguments fail to meet the objective of arguing. Thus, a dialogue system DS respects non-triviality *if and only if* every argument, in every possible dialogue, brings some new information.

6.4.2 Evaluating strategies in persuasion dialogues

Before concluding this chapter, inspired by the previously presented evaluation criteria of dialogue protocols found in the literature, we propose some additional criteria which evaluate agent strategies in such dialogues. The strategy evaluation criteria we propose are the following:

- **Finiteness:**
  As in [AdSC13], finiteness is usually studied at the level of the protocol, in order to know if the debate will always terminate after a finite number of moves, regardless of the agents' strategies. But, if a protocol does not satisfy finiteness, then it might be useful to know if finiteness holds in the case where specific strategy profiles are used by the agents (e.g. when the agents' strategies never propose retract moves).

- **Non-Determinism:**
  Once again, non-determinism is usually studied at the level of the protocol. Nonetheless, if non-determinism is proven, we might ask if non-determinism also holds in the case of some specific strategy profiles.

- **Convergence to state satisfying a property:**
  Is it guaranteed, possible, or impossible (and under which conditions) to obtain, at the end of the debate, a collective outcome satisfying some property? Do the agents' strategies play an important role in obtaining that outcome? Let us give some examples of interesting properties that a debate may satisfy. We assume the use of a GB.

  1. The debate's stable GB is identical to the merged system.
  2. The debate's outcome coincides with the merged outcome.[5]
  3. The debate's outcome coincides with the opinion of the agents' majority.
  4. The debate's outcome coincides with the unanimous opinion of the agents.

  Naturally, if the debate's stable GB is identical to the merged system, then the two outcomes coincide. On the other hand, if the outcomes coincide, it does not follow that the two systems are obligatorily identical.

[5] They indicate the same winning and losing agents.

- **Debate length:**
  Agent strategies can affect debate length.[6] Usually, shorter debates are preferred to longer ones, for various practical reasons. In shorter debates, the conclusion is drawn more quickly, and the agents, as well as the central authority, spend less resources (e.g. time, computational resources). Moreover, real systems often face stability issues, for example some agents may lose their connection with the central authority of the debate, before the debate has ended. Short debates help a lot in avoiding such types of inconveniences. Let us also note that debates usually allow pass moves, which are "empty" utterances, simply giving the token to the next agent. Therefore, as far as debate length is concerned, we may focus on:

  1. The total number of timesteps (including timesteps when pass moves were played).
  2. The number of "informative" timesteps only (excluding timesteps when pass moves were played).

- **Efficiency:**
  While the four previous criteria arguably interest both the agents and the central authority, this criterion mainly interests the agents. Here, the question is how a strategy fares against the others. Is there a dominant strategy? Are there any Nash equilibria? Which strategies maximize an agent's probability to win the debate?

At this point we have finished our overview of the key points of multi-party dialogues, having focused on the notion of protocol for argumentative multi-party dialogues. Based on the previously mentioned literature, as well as on the protocol typology we have proposed, in the next chapter we will define and study specific, and diverse, debate protocols. Also, we will define and evaluate agent strategies in the context of these protocols. Some of these strategies will be based on the notion of minimal change achieving a goal, which has been extensively analyzed in Chapter 5.

[6] We will later see specific examples.

Chapter 7

Argumentative Debate Protocols[1]

In this chapter we present our work on argumentative debate protocols. We define and study protocols with different characteristics as, depending on the application and context, some types of protocols may be preferred over others.

First, we address the following question, related to the issue of debate controversy: given a Gameboard which contains a non-persistent argument, how can we turn it persistent, as soon as possible, by asking the opinions of experts on weak attacks?

Then, we concentrate on debate protocols which oppose two (or more) parties:

- In the first type of protocols, abstract (Dung) argumentation systems are used, while argument acceptability is extension-based (specifically grounded semantics is used). One specific protocol is defined, which focuses on target sets. A number of agent strategies and heuristics are proposed for that protocol.
- In the second type of protocols, abstract bipolar argumentation frameworks are used, while argument acceptability relies on numerical valuations. The two subtypes of protocols, which have already been presented, are: 1) *category-based protocols*, and 2) *cluster-based protocols*. For each subtype, one representative protocol is defined, and their properties are studied.

Finally, the protocols, as well as the strategies and heuristics, are evaluated on the basis of criteria presented in Chapter 6. Regarding the protocol which focuses on target sets, the evaluations are backed-up by experimental results.

Though the two types of protocols we will present have important differences, they share the following common features with the protocol defined in [BM11]:

1. A debate has an arbitrary number of participants, each one equipped with his own private argumentation system, representing his viewpoint on the discussion.
2. All the agents share the same sets of arguments, but they may disagree on the attack and support relations over them.
3. All the agents focus on a single argument of the debate which is called *issue*.
4. All the agents use a common Gameboard (GB), where they assert arguments and relations between them.[2] The debate takes place in discrete timesteps. At every timestep, the GB represents the actual state of the debate, so the issue's status is computed from the GB. Also, at every timestep, an agent is either winning or losing the debate, depending on the relation between: the issue's status on the GB and the issue's status on his private system.

[1] Our publications relevant to this chapter are: [KBMM11, KBMM14a].
[2] In our work, we will consider attack and support relations.

5. Disagreement over relations is addressed by letting the agents vote on them. A positive vote on a relation expresses the belief that this relation holds, while a negative vote expresses the opposite belief. Moreover, the agents are assumed to vote in accordance with their beliefs (they are called "truthful").

6. The agents' private systems are static, thus arguments and relations cannot be modified.

Let us make some comments on the above points.

First, persuasion dialogues where the agents' private argumentation systems are dynamic (true persuasion dialogues) constitute arguably a much more complex case than the one considered in this work (which is dispute resolution). They are left for future research.

Another choice promoting simplicity is that all the agents share the same set of arguments. Consider the case where an agent is not aware of argument x, and at some point of the debate, another agent puts forward x. Then, the first agent is faced with multiple questions: should he include it in his own system? Even if we consider static private systems (where the answer would be "no"), more questions arise: what should the agent think about the relations of argument x with the other arguments in the debate? It is not easy to provide a meaningful answer, unless the arguments' structures are analyzed.[3] Given that in our work abstract argumentation is used, we have chosen to make the simplifying assumption that all agents share the same set of arguments.

We have already explained why agents may disagree on the existence of an attack relation between two arguments. We remind that the two main reasons are: 1) the fact that arguments are usually enthymemes [Wal08], which contain implicit information, and 2) the fact that the link between an argument's conclusion and another argument's premises may be implicit. These two reasons may also cause disagreement on the existence of a support relation between two arguments. Let us see an interesting example focusing on the second reason.

Consider the following arguments:
Argument a: Team X has great chemistry, therefore they will be champions.
Argument b: Team X has recently bought some expensive players.

Notice that the link between b's conclusion and a's premise is not clear. A first agent may believe that the arrival of some expensive players will improve the team's chemistry; therefore he believes that b supports a. A second agent may believe that adding some expensive players, will have no direct effect on the team's chemistry. Therefore, the second agent believes there is no support from b to a. Finally, a third agent may believe that the addition of some expensive players will impact the team's chemistry, but in a negative way, due to the big egos of the players. As a result, the third agent believes that b attacks a.

We also want to stress the importance of letting the freedom to the agents, to vote on whichever attacks they wish (provided some conditions are met, as we will shortly see). Indeed, the agents may not want to disclose some of their beliefs, at least not from the start of the debate. They may wish to hide such information from the other participants for various reasons. For example, we can easily think of cases where valuable information should not be revealed, even if it could be helpful during a debate. Another example could be that some information might, at some point, be used against the agent who disclosed it. We want to give the agents the ability to "play safe" and to reveal, at every step of the debate, a minimum amount of information, able to advance the debate. This is why debate protocols, which lead to a gradual merging of information, differ significantly from other, more direct merging approaches, as those proposed in [CMDK$^+$07].

[3] In [CMDK$^+$07] the authors address this problem in an abstract argumentation setting. The difference, compared to our work, is that they assume that an agent can "see" the private systems of the other agents. This is quite unlikely in a debate, where being able to hide information is common and crucial. Note that [CMDK$^+$07] also introduces a relation of indecisiveness on whether an attack holds or not, which we do not use.

Finally, we would like to underline that, in this work, we make the assumption that agents are truthful. This means that every agent's votes must be in accordance with his private system. Of course, in many settings, it could be possible for an agent to intentionally lie, if this would help him during the debate. This could lead to even richer protocols where lying is an additional weapon in the arsenal of a debating agent.

7.1 Converging quickly to a persistent issue

In this section, we address the following problem: given a debate modelled in the form of a Gameboard, and given a single issue which is non-persistent, the debate's administrator strives to turn the issue persistent (by turning some unstable attacks into stable). In order to achieve this, he can ask the opinions of (sufficiently many) experts on the unstable attacks. The question is, in what order should the unstable attacks be put into question? We specifically search for an ordering which will lead to a persistent issue, as soon as possible. We present, with the help of an example, a technique which can be used to find a good ordering of the questions.

The example is inspired from the Parmenides system [CA09]. Parmenides is build upon the value-based extension [BC03] of Dung's abstract argumentation frameworks, although the arguments are instantiated as specific schemes for deliberation [WRM08]. More specifically, we shall follow the *"Speed Camera Debate"* appearing in [CA09].

Upon entering the Parmenides system, the user is confronted with an "official" position, as set-up by the *administrator*: It is a proposal of an action to be undertaken (e.g. "Install more cameras"). If he disagrees with it, the user is then lead through a series of critical questions *CQ*s (which consists in "the critique phase"), prepared beforehand by the system administrator. Specifically, each of the premises of the official position is challenged by a critical question. For instance, in the *"Speed Camera Debate"* we consider here, the user would be asked, in sequence, whether he believes all the circumstances are indeed as stated, so we may have the following critical questions: (i) *Is there a high death toll on UK roads?* (ii) *Do many drivers break the speed limits?* and (iii) *Does the government make money from fining speeders?* The user may reply positively or negatively to any of these questions. Next, the user is asked whether the different social values involved in this debate (e.g. *saving lives*, *improving the government wealth*, and *enforcing law and order*) should indeed be promoted. Then, it is checked whether the user believes that the consequences of the action indeed promote the different social values, whether the action could instead demote some of the social values, whether the action will have the stated consequences, and finally, whether alternative actions should be considered to reach the goal. Interestingly, the order of questioning is based on importance (e.g. the user is given the opportunity to critique the circumstances first). After this phase, the user is asked whether he would like to submit an alternative position ("Alternative Position Phase"). We do not describe here this phase, since it is out of the scope of the current discussion.

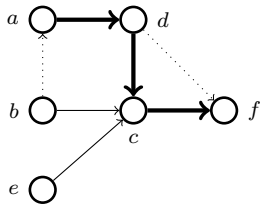
Let us highlight a number of important points: (i) the official position and the potential arguments are prepared beforehand by the administrator, and (ii) the official position does not change, and the user does not get to see how his contribution modifies the current state of the debate.[4] The administrator does not necessarily seek to obtain a *minimal* amount of information. However, as debates become richer and involve more and more justifications, some choice has to be made as to which part of the initial position must be challenged. More complex interactions can occur, e.g. some circumstances may be incompatible with other ones. In this context, it is valuable to help the administrator select a *good ordering* of questions. The goal of the following analysis is to provide insight on how to find the best ordering of questions.

Now we present the example of the *"Speed Camera Debate"* in the form of a debate modelled by a Gameboard and containing some unstable attacks. In our case, a "critical question" is an unstable attack, which must be "answered" either positively (meaning that the attack is valid),

[4] Although some tools allow the administrator to do so.

or negatively (meaning that the attack is invalid). We show how the debate's administrator can order the critical questions he will ask to a sufficient number of additional experts, in order to turn the debate's issue persistent, as soon as possible.

Example 30 *Let $\mathbb{W} = \langle A, R, Eval \rangle$ be a GB such that $A = \{a, b, c, d, e, f\}$ is the following set of arguments from the Parmenides system. a: install more speed cameras is useful for social marketing, b: there is a high death toll on UK roads, c: the circumstances are not true, d: install more speed cameras only enriches the UK Government, e: there is a high number of high speed limits offences on UK roads, and the argument-issue is f: Install more speed cameras. Assume that, at some point of the debate, the weak attacks are $R^{\mathbb{W}}_{wkP} = \{(b, c), (e, c)\}$, $R^{\mathbb{W}}_{wkN} = \{(b, a), (d, f)\}$, while the strong attacks are $R^{\mathbb{W}}_{strP} = \{(a, d), (d, c), (c, f)\}$. No other attacks have been proposed during the debate. Therefore, the attacks (b, c) and (e, c) are present in the counterpart system, but they can be removed, while the attacks (d, f) and (b, a) are not present, but they can be added. Notice that f is currently accepted, as $lab^{\mathbb{W}}(f) = in$, but it is not persistent ($f \notin A^{\mathbb{W}}_{pers}$), due to the four unstable attacks of the system. Assume that our goal is to turn argument f persistent, by asking a number of experts their opinions on the unstable attacks, and that we want to achieve this goal as soon as possible.*



In the following, we present the policies of questions on unstable attacks, in the form of binary decision diagrams.[5] *Two such decision-diagrams are illustrated in Figure 7.1.*

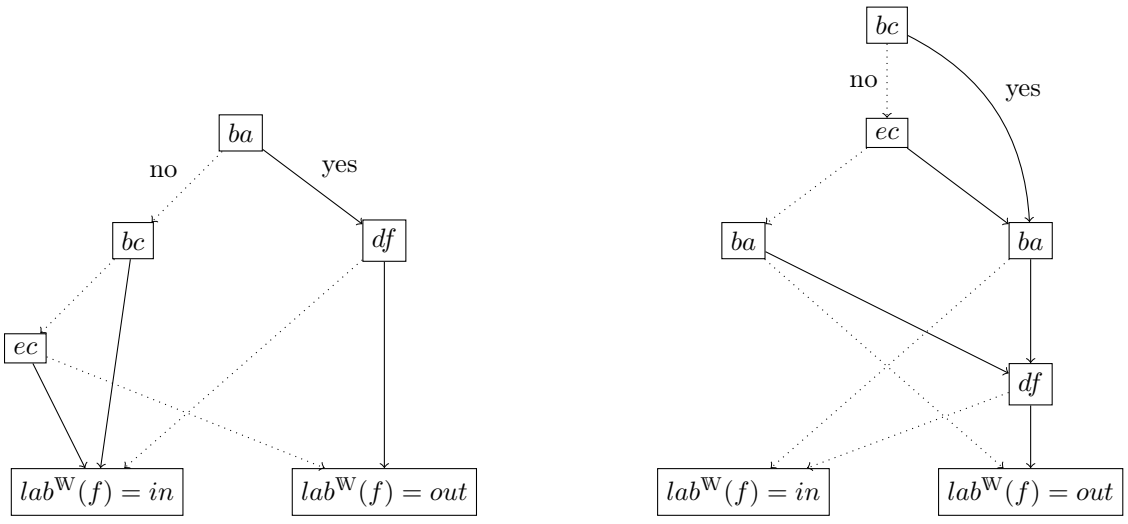


Figure 7.1: Two different orderings of questions, visualized as binary decision diagrams. Dotted (resp. normal) arrows correspond to the answer "no, the attack does not hold" (resp. "yes, the attack holds"). At the bottom nodes, argument f is turned persistent (either *in* or *out*).

[5]Note a slight difference though, since variables have initially a default value.

In Figure 7.1 there are two decision diagrams. Firstly, in the decision diagram on the left, we start by asking additional opinions of experts on the attack (b, a). We hope that after a sufficient number of opinions, (b, a) will be turned strong. In that case, we will have either $(b, a) \in R^{\mathrm{W}}_{strP}$, or $(b, a) \in R^{\mathrm{W}}_{strN}$. If we obtain $(b, a) \in R^{\mathrm{W}}_{strP}$ (the experts have decided that (b, a) is valid), then we will seek info on (d, f). Similarly, we expect that at some point we will obtain either $(d, f) \in R^{\mathrm{W}}_{strP}$, or $(d, f) \in R^{\mathrm{W}}_{strN}$, but in either case, the argument f will be turned persistent (either in or out), as the attacks (b, c) and (e, c) will have become irrelevant to the status of f. Secondly, in the decision diagram on the right, an alternative order of questions is considered.

Let us now compare the efficiency of these two decision diagrams. The metric used is the average path length from the initial node to a final node. In the first decision diagram, there are 5 possible paths, which contain (respectively) 2,2,2,3 *and* 3 *edges (questions). Therefore, the average path length is* 2.4 *questions. Now let us examine the second decision diagram, which first seeks info on the attack (b, c). The average path length here is greater,* 3.333 *questions. We conclude that, if we want to turn f persistent, then among the two above orderings of questions, the first one is more efficient, as it is expected to lead to less questions.*

In this section we have presented some basic ideas to address the following problem: given a Gameboard, how can we focus on the right questions to ask in order to converge, as quickly as possible, towards a persistent issue. We believe that these ideas could form the basis of a rigorous method, able to address this problem for any arbitrary Gameboard. Also, it could be interesting to examine whether (and how) such a method could rely on the computation of target sets.

7.2 A Debate Protocol Focusing on Target Sets

In this section we define a first type of debate protocol among experts. In this protocol we use extension-based acceptability and, more specifically, grounded semantics in order to draw conclusions. A debate's designer may find the use of grounded semantics preferable in some cases, for two main reasons: the simplicity of having a unique grounded extension, and the computational efficiency of finding it. Moreover, we will use the previously presented work on target sets, in order to define agent strategies focusing on the minimal change achieving a goal.

7.2.1 Modeling the debate's setting

In this subsection we define the debate's setting. Let us start with the participants.

The participants

A finite set of agents, denoted N, take part in a debate. Each agent $i \in N$ has an abstract (Dung) argumentation system. All the agents share the same set of arguments A, and they focus on the acceptability of a specific argument, which is the *issue* of the debate. For each agent, if the issue belongs (resp. does not belong) to the grounded extension of his system, then he belongs to the PRO (resp. CON) group.

Moreover, the agents may disagree on the attack relation over the arguments. We will assume that exists is a set of attacks on which all agents agree, and a set of attacks on which they have different opinions. The *master argumentation system* contains all the attacks on which the agents agree, as well as all the attacks on which they disagree. Since ASMAs separate attacks into fixed and debated, we define the master argumentation system as an ASMA.

Definition 59 *A* **master argumentation system** *is an ASMA $SM = \langle A, R, R^+, R^- \rangle$, where $R^{fix} = R \setminus R^-$ denotes its set of fixed attacks, and $R^{deb} = R^+ \cup R^-$ denotes its set of debated attacks. We say that an abstract (Dung) argumentation system $AS_i = \langle A, R_i \rangle$ is* **in accordance with** *SM if and only if $R^{fix} \subseteq R_i \subseteq (R^{fix} \cup R^{deb})$. In other words, AS_i is in accordance with SM if and only if it contains all the fixed attacks of SM, and it contains only fixed and debated attacks of SM.*

Let us see an example of a master system and some private systems which are in accordance with it.

Example 31 *Let $SM = \langle A, R, R^+, R^- \rangle$ be a master system, with $A = \{a, b, c, d\}$, $R = \{(a, b), (b, c), (d, c)\}$, $R^+ = \{\}$, and $R^- = \{(a, b), (d, c)\}$.*

Also, let $1, 2, 3$ be three agents with their corresponding private systems $AS_1 = \langle A, R_1 \rangle$, with $R_1 = \{(a, b), (b, c)\}$; $AS_2 = \langle A, R_2 \rangle$, with $R_2 = \{(b, c), (d, c)\}$; $AS_3 = \langle A, R_3 \rangle$, with $R_3 = \{(b, c)\}$.

The master system as well as the three private systems are illustrated in Figure 7.2. The fixed attack of the master system is represented with a bold arrow, while the two debated attacks are represented with normal arrows. Notice that all the agents' systems are in accordance with SM, as they all contain the fixed attack (b, c), and they contain only fixed and debated attacks. It holds that: $E_{Gr}(AS_1) = \{a, c, d\}$, $E_{Gr}(AS_2) = \{a, b, d\}$ and $E_{Gr}(AS_3) = \{a, b, d\}$. This means that if argument c is chosen to be the issue, then the agents will be separated into $PRO = \{1\}$ and $CON = \{2, 3\}$.

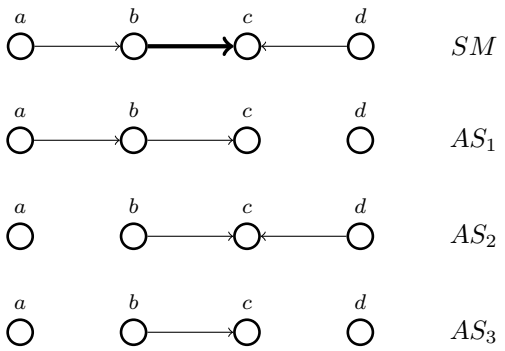


Figure 7.2: The master system SM and three private systems AS_1, AS_2, AS_3 which are in accordance with it. The fixed attack (b, c) is copied in all the private systems, while the debated attacks may or may not belong to the private systems.

As in Subsection 3.1.1, we consider that each argument is associated with a set of keywords specifying which topics this argument is about. We assume a fixed set of topics, denoted T, and the set of *topics of an argument* $a \in A$ is given by the function $top(a) \subseteq T$. The set of *topics of an attack* $(a, b) \in R$ is given by the function $top(a, b) = top(a) \cup top(b) \subseteq T$.[6]. Finally, each agent is expert in a subset of topics. The *expertise of agent* $i \in N$ is given by function $exp(i) \subseteq T$.

The Gameboard with respect to a Master System

As previously, we will define debates which use a central argumentation system, called Gameboard (GB), which aggregates all the opinions expressed by the agents during the debate. From the GB, at every time, the debate's conclusions can be drawn.

Sometimes, in debates among humans, the central authority may desire to start a debate with a relatively narrow scope. Assume that the central authority knows *a priori* the arguments which are relevant to the issue, the attacks which certainly hold over them, as well as the non-attacks. Then, it starts a debate where the participants can only contribute with their opinions on some (debated) attacks. How could such a debate be modelled?

Inspired by the ASMAs, which consider both fixed and debated attacks, and based on Definition 23 of a Gameboard, we will consider a slight variation of the Gameboard. This Gameboard

[6]Here, for simplicity, we shall use the union ($\cup$) instead of the multiset union ($\uplus$).

contains a fixed set of arguments. The agents cannot introduce new arguments, they can only discuss about the attacks. Moreover, some of the attacks are fixed (they are not under discussion), and only a subset of the attacks are debated. Agents can vote either for, or against any debated attack. The role of the GB is to gather and aggregate the votes cast during the debate. Moreover, the voters' relevant expertise will play a crucial role in determining the result of the aggregation.

Definition 60 *Let $SM = \langle A, R, R^+, R^- \rangle$ be a master system, where R^{fix} and R^{deb} denote its fixed and debated attacks, respectively. The* **Gameboard with respect to** SM[7]*, at timestep $t \in \mathbb{N}$, is denoted by $GB^t = \langle A, R^{fix}, R^{deb}, Eval^t \rangle$ where: A, R^{fix}, R^{deb} are, respectively, the set of arguments, the set of fixed attacks, and the set of debated attacks of SM, and $Eval^t$ is the set of evaluation vectors of the debated attacks in R^{deb}.*

We consider that only debated attacks have evaluation vectors, as the agents can only vote on them. Also, when we refer to no specific timestep, or when the timestep we refer to is evident from the context, we will drop the superscript t from the notation.

A vote on Gameboard $\langle A, R^{fix}, R^{deb}, Eval \rangle$ is a tuple $\langle (a, b), s, i \rangle$ where $(a, b) \in R^{deb}$ is the debated attack concerned by the vote, $s \in \{-1, +1\}$ is the sign of the vote, and $i \in N$ is the voter. As usual, a positive vote by an agent means that he supports that the attack holds, while a negative vote means that he supports the opposite. Every debated attack $(a, b) \in R^{deb}$ has an evaluation vector, denoted $\vec{v}(a, b) = \langle w(a, b), mw(a, b) \rangle$. The evaluation vectors of debated attacks are updated after every vote, as described in Subsection 3.2.1.

Given a Gameboard, in order to compute the acceptable arguments we use its counterpart system. Here, the counterpart system contains simply all the fixed attacks, and the debated attacks with positive weights.

Definition 61 *Let $GB^t = \langle A, R^{fix}, R^{deb}, Eval^t \rangle$ be a Gameboard with respect to a master system SM. The* **non-weighted counterpart argumentation system** *of GB is an abstract argumentation system denoted by $GB_{cp} = \langle A, R_{cp} \rangle$, with $R_{cp} = R^{fix} \cup \{(a, b) \mid (a, b) \in R^{deb}, w(a, b) > 0\}$.*

Example 31 (cont.) *Let $T = \{t_1, t_2, t_3\}$, with $top(a) = \{t_1, t_2, t_3\}$, $top(b) = \{t_2\}$, $top(c) = \{t_2, t_3\}$, $top(d) = \{t_1\}$. Also, let $exp(1) = \{t_1, t_2\}$, $exp(2) = \{t_2, t_3\}$ and $exp(3) = \{t_1\}$. At the beginning of the debate, no votes have been cast on any attack. Below every attack $(x, y) \in R^{deb}$, we indicate its evaluation vector $\vec{v}(x, y)$. Figure 7.3 illustrates the initial Gameboard and its counterpart system. Fixed attacks are drawn with bold arrows (as usual); debated attacks with non-positive weights are drawn with dotted arrows; debated attacks with positive weights are drawn with simple arrows.*

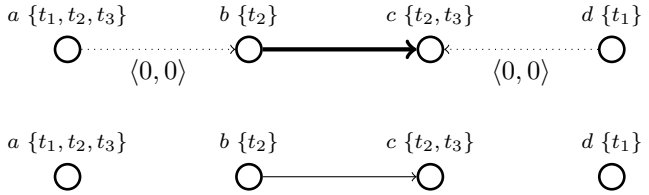


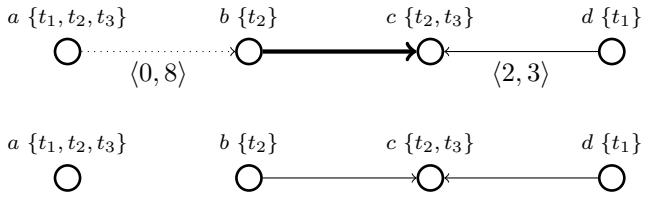
Figure 7.3: Above: The initial Gameboard of the debate (with the arguments' topics). Below: Its counterpart system.

At the beginning of the debate it holds that $c \notin E_{Gr}(GB)$, so the issue is being rejected.

Let us assume that agent 1, who belongs to PRO, votes for attack (a, b). As $top(a, b) = \{t_1, t_2, t_2, t_3\}$ and $exp(1) = \{t_1, t_2\}$, we get $\vec{v}(a, b) = \langle 3, 4 \rangle$. The agent has succeeded in turning the attack's weight

[7]It will be simply called *Gameboard* when there is no danger of confusion.

positive, thus putting it into the counterpart system. Currently, the issue c is being accepted. Next, assume that agent 2*, who belongs to CON, votes against attack* (a, b)*. As* $top(a, b) = \{t_1, t_2, t_2, t_3\}$ *and* $exp(2) = \{t_2, t_3\}$*, we obtain* $\vec{v}(a, b) = \langle 0, 8\rangle$*. The issue c is now rejected, as the attack* (a, b) *has been removed from the counterpart system. Finally, agent* 2 *plays once again and tries to reinforce the position of the CON group, by voting for attack* (d, c)*. As* $top(d, c) = \{t_1, t_2, t_3\}$ *and* $exp(2) = \{t_2, t_3\}$*, we obtain* $\vec{v}(d, c) = \langle 2, 3\rangle$*. The updated Gameboard* GB'*, as well as its counterpart system* GB'_{cp}*, are shown below:*



As we can see, the argument c is being rejected, and CON is winning.

The merged system

When participating to the debate, agents are assumed truthful, and they cannot vote for (resp. against) an attack if they think that it does not (resp. does) hold. Certainly, absolute truthfulness is not often encountered in real-life debates, but it is an assumption preventing the agents from stating anything that may help them in the debate. A more refined approach would be to define a set of beliefs (in our case attacks) upon which an agent is able to lie, if he considers it favourable at some point. This kind of situation has already been studied by Rahwan et al. [RL09].

On the other hand, we allow agents to *not* express their opinion on some attacks, because that could harm their purpose, or make them disclose information they wish to hide. We thus need a way to compare the results obtained in our debates with a collective view of the argumentation systems of the agents. We rely on two different notions. The first one is the notion of *merged* argumentation system [CMDK+07]. In the specific case we discuss here, it turns out that a meaningful way to merge is to take the vote of all agents on all the debated attacks.

Definition 62 *Let* $SM = \langle A, R, R^+, R^-\rangle$ *be a master system. Let* N *be a set of agents, whose systems are in accordance with* SM*. Let* $GB^0 = \langle A, R^{fix}, R^{deb}, Eval^0\rangle$ *denote the Gameboard with respect to* SM*, at the start of a debate. Let* $GB^k = \langle A, R^{fix}, R^{deb}, Eval^k\rangle$ *denote the Gameboard after all the agents in* N *have voted on all the debated attacks.* GB^k *is called* **merged system**.

Another notion which can be useful for analyzing the collective view of the debate is the one of *happiness*: we could want to see a majority of agents satisfied at the end of the debate, in the sense that they agree with the status of the issue of the debate.

Definition 63 *Let* N *be a set of agents, and* $\forall i \in N$*, let* AS_i *denote the system of agent* i*. Also, let* $d \in A$ *be an argument called* issue of the debate. *The* **majority result** *is denoted* $majIn(d) \Leftrightarrow |\{i \in N \mid d \in E_{Gr}(AS_i)\}| \geq |N|/2$*, and it is denoted* $majOut(d) \Leftrightarrow |\{i \in N \mid d \notin E_{Gr}(AS_i)\}| > |N|/2$.

Note that ties for the majority are broken in favour of the agents who want to see the issue in the grounded extension.

7.2.2 Defining the protocol

In this section we define a specific debate protocol. As we said before, a set of agents N focus on the status (under grounded semantics) of a single argument $d \in A$, which is the issue of the debate. The goal of agent $i \in N$ is that, at the end of the debate, the same issue's status, in the GB and in his private system, at the end of the debate. We can therefore distinguish two groups

of agents: the agents of the group PRO (resp. CON) who have (resp. do not have) the issue in the grounded extension of their systems.

The protocol proceeds in timesteps. The debate starts at timestep $t = 0$, when we have $GB^0 = \langle A, R^{fix}, R^{deb}, Eval^0 \rangle$, with $\forall (a, b) \in R^{deb}$, $\vec{v}(a, b) = \langle 0, 0 \rangle$. During the debate the agents are able to cast votes on debated attacks. In order to ensure the termination of our protocol, we assume that an agent cannot vote on the same attack twice. To account for this, each agent $i \in N$ is equipped with a set RP_i^t which contains all the attacks agent i has voted on, until timestep t. The protocol is defined by the following:

- **Participants**: A finite set of agents N, each one being either PRO or CON, according to his opinion on the issue's status.
- **Turn-taking**: Round-robin. The token is given to each agent, in turn, and comes back to the first agent once all agents have played.
- **Permitted moves**: Agent i at timestep t can either:
  - Cast the vote $\langle (a, b), +1, i \rangle$, if $(a, b) \in R^{deb}$ and $(a, b) \notin RP_i^t$
  - Cast the vote $\langle (a, b), -1, i \rangle$, if $(a, b) \in R^{deb}$ and $(a, b) \notin RP_i^t$
  - Play a **pass move** (giving the token to the next agent).
- **Stopping condition**: $|N|$ pass moves have been played in a row.
- **Winning condition**: Once the debate has stopped, all PRO (resp. CON) agents win if and only if the issue belongs (resp. does not belong) to $E_{Gr}(GB)$.

7.2.3 Strategies using target sets

When an agent has the token, he can vote on any debated attack, but which one should he choose? In general, a *strategy* states, for each agent, what move should be uttered next in the course of the debate. Depending on the information required to take this decision, we now remind two types of strategies which we have already described in Chapter 6: history-based strategies and state-based strategies.

- *(k)-history-based strategies*: the strategy selects moves based on the last k moves uttered in the debate, noted *h(k)-strategies*. For instance:

  "If someone just attacked argument a, I will try to defend it."

- *(k)-state-based strategies*: the strategy selects moves based on the last k states of the Gameboard, noted *s(k)-strategies*. For instance:

  "If $a \in E_{Gr}(GB)$, then I will utter the attack (d, a)."

In what follows, we study a natural class of *s(1)-strategies* which are based on target sets. Given a GB and its counterpart system GB_{cp}, a target set indicates which attacks should be added and which ones should be removed from GB_{cp}, in order to achieve a given goal (put the issue in the grounded extension, or drop it from the grounded extension).

In Section 5.4 we presented some properties of target sets and we showed the practical usefulness of "playing" in target sets, which is the following: by changing (adding or removing) attacks in a target set, we are certain that we shrink *at least one* target set. But, in a multi-agent debate setting, is this sufficient in order to use target sets for agent strategies? Is it always the best choice to play in target sets, no matter the agents' systems? Or is it advantageous in most cases? We will now address these questions, in the debate setting which we have presented.

Lack of dominance and equilibrium guarantees

Dominance. One may wonder whether "playing within target sets" of the system is a dominant strategy, that is, whether agents can never be better off playing a different strategy, whatever the strategy of the other party is. Note first that "playing within target sets" does not constitute a single strategy, because at some points of a debate there may be several permitted moves on target sets. Instead, it constitutes a class of strategies, which is in fact a subclass of *s(1)-strategies*. So when say "playing within target sets is a dominant strategy", we abuse language and we actually refer to *any* strategy belonging to this class. This turns out to be a too demanding notion, because the strategy of the other player can be of any kind, in particular, it may be such that moves played outside a target set will precisely be the moves required to lead to a winning result.

This may be illustrated in the following scenario. We note that, for convenience, in what follows we denote the addition/removal of an attack (a, b) simply by (ab, s) with $s \in \{+, -\}$, instead of $(ab, s, \#)$.

Example 31 (cont.) *Consider a slightly modified GB which contains also argument e and the debated attack (b, e). The system is illustrated in Figure 7.4. Assume that we are in the beginning of the debate, and no moves have been played yet. Let the Gameboard be $GB^0 = \langle A, R^{fix}, R^{deb}, Eval^0 \rangle$. We have $\vec{v}(a, b) = \vec{v}(d, c) = \vec{v}(b, e) = \langle 0, 0 \rangle$.*

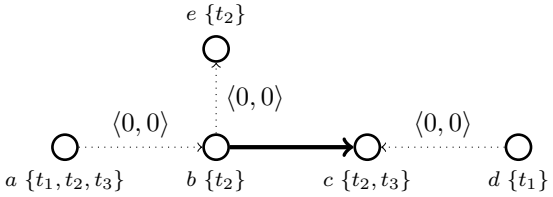


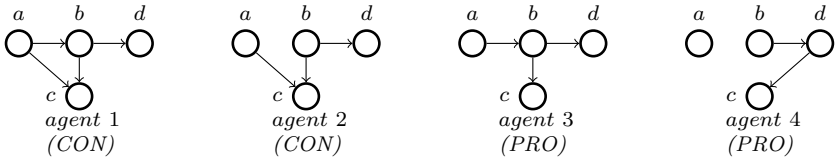
Figure 7.4: Agents who want c to be rejected, would better not try to defend e.

Assume that the debating agents argue on the acceptability of argument c. Agent 1, who belongs to PRO, is currently losing the debate, and he focuses on target set $\{(ab, +)\}$, because adding (a, b) will make c accepted. However, it is impossible for him alone to impose the attack (a, b), as both agents 2 and 3 will vote negatively on it. But now suppose that an agent of the CON team (e.g. agent 3) is very picky on the issue of the (new) argument e, and he has a strategy which says: "If e is attacked, then I will defend e" [8]*. Of course this strategy is not directly focused on the topic of the debate (which is c), but this kind of rhetorical move is common in real-life argumentation. In this case, agent 1 has an incentive to vote for (b, e) (provided that he is able to), and lure agent 3 in responding with a positive vote on (a, b). This way, thus not focusing always on target sets, agent 1 can eventually make c accepted and win the debate.*

Symmetric Equilibrium. The previous example showed that not focusing on target sets may in some cases lure the other group to make a "bad" move. Of course this relies on the rather artificial construction consisting of an agent playing a somewhat irrational strategy. We may then ask whether a weaker property can be guaranteed: is it the case that, *if the other agent follows a strategy consisting of playing within target sets*, then agents of the other side will not have an incentive to play differently, i.e. whether this constitutes a symmetric equilibrium. The following example, shows that this is not the case either.

Example 32 *Four agents have the argumentation systems shown below. We assume, for the sake of simplicity, that all the illustrated attacks are debated, that all the arguments concern a single topic, and that all the agents are expert on that topic.*

[8] For the sake of simplicity, let us assume that here agent 3 may violate his truthfulness.



The dialogue's issue is argument c. We have $CON = \{a_1, a_2\}$, $PRO = \{a_3, a_4\}$. *If both teams of agents play only in the targets sets, agents in PRO cannot win: at the beginning of the debate, agents in CON have two target sets* $\{(bc, +)\}$ *and* $\{(ac, +)\}$. *If they vote on* (a, c), *agents in PRO will be able to remove that attack (by voting twice if it is necessary). The remaining target set for CON will then be* $\{(bc, +)\}$. *Once CON agents vote on* (b, c), *agents in PRO will have two target sets:* $\{(bc, -)\}$ *and* $\{(ab, +)\}$. *Assume that* a_4 *votes against* (b, c). *Then agents in CON can vote again to reinstate it. Agents in PRO have then one remaining target set:* $\{(ab, +)\}$. *Once this vote is cast, the target set for CON is* $\{(ab, -)\}$. a_2 *votes against* (a, b), *and the agents in PRO cannot do anything else. In this case, PRO agents cannot win the debate.*

Assume now that agents in PRO do not play only in the target sets. As previously, at the beginning, agents in CON have two target sets, $\{(bc, +)\}$ *and* $\{(ac, +)\}$. *Once again, they can vote on* (a, c) *but these votes will be removed by agents in PRO. Once CON agents vote on* (b, c), *assume that* a_4 *votes on* (b, d). *The target set for CON is empty (as their goal is satisfied).* a_4, *for the PRO team, can play once more, so he chooses to add* (d, c), *and then to remove* (b, c). *The group CON has now two target sets,* $\{(ab, +)\}$ *and* $\{(bc, +)\}$. *Assume that* a_1 *votes for* (a, b). *Agents in PRO have now two target sets,* $\{(ab, -)\}$ *and* $\{(dc, -)\}$. *If* a_3 *votes against* (d, c), *agents in CON will have one target set,* $\{(ab, -), (bc, +)\}$. *Assume that* a_2 *votes against* (a, b), *and after everybody passes, he votes again for* (b, c). a_3 *can now vote for* (a, b). *Agents in CON cannot do anything else, as* a_2 *has already voted against* (a, b) *once. PRO wins the debate.*

Defining strategies and heuristics

From the above examples, we see that it is difficult to obtain theoretical guarantees on the efficiency of strategies focusing on target sets. This motivated us to study these strategies *experimentally*. We define 5 strategies, from the simpler to the more complex, focusing on target sets. Strategy 0 is the exception, as it is a *random strategy*, which will allow us to check whether playing in target sets is useful. We remind that, at any timestep, an agent is winning (resp. losing) the debate if the status of a given issue is the same (resp. is not the same) both in his private system and in the argumentation system associated to the GB. Note that when there are no available moves for an agent (we remind that an agent cannot vote on the same attack twice), that agent obligatorily passes.

Strategy 0: This is a random strategy, where: (1) if the agent is winning, then he plays pass; (2) otherwise, he votes randomly on an attack on the Gameboard (and if he is unable, then he plays pass).

Strategy 1: The idea of this strategy is that the agent can vote only if he is not satisfied by the current state of the Gameboard. Moreover, the agent can vote only if he can change the status of the issue (in other words, if he can change the sign of the weight of an attack in a target set of cardinality 1).[9] More precisely: (1) if the agent is winning, then he plays pass; (2) otherwise, he votes on an attack, but only if this vote changes the status of the issue (and if he is unable, then he plays pass).

Strategy 2: This strategy "improves" the previous one, as the agent can vote on a target set of cardinality greater than 1. In other words, the agent can vote on an attack of a target set if he can change the sign of its weight, even if the status of the issue remains unchanged.

[9] This strategy is the one studied in [BM11].

More precisely: (1) if the agent is winning, then he plays pass; (2) otherwise, he votes on an attack, but only if this attack belongs to a target set, and that vote changes the sign of the weight of that attack (and if he is unable, then he plays pass).

Strategy 3: With this strategy the agent can vote on an attack belonging to a target set, even if he cannot change the sign of the weight of that attack. More precisely: (1) if the agent is winning, then he plays pass; (2) otherwise, he votes on an attack which belongs to a target set, and towards changing the sign of its weight (and if he is unable, then he plays pass).

Strategy 4: This strategy "improves" the previous one as the agent, if he is currently winning, then he plays a move which renders the goal of the other team more difficult to be reached. More precisely: (1) if the agent is winning, then he votes on an attack which belongs to a target set for the goal of the other team and "reinforces" that attack (and if he is unable, then he plays pass).[10] (2) otherwise, the agent votes on an attack which belongs to a target set (for his goal), towards changing the sign of its weight (and if he is unable, then he plays pass).

As we can have several target sets, and several attacks in a target set, an agent can have several possible votes for each of these strategies. We thus introduce three heuristics to help an agent to choose a vote.

An agent can compute a set of possible votes, using any of the above strategies. Then, he can either randomly choose a vote among them, or use a more subtle heuristic. We have defined three heuristics which can be used for filtering the initial set of possible votes.

- **Heuristics A:** the agent randomly chooses a possible vote.
- **Heuristics B:** the agent filters out all possible votes on non-minimal (with respect to cardinality) target sets.[11] Then, he randomly chooses a vote.
- **Heuristics C:** the agent filters out all possible votes on non-minimal (with respect to cardinality) target sets. If he can change the sign of the weight of an attack among the remaining ones, then he filters-out all the attacks whose signs he cannot change. Then, he randomly chooses a vote.

The following example illustrates which moves are allowed by the different strategies (and heuristics) in a specific scenario.

Example 33 *Let there be a debate whose current state is represented by the Gameboard GB in Figure 7.5. The Gameboard has no fixed attacks, and it has four debated attacks. The weights of the debated attacks are illustrated next to them (their max-weights are not of interest to us). Let the argument c be the issue of the debate. Let T be a singleton, so every argument of the debate refers to the same topic. Also, let i be an agent who is expert on the sole topic ot T. Agent i has not yet voted on any attacks, and his private system is AS_i.*
At this point of the debate, the issue c is being accepted. There are three target sets for dropping c from the grounded extension: $\{(dc,+)\}$, $\{(ec,+)\}$ and $\{(ac,+),(ba,-)\}$.
Let us see the possible moves for agent i, according to the different strategies:

- *With* **strategy 1***: i will vote* **for (e,c)**.
- *With* **strategy 2***: i will vote* **for (e,c)**, *or* **for (a,c)**.
- *With* **strategy 3***: i will vote* **for (e,c)**, *or* **for (d,c)**, *or* **for (a,c)**, *or* **against (b,a)**.

[10] The motivation is to make it more difficult for the other team to later change the sign of the weight of that attack.

[11] For example, if an agent can vote on two attacks, the first being in a target set of cardinality 1, and the second in a target set of cardinality 2, then he will filter out the second option.

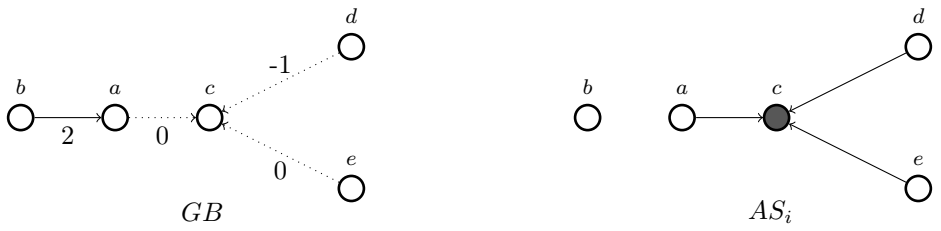


Figure 7.5: On the left: The Gameboard GB. For every debated attack (x, y), only its weight $w(x, y)$ is illustrated. On the right: The private system of agent $i \in CON$. Arguments belonging to $E_{Gr}(AS_i)$ are shown in white.

- *With* **strategy 4***: i will vote* **for (e,c)**, *or* **for (d,c)**, *or* **for (a,c)**, *or* **against (b,a)**.[12]

As far as the heuristics are concerned, let us focus on strategy 3.

- *With* **strategy profile 3A***: i will vote on a target set (he has 4 choices).*
- *With* **strategy profile 3B***: i will vote* **for (e,c)** *or* **for (d,c)**.
  The reason is that he focuses on target sets with minimum cardinality.
- *With* **strategy profile 3C***: i will vote* **for (e,c)**.
  The reason is that he focuses on target sets with minimum cardinality, and leading to attack additions/removals.

Strategies 3 and 4 propose different votes when the agent with the token is currently winning the debate, as shown next:

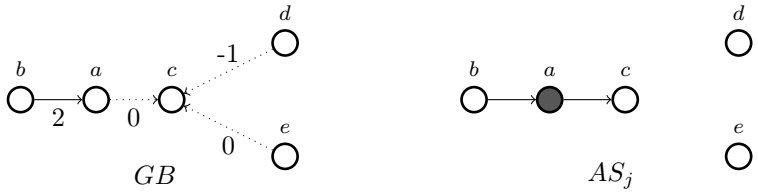


Figure 7.6: On the left: The Gameboard GB. For every debated attack (x, y), only its weight $w(x, y)$ is illustrated. On the right: The private system of agent $j \in PRO$. Arguments belonging to $E_{Gr}(AS_j)$ are shown in white.

- *With* **strategy 3***: j will vote* **pass**.
  The reason is that he is currently winning the debate.
- *With* **strategy 4***: j will vote* **against (e,c)**, *or* **against (d,c)**, *or* **for (b,a)**.
  The reason is that he can "reinforce" the target sets, instead of playing pass.

Coupling a strategy with a heuristics gives us a specific **strategy profile**. As Strategy 0 does not use target sets, it can not be coupled with any heuristics. Also, in Strategy 1 an agent can only vote on an attack if it belongs to a target set of cardinality 1 and he can change the sign of its

[12]The difference between Strategies 3 and 4 will become apparent later.

weight, so it does not make any sense to associate Strategy 1 with heuristics B or C. In the same way, in Strategy 2 an agent can only vote on an attack if he can change the sign of its weight, so it does not make sense to couple Strategy 2 with heuristics C. We thus have the following strategy profiles to consider (the number indicates the strategy type and the capital letter the heuristics): $SP = \{0, 1, 2A, 2B, 3A, 3B, 3C, 4A, 4B, 4C\}$.

We assume that the agents of the same group (PRO or CON) are using the same strategy profile during a debate. This is done in order to draw more easily conclusions on how the strategy profiles fare against each other. We can thus introduce the notion of **debate profile**. A debate profile is defined as a couple (SP_{PRO}, SP_{CON}) with $SP_{PRO}, SP_{CON} \in SP$. It indicates that all agents in the PRO (resp. CON) group are using the strategy profile SP_{PRO} (resp. SP_{CON}). Since there are 10 strategy profiles, there exist $10 \times 10 = 100$ different debate profiles. In the following, we first examine Strategy 0, and then we turn our attention to the 9 other strategy profiles which use target sets (thus on their corresponding $9 \times 9 = 81$ debate profiles).

An important remark is that the debate framework is quite rich, as the arguments' topics may differ a lot, the debating agents may have varying topics of expertise, and they may have multiple disagreements on the considered attacks. Also, at any given point, many moves may be available for an expert. As a result, it is relatively hard to come up with strict properties of the strategies and the heuristics, unless restrictive assumptions are made (on the structure of the experts' systems and on their expertise). This is why we have decided to evaluate the strategies and the heuristics, following an experimental approach. We generated many debate configurations (which roughly consist on the experts' private systems and expertise) and then we launched a significant number of debates. The results have been analyzed statistically, with respect to a number of criteria.

The debate framework has been implemented in Java. Next, we explain the details of the implementation.

The first key point is the (random) generation of the agents' private systems.

7.2.4 The experimental setting

In order to perform an important number of debates, our program is able to randomly generate different *debate configurations*. A debate configuration consists of three elements:

- A set of topics.
- A master argumentation system.
- A set of agents with their private systems and expertise. The private systems are in accordance with the master system.

For the experiments, the following choices were made:

Topics: There are $|T| = 6$ topics.

Master argumentation system: The master argumentation system (therefore also the GB and all the agents' systems) contains $|A| = 20$ arguments, each one randomly attached to one or two topics. The master system's graph has a density of attacks equal to 0.1. Among the attacks, there are 10 which are debated.[13] Finally, the issue is randomly chosen among the arguments in A.

Agents: Each debate involves 10 agents. Each agent is expert in one, two, or three topics, randomly chosen. For each agent i, his private system $AS_i = \langle A, R_i \rangle$ is in accordance with the master system, so it includes all its fixed attacks. Also, each debated attack of the master system belongs to R_i with a probability of 50%.

[13] We chose a small number of debated attacks, as this element causes an overhead in the computations of target sets.

The debating groups

A number of random configurations were generated using the above parameter values. When the difference of the PRO and CON cardinalities was important, the debates were trivial, as the majority easily won. In order to control the groups' cardinalities, we proceeded as follows: We generated 10 random configurations for each combination of PRO/CON cardinalities. So, we generated 10 configurations with 9 PRO and 1 CON agents (denoted 9/1), then 10 configurations with 8 PRO and 2 CON agents (denoted 8/2), and so on (7/3, 6/4, 5/5, 4/6, 3/7, 2/8, 1/9). The only PRO/CON cardinalities excluded from the experiments, were the extreme cases of 10/0 and 0/10, where all the agents belong to the same group.

The exact steps we took in order to generate 10 configurations of X PRO, and $10 - X$ CON agents, are the following:

1. Using the given parameter values, generate a random master system.
2. From that master system, randomly generate 10 private systems (remember that every debated attack has 50% probability of being in a private system).
3. If the groups' cardinalities are $X/10-X$ (as wished), then save this configuration. Otherwise, go to step 1.
4. If 10 configurations have been found, then stop. Otherwise, go to step 1.

A last remark on the generation of configurations just described: When we want to find groups with cardinalities $X/10 - X$, if we do not get these cardinalities immediately from the generation of 10 private systems (step 2), then we see that we discard the current master system (step 3), and we generate another one. The reason is that we may have a master system from which it is impossible to generate the cardinalities $X/10 - X$. For example, imagine a master system from which we can never obtain a PRO agent. In that case the program could get in an infinite loop (if that master system was not discarded). Another option would have been to examine, whether for the current master system, it is possible to obtain some PRO and some CON agents, and then to generate them according to our needs.

So this was the exact procedure with which we generated 90 debate configurations. Now, for each configuration, we tested all the 81 debate profiles focusing on target sets. Moreover, for every debate profile, the debate was repeated 10 times. The reason is that, as agents randomly choose their moves among a set of possible moves, the results of these 10 debates may differ. Practically, this happened quite rarely. Concluding, we have performed in total $90 \times 81 \times 10 = 72,900$ debates focusing on target sets.

The histogram in Figure 7.7 shows the percentage of agreement between the debates' results and the majority results, for each possible PRO/CON cardinalities. There are two interesting remarks to be made: First, when a group contains the vast majority of agents (8 or 9, out of 10), then the debate's result almost always agrees with the majority result. Indeed, the columns corresponding to 1/9, 2/8, 8/2 and 9/1, indicate an agreement with the majority close to 100%. These debates are, arguably, not very interesting, as their result is (almost) always predetermined. Therefore, we decided to filter out the cases of near-unanimity, and keep only the configurations where the PRO/CON cardinalities are 3/7, 4/6, 5/5, 6/4 or 7/3. In these configurations, though the majority always won more debates, the minority was also able to win sometimes. As a result, we focused on the corresponding 50 configurations (among the 90 initially generated).

Second, another interesting information available in the histogram is that the column of 7/3 (resp. 6/4) is bigger than the column of 3/7 (resp. 4/6). Also, the column of 5/5 is also relatively big, meaning that if there are 5 PRO and 5 CON agents, then PRO (by definition the majority in this case) usually wins the debate. In order to explain this fact, we conjecture that the master systems from which balanced PRO/CON groups were generated, slightly favoured the PRO group, as far as winning the debate is concerned. This fact merits a deeper study in the future. For example, it will be interesting to change some elements of the master system (e.g. increase the density of its attacks) and see if the winning percentage of PRO decreases.

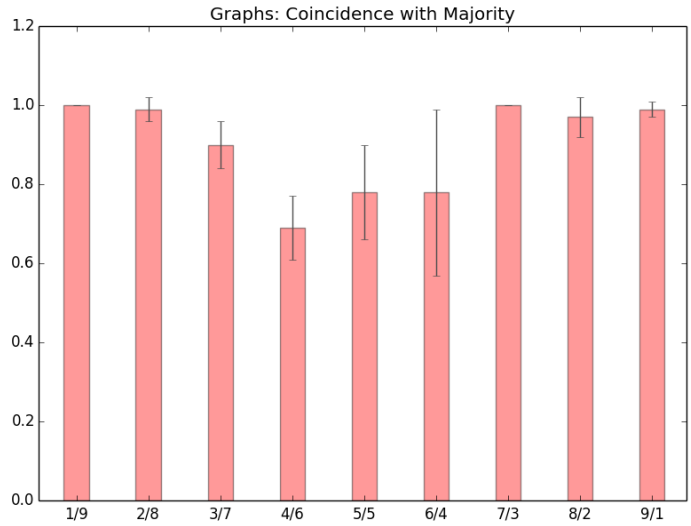


Figure 7.7: Histogram showing the coincidence of the debates' results with the majority results, for configurations having different cardinalities of PRO and CON groups.

7.2.5 Analyzing the results

In order to evaluate the quality of the strategies and heuristics, we first define four evaluation criteria of strategies and heuristics:

- **Probability to win**:
  Naturally, the agents prefer to use strategies and heuristics which will maximize their chances to win the debate.

- **Debate length**:
  It is the average number of timesteps in the debate. Normally, both the debating agents and the central authority of the debate prefer shorter debates, as longer debates need more resources, and in many cases there is the danger that the debate gets abruptly stopped, due to exogenous reasons, before the conclusion is drawn.

- **Happiness**:
  It is the percentage of coincidence between the debate's outcome and the majority's opinion. The key element here is the power of the minority to influence the debate's result. Will it be "powerless" against the majority, or will it have non-negligible chances of winning?

- **Rationality**:
  It is the percentage of coincidence between the debate's outcome and the merged outcome. We remind that the merged outcome is considered *rational*, because it aggregates all the agents' opinions on all the attacks. Therefore, as a rule of thumb, it is preferable to have debates whose results agree with the merged outcome.

Now that we have defined four evaluation criteria for the strategies and heuristics, we are able to evaluate them, by analyzing the results of our experiments. We first analyze the results of the random strategy, and afterwards we focus on the 81 debate profiles which are based on target sets.

The random strategy

We begin our analysis with the random strategy. We remind that by following this strategy, an agent who has the token will just randomly choose a vote he has not cast yet. This strategy is

considered in our experiments with the goal of underlining the advantage of focusing on target sets.

As far as the probability of winning the debate is concerned, the random strategy did not fare worse than the quite simple strategy profiles 1 and 2X (with X $\in \{A, B\}$). The reason is that its drawback (the fact that agents playing random attacks could harm their own group), was balanced by the drawback of strategy profiles 1 and 2X, which can "block" a group in a losing position, even if the combined votes of two (or more) agents would have changed the issue's status (we remind that, in these cases, strategy profiles 1 and 2X prevent the first voter from casting his vote).

On the other hand, we expected that the winning percentage of a group would increase, if instead of the random strategy, he used the elaborated strategy profiles 3X and 4X. This was indeed verified, as the winning percentage always increased, up to 25% in some cases (although less in others). Furthermore, we conjecture that the more attacks the GB has, the worse the results will be for the random strategy, compared to any other strategy profile. The reason is that the more attacks the GB has, the more dangerous it is for a group to randomly vote on attacks, as the probability of votes backfiring is increased.

Another major disadvantage of the random strategy is that, if a group uses it, then the number of timesteps of the debate explodes. In most cases, when one group adopted the random strategy, the number of timesteps increased by a factor of 10 (e.g. from 25 timesteps, into 250 timesteps). This was expected, as the strategy profiles based on target sets, propose moves focused on quickly changing the issue's status. This is not the case in the random strategy, where a group can play a lot of "dummy" moves before (by chance) it achieves its goal. Even worse, when both groups used the random profile, the number of timesteps doubled.

On a positive side, if a group used the random strategy, then the percentage of agreement with the merged outcome was quite high (in almost all cases this percentage was greater than 90%). Naturally, the reason behind this, is that a group using the random strategy will cast too many votes during the debate, and as a result, the GB will resemble more to the merged system. This was even clearer when both groups used the random strategy, when the percentage of agreement with the merged outcome went up to 97.6%.

Concluding, the fact that the number of timesteps increases dramatically when a group uses the random strategy, as well as the fact that it fares worse (as far as winning the debate is concerned) than strategy profiles 3X and 4X, are two good reasons to prefer, in practice, strategies focusing on target sets, than the random strategy. Therefore, in the following tests we do not include the random strategy, and we just consider the 9 strategy profiles focusing on target sets.

Strategies based on target sets

We now turn our attention to the 9 strategy profiles focusing on target sets and to the corresponding 81 debate profiles. Each of the four following graphics contains information on *all* debate profiles focusing on target sets. The top left shows the percentage of PRO wins (for every profile), the top right shows the average number of timesteps of the debates, the bottom left shows the percentage of agreement between the debate's outcome and the merged outcome, and the bottom right shows the percentage of agreement between the debate's outcome and the majority's opinion. PRO strategies are shown on the left side, and CON strategies on the right side of every graphic.

- Criterion of **winning probability** (Figure 7.8, top-left):
  Regarding the strategy which is most likely to win a debate, the most elaborated strategies 3 and 4 provide a clear advantage. PRO's best chance to win is when the profile (4X,1) is used (75% probability of PRO winning). Similarly, CON's best chance to win is in profile (1,4X) (38% probability of PRO winning). In general, no matter what strategy a group is using, the other group increases its winning percentage if it uses strategy 3 or 4, instead of the simpler strategies 2 and 1 (1 being the worst choice). Finally, it is clear in the graph, that (as mentioned before) PRO wins more debates than CON. This is apparently related to the nature of the random master systems from which balanced PRO/CON groups are generated.

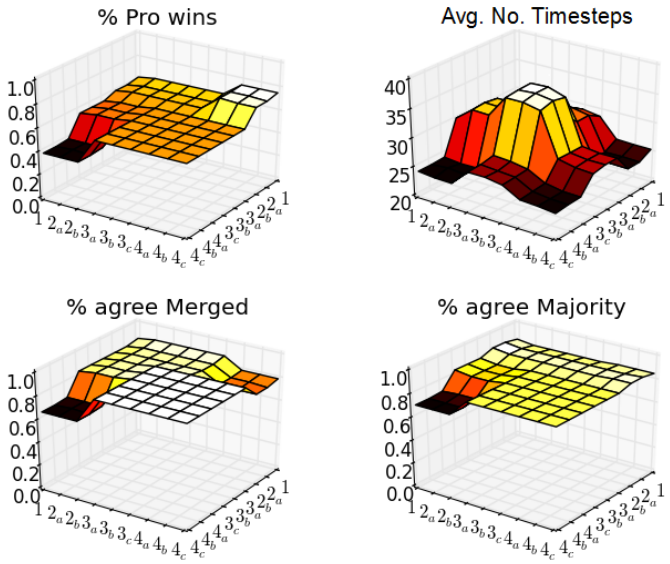


Figure 7.8: Top-Left: Percentage of wins by PRO. Top-Right: Average number of timesteps of the debates. Bottom-Left: Percentage of agreement with the merged outcome. Bottom-Right: Percentage of agreement with the majority's opinion. PRO strategies are shown on the left side, and CON strategies on the right side of every graphic.

- Criterion of **debate length** (Figure 7.8, top-right):
  The lowest number of timesteps is found when both agents employ strategy 4. A small number of timesteps is also obtained in profiles (1,4X), (4X,1) (where X $\in \{A, B, C\}$) and (1,1). For the latter, the reason is that there are cases where a group cannot vote on an attack because no single agent can change it (and thus the debate stops). For profiles (4X,4Y) the reason debates are short is that agents are not forced to play (useless) pass-moves when they are winning. Instead, they can reinforce attacks on the GB, something that is proven useful later. This is not possible with profiles (3X,3Y) which give the longest debates. Note that agents using the strategy profiles 4X have incentive to disclose more information than with the other strategy profiles. This can be seen as a disadvantage of the strategy, in case the agents wish to hide information. More specifically, if we concentrate only on timesteps which do not contain pass moves (let us call them *no-pass timesteps*), then the results of strategy profiles 3X and 4X are inversed. Strategy profiles 4X lead to more no-pass timesteps, than profiles 3X. E.g. (4C,4C) leads in average to 11.97 no-pass timesteps, while (3C,3C) leads in average to 10.36 no-pass timesteps. We clearly see that when profiles 3X are used, many timesteps involve pass moves, and this is the reason why profiles 3X have the greatest total number of timesteps.

- Criterion of **rationality** (Figure 7.8, bottom-left):
  The most "rational" debate outcomes (closer to the outcomes of the merged system) are obtained when both groups use one of the strategies: 3A, 3B, 3C, 4A, 4B, 4C (the probability of agreement being 88%). The only cases where the outcomes of the debates are farther from the merged outcomes are when a group uses strategy profile 4X and the other group uses a simple strategy profile, 1 or 2X. So, we pull away from the merged outcome when a group uses the most advanced strategy (4X), while the other a simple one (1 or 2X). The smallest probability of agreement is 66%, corresponding to profile (1,4X).

- Criterion of **happiness** (Figure 7.8, bottom-right):
  Similar results are obtained when we focus on happiness. Almost all profiles give a similar value of agreement with the majority (the probability is close to 85%). However, when PRO uses strategies 1, or 2X, and CON uses 4X, the debate's outcome starts to move away from the majority's opinion (its minimum value is 70%). Interestingly, we were expecting the same to happen at the other end of the graphic (as in the bottom-left graphic), but it did not. This might be related to the relative advantage PRO agents have, as far as winning is concerned. Testing more configurations will help understand this behaviour.

The role of the heuristics

Let us make some remarks on the three heuristics A, B and C. The main reason for using heuristics is to accelerate the debate. Heuristics C which focuses on the smallest target sets, and prefers moves able to add/remove an attack, was expected to lead to the quickest debates. This was verified, although its results were not significantly better than the results of the simpler heuristics B and A. For example, the debate profile (4C,4C) lead to 23.88 timesteps in average, while the profile (4A,4A) lead to 24.81. Also, the debate profile (3C,3C) lead to 35.29 timesteps in average, while the profile (3A,3A) lead to 36.29. We conjecture that, when heuristics C is used instead of B or A, the decrease in the number of timesteps is small, due to the fact that the randomly generated systems do not contain many target sets, and these target sets do not have great differences in size. We expect that in the case of master systems with target sets of considerably different sizes, heuristics C will lead to a more significant decrease in the number of timesteps, compared to heuristics B and A. Finally, no particular effect of the heuristics on the winning probability percentage has been noticed.

7.2.6 Conclusion on strategies and heuristics

To conclude, a general observation is that the more sophisticated strategy profiles (3X and 4X) are the best choices for the agents who want to increase their probability of winning the debate. Their main difference lies on the average number of timesteps, and on the amount of information disclosed during the debate. Surprisingly, the simpler strategy profiles (1 and 2X) offer an interesting alternative, provided that the debate's central authority can ensure that both groups will use a simple strategy profile, and that no group will switch into using a sophisticated one. This is supported by the fact that, in our experiments, the probability that the winner is the same, in case of profile (1,1) and of profile (3C,3C), was almost 95% (this information does not appear in the previous graphics). Finally, the use of heuristics C shortens the length of the debates, though more tests are needed in order to evaluate its impact.

7.3 Debate Protocols Using Bipolar Argumentation Frameworks

Many protocols for argumentative debates rely on Dung's original framework of argumentation, as for example the previous protocol and the protocol defined in [BM11]. In the following, we will use bipolar argumentation frameworks, which allow support and attack relations between arguments. Given the common use of both types of relations in everyday argumentation, we argue that our work may have interest in practical debate settings. Moreover, we will be using numerical argument valuation, and we believe that in some settings, this choice can be advantageous, as it provides a fine-grained characterisation of an argument's status.

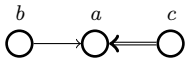
7.3.1 Brief reminder of bipolar argumentation

Bipolar argumentation considers two types of relations between arguments. The first one is an *attack* relation over arguments (denoted R), as in the classical framework proposed by [Dun95].

The second type of relation over arguments is a *support* relation (denoted S). If (bSa) then argument b supports argument a, in other words b is a reason to believe that a holds. It is reminded that the support relation is different than the defence relation defined in [Dun95], because the notion of defence depends on the notion of attack, while the notions of support and attack are independent. Support relations between arguments are often used in everyday life and this makes them particularly interesting for modeling debates.

According to Definition 12 of a bipolar argumentation framework (BAF), a BAF is a triplet $\langle A, R, S\rangle$ which consists of a set of arguments A, an attack relation R over A, and a support relation S over A. If $a, b \in A$, then aRb (resp. aSa) means that a attacks (resp. supports) b. A BAF may be represented by a directed graph called *bipolar graph.*
In the following, the attacks (resp. supports) on a bipolar interaction graph are represented by simple (resp. double) arrows, as in the following example.

Example 34 *The* $BAF = \langle\{a, b, c\}, \{(b, a)\}, \{(c, a)\}\rangle$ *is represented by the following bipolar graph. The argument a is attacked by b and it is supported by c.*



Using gradual local valuation

In Subsection 2.1.3 we presented some of the existing approaches for argument valuation on a BAF. Here, we shall use the Definition 13 of local gradual valuation. The reason is that, in the debates we define later, we will assign numerical valuations to the arguments which have been put forward. In a real debate setting, this can provide an indication of *how much* an argument is accepted or rejected. For example, if the possible valuations of an argument are in the interval $[-1, 1]$, then an argument with a valuation of 0.9 could be considered as *more accepted* than an argument with a valuation of 0.1. In some cases we may prefer to have such expressivity in the results, instead of (for example) partitioning arguments into accepted and rejected. Also, we will make the simplifying assumption that the arguments and the attacks put forward in the debate do not have any initial weights.

We remind that, when local gradual valuation is used, the value of an argument only depends on the values of its direct attackers and supporters. Note that the above definition produces a *generic* local gradual valuation, and there exist several instances for it. In the remainder of this work, we will use the following instance, which has been proposed in [ACLSL08]:

- $\mathcal{V} = [-1, 1]$ interval of reals;
- $H_R = H_S = [0, \infty]$ interval of reals;
- $h_R(x_1, \ldots, x_n) = h_S(x_1, \ldots, x_n) = \sum_{i=1}^{n} \frac{x_i+1}{2}$;
- $g(x, y) = \frac{1}{1+y} - \frac{1}{1+x}$.

According to this instance, if an argument has no attackers or supporters, then its valuation is 0. It can also be 0 when its attackers and supporters "balance" each other. If the aggregated attack it receives (computed by h_R), is greater than the aggregated support it gets (computed by h_S), then the argument will have a negative valuation, otherwise it will have a positive valuation. It could be argued that these characteristics make the above instance useful in some real-life argumentation settings.

7.3.2 Merged Bipolar Argumentation Framework

The notion of *merged argumentation system* has been introduced in [CMDK+07] for classical abstract argumentation frameworks. A meaningful way to merge several argumentation systems which share exactly the same arguments but with possible conflicting views on the attack relations between them is to take the *majority argumentation system*. In this system, defined in [CMDK+07], attacks supported by the majority of agents are kept, and ties are broken in favour of the absence of the attack.

We propose here a similar process for bipolar argumentation frameworks. We consider a set of agents N, each one having an abstract bipolar argumentation framework $BAF_i = \langle A, R_i, S_i \rangle$. All these bipolar argumentation frameworks share exactly the same arguments, but express possibly conflicting views on the attack and support relations. The merged bipolar argumentation framework (denoted $MBAF_N$) is then defined in the following way:

Definition 64 *Let N be a set of agents and let $\langle BAF_1, \ldots, BAF_n \rangle$ be the collection of their private bipolar argumentation frameworks, sharing the same set of arguments A. Let $Attack(a,b) = \{i \in N \mid (a,b) \in R_i\}$ and let $Support(a,b) = \{i \in N \mid (a,b) \in S_i\}$. Then, the* **merged bipolar argumentation framework** *is denoted by $MBAF_N = \langle A, R_N, S_N \rangle$, where $R_N \subseteq A \times A$, $S_N \subseteq A \times A$ such that:*

- aR_Nb *iff* $|Attack(a,b)| > |N| - |Attack(a,b)|$
- aS_Nb *iff* $|Support(a,b)| > |N| - |Support(a,b)|$

Therefore, in order for an attack or support to appear on $MBAF_N$, it must be supported by a strict majority of the agents. Let us see an example of computation of a merged bipolar framework.

Example 35 *Let $N = \{a_1, a_2, a_3\}$ be a set of three agents, and let $BAF_{a_1} = \{\{a,b,c\}, \{(b,a)\}, \{(c,a)\}\}$, $BAF_{a_2} = \{\{a,b,c\}, \{(b,a),(c,b)\}, \{\}\}$, $BAF_{a_3} = \{\{a,b,c\}, \{(c,b)\}, \{\}\}$ be the agents' private frameworks. In Figure 7.9 the agents' frameworks and the merged framework $MBAF_N$ are illustrated, and the valuations of their arguments are shown.*

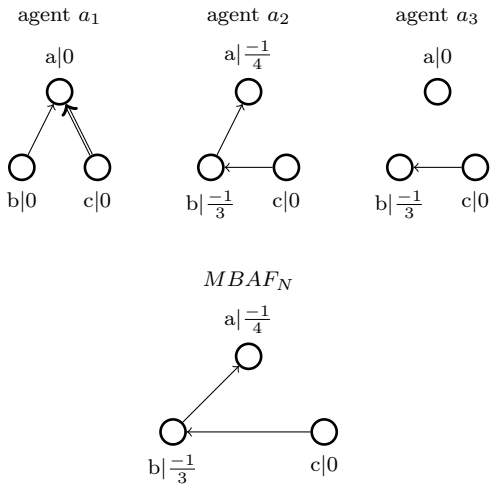


Figure 7.9: Three BAFs and their corresponding merged framework.

Let us now analyze the arguments' valuations. In the first graph, the attack bRa and the support cSa "balance" each other, in the sense that $v(a) = 0$. In the second graph, argument b is "weakened" by the attack cRb (so we have that $v(b) = \frac{-1}{3} < 0$), therefore the effect of the attack

bRa on argument a is also weakened. Still, the valuation of a is negative, as $v(a) = \frac{-1}{4}$. *Finally, in the third graph, argument a receives no attacks or supports, so* $v(a) = 0$.

As far as the merged framework $MBAF_N$ *is concerned, we see that an attack or support appears on* $MBAF_N$ *if and only if it is supported by a strict majority of agents. Here, this is the case for the attacks cRb and bRa. The valuation of a is* $v(a) = \frac{-1}{4}$.

The merged framework is arguably the most natural way to obtain a rational collective result, in the sense that the focus is on the elementary parts of the debate (here the attacks and supports) and disagreement on every single one is first addressed. Once the collective opinion on every attack and support is obtained, it is possible to put every partial conclusion together and to compute the valuations of the arguments.

On the other hand, as shown in [CMDK+07] for the case of abstract argumentation systems and extension based acceptability, the results of the merged system may be, in a sense, controversial. More specifically, if we focus on the acceptability of a single argument, the merged framework may contradict the majority's opinion and, even worse, in some cases it may even contradict the unanimity's opinion. In our setting, where bipolar frameworks and numerical valuation of arguments are used, we face a similar problem: an argument's valuation in the merged framework may be quite different than its valuation in the majority of the private systems. Also, it may be different than the valuation in all the private systems, thus the unanimous opinion of the agents is not obligatorily respected. This can be seen in the following example.

Example 36 *Let* $N = \{a_1, a_2\}$ *be a set of two agents whose private frameworks are illustrated in Figure 7.10.*

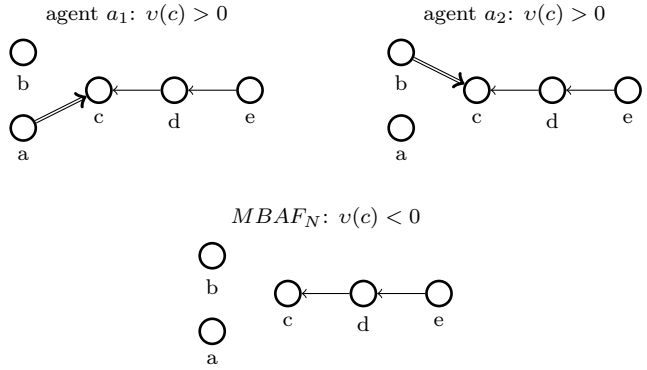


Figure 7.10: The valuation of argument c in the $MBAF_N$ differs from the agents' valuations.

Notice that, for both agents, $v(c)$ *has a positive value, because the effect of the support aSc (resp. bSc) is greater than the effect of the attack dRc (which is weakened by eRd). Nonetheless, in the merged framework, argument c has a quite different (negative) valuation, as it has no supporters. We see that, in the general case, even if the unanimity of the agents agree on the valuation of an argument, its valuation on the merged system does not coincide with the agents' valuations.*

7.3.3 Designing protocols - focus on the agents' goals

An essential element of multi-agent debates is the definition of the types of goals that participants strive to achieve. In the following, as in Section 7.2, all the agents are *focused* on the same argument which is called *issue* of the debate. Moreover, an agent is able to estimate, at every point of the debate, how satisfied he is with the current situation, by comparing his personal viewpoint (argumentation system) with the collective viewpoint (the Gameboard). While in Section 7.2 an agent wanted the issue to belong (or to not belong) in the grounded extension of both his personal system and of the Gameboard, in this section another approach will be tried: numerical argument

valuation will be used instead of argument extensions. Therefore, in the bipolar argumentation setting we present next, an agent compares the valuation of the issue in his personal system with its valuation on the Gameboard. The agent goals we define are based on the idea that, an agent wants his valuation of the issue to be "similar" to the collective valuation of the issue.

In fact, after careful inspection, we see that when using gradual valuation of arguments, different approaches can be chosen. Let us provide a motivating example, where there are four debating agents. Suppose that the issue of the debate is argument d, and that the valuations of the issue by four agents a_1, a_2, a_3 and a_4, are respectively: $v_1(d) = -0.9, v_2(d) = -0.1, v_3(d) = +0.1, v_4(d) = +0.4$. In a debate, it seems natural and intuitive to separate the agents in different groups, according to their viewpoint on the issue. In Chapter 6 we have defined two main types of protocols, according to the way agents are regrouped: category-based protocols and cluster-based protocols. Let us provide a brief reminder.

- **Category-based protocols**: In these protocols the agents are divided in k predefined categories $K_1, \ldots, K_k$. This is done with the help of $k-1$ threshold values, which partition the range of an argument's possible values into k intervals. Notice that we use the word "category" (or "group"), instead of the word "team", because we assume that the agents of the same category cannot coordinate. In the extreme case where $k = 2$, we can see the debate as opposing two groups, named PRO and CON, such that: $CON = \{a_i \in N \mid v_i(d) < thr\}$ and $PRO = \{a_i \in N \mid v_i(d) \geq thr\}$.
- **Cluster-based protocols**: In these protocols no agent categories are defined beforehand, but agents may be clustered in k groups $C_1, \ldots, C_k$, on the basis of the similarity of their valuations of the issue (in the sense of minimizing the maximum difference between two agents of a group).

We make no claim on the relative merits of these approaches, and we believe that they may be justified in different contexts. If there are some exogenously given, meaningful threshold values, then dispatching agents in the adequate categories seems to be the best choice. The simplest option, which we later examine, is to define just two categories, thus simply partitioning the agents into those in support of the issue, and into those against it. If three categories are defined, then it is possible to have a group of agents who neither accept, nor reject the issue, but they are undecided towards it.

Instead, if the purpose is to regulate the dialogue in a more flexible way, thus not using predefined "opinions" (categories) regarding the issue, then it may be more appropriate to opt for a cluster-based approach.

In the following we focus on two extreme versions of these types of protocols, as already mentioned: a category-based protocol with $k = 2$ (named π_0), and a cluster-based protocol with $k = n$ (named π_1). Note that π_0 is more in line with persuasion dialogues which confront two opposing views.

The bipolar Gameboard

But, what exactly are the agents allowed to do during a debate? How do they pursue their argumentative goals? Here we follow the work in [BM11], as we assume that the debating agents contribute step-by-step to the debate, by inserting arguments on a common *Gameboard* and by stating their opinions on the validity of attack and support relations. Moreover, in accordance to [BM11], the agents do not disclose their private argumentation systems to a central authority, neither to the other agents. Therefore, in our setting agents are not able to directly coordinate with others and to collectively devise a strategy. Nonetheless, strategic considerations can be taken into account, as we shall see later, based on the state of the *Gameboard*.

The bipolar Gameboard (or simply Gameboard) has a couple of important differences compared to the Gameboard in Definition 23. First, the bipolar Gameboard has a support relation over its arguments. Second, we do not take agent expertise into account, and the aggregation of opinions is simplified. Therefore, the bipolar Gameboard is a BAF.

Definition 65 *The* **bipolar Gameboard**[14]*, at timestep* $t \in \mathbb{N}$*, is a BAF denoted by* $GB^t = \langle A^t_{GB}, R^t_{GB}, S^t_{GB} \rangle$*, where:* A^t_{GB} *is the set of arguments which have been introduced in the debate until* t*,* $R^t_{GB} \subseteq A^t_{GB} \times A^t_{GB}$ *is the set of attacks between arguments in* A^t_{GB}*, and finally* $S^t_{GB} \subseteq A^t_{GB} \times A^t_{GB}$ *is the set of supports between arguments in* A^t_{GB}*.*

The debate starts at timestep $t = 0$, when the Gameboard contains just the issue and no attacks or supports, so: $GB^0 = \langle \{a\}, \{\}, \{\} \rangle$, where a is the issue.

We shall denote by $v^t_{GB}(a)$ the valuation of argument $a \in A_{GB}$, at timestep t.

Using attraction values

We repeat that agents are *focused*, in the sense that they all concentrate their attention on a specific argument which is the *issue* of the debate [Pra05]. Depending on the type of debate considered, the agents may have different preferences over the issue's valuation. Here we define agent preferences with respect to some *attraction values*. Roughly, an agent wants the issue's valuation on the Gameboard to be as close as possible to his attraction value.

In principle, in category-based protocols, agents of category K_j prefer any state of the debate where the issue's valuation lies in K_j, regardless of the exact value. But, if an agent takes into consideration the dynamic nature of the *Gameboard* and wants to increase the probability that the issue's valuation stays in the wished interval[15] $K_j = [K^l_j, K^u_j]$, then he may use as attraction value the central value $(K^u_j + K^l_j)/2$. The reason is that this value has a maximum distance from the other "neighbour" categories. Categories K_1 and K_k are somewhat different, as they have only one "neighbour" category (K_2 and K_{k-1}, respectively), so for them it is reasonable to choose the extreme values -1 and 1 as attraction values.

In cluster-based protocols, the interpretation of the attraction value is similar: The agents seek to turn the issue's valuation on the GB as close as possible to the *representative value* of their cluster, which can be defined in different ways, e.g. it may be the mean or the median of the agents' valuations.

Getting back to our protocols π_0 and π_1, the question is how we shall instantiate the notion of attraction value. In the following, let d be the issue of the debate. Also, a move which occurs at round t yields a value $v^{t+1}_{GB}(d)$ after it is played on the GB. We make the following choices:

- For π_0, where we have two categories, we shall consider that the attraction values are 1 (for PRO) and -1 (for CON). The intuition behind this choice is that the group PRO (resp. CON) increases its probability to win the debate if its agents try to increase (resp. decrease) the issue's valuation as much as they can. Therefore, a state at $t+1$ is preferred over a state at t by PRO (resp. by CON) iff $v^{t+1}_{GB}(d) > v^t_{GB}(d)$ (resp. if $v^{t+1}_{GB}(d) < v^t_{GB}(d)$).

- For π_1, the attraction value will be simply $v_i(d)$, as we consider that a state at $t+1$ is preferred over a state at t (by agent i) if and only if $|v^{t+1}_{GB}(d) - v_i(d)| < |v^t_{GB}(d) - v_i(d)|$. In words, each agent seeks to turn the valuation of the issue on the GB as close as possible to its own valuation.

These different choices can yield very different behaviours, as it can be seen in the following example. Let i be an agent who holds the view that $v_i(d) = -0.1$. If the threshold is set to 0, then under the protocol π_0, agent i prefers a state of the debate with $v_{GB}(d) = -0.9$, than a state of the debate with $v_{GB}(d) = +0.1$. On the other hand, under the protocol π_1, agent i will have the opposite preference.

[14] It will be simply called *Gameboard* when there is no danger of confusion.

[15] K^l_j (resp. K^u_j) denotes the *lowest* (resp. *highest*) value of the interval K_j

Formal definition of π_0 and π_1

We now define the two protocols which are built on the same underlying principles. No coordination takes place among the agents, they may for instance play asynchronously, and then the central authority will simply pick a permitted and relevant move, and place it on the *Gameboard*. The moves that agents are allowed to make during a debate are called *permitted moves*. They are sent to the central authority which will choose, at every time, one and place it on the GB. Before choosing a move, the central authority will filter the received permitted moves and it will only consider the moves among them which are considered *relevant*. Roughly, relevant moves are those moves which "advance" the debate, given its current state and the participants. Therefore, at timestep t, we define permitted and relevant moves as follows.

- **Permitted moves**: They are simply positive assertions of attacks xRy (resp. of supports xSy) with $y \in A^t_{GB}$, or contradictions of (already introduced) attacks (resp. supports) with $(x, y) \in R^t_{GB}$ (resp. $(x, y) \in S^t_{GB}$). This means that the attacked (resp. supported) argument must already be on the GB, at timestep t, while the attacking (resp. supporting) argument may or may not be there, at timestep t. Note that arguments are progressively added on the GB via adding attacks and supports, and that it may not contain the whole set of arguments when the debate finishes. Furthermore, the protocols prevent an agent from repeating the same move. To account for this, each agent a_i is equipped with a commitment store. It is a set $RP^t_i \subseteq A^t_{GB} \times A^t_{GB}$ (resp. $SP^t_i \subseteq A^t_{GB} \times A^t_{GB}$) which contains the attack and non-attack relations (resp. support and non-support relations) he has already expressed his opinion on,[16] until timestep t. This way agents can be prevented from stating twice the same relation.

- **Relevant moves**: The central authority filters the permitted moves proposed by the agents, and only considers the relevant moves among them. In a sense, any move which changes the current valuation of the issue might be deemed relevant, as it brings a modification of the current status of the issue. However, more specifically, we make the following choice: a move of agent i will be considered *relevant* at round t if the updated Gameboard is preferred by the group (category or cluster) of agent i. In other words, a move is relevant for a group if it gets the valuation closer to the attraction value of the group.

When a move is played on the Gameboard, the update operation given in Table 7.1 takes place.

Let $GB^t = \langle A^t_{GB}, R^t_{GB}, S^t_{GB} \rangle$	
1. After the **assertion of** xRy:	$GB^{t+1} = \langle A^t_{GB} \cup \{x\},\ R^t_{GB} \cup \{(x, y)\},\ S^t_{GB} \rangle$
2. After the **assertion of** xSy:	$GB^{t+1} = \langle A^t_{GB} \cup \{x\},\ R^t_{GB},\ S^t_{GB} \cup \{(x, y)\} \rangle$
3. After the **contradiction of** xRy:	$GB^{t+1} = \langle A^t_{GB},\ R^t_{GB} \setminus \{(x, y)\},\ S^t_{GB} \rangle$
4. After the **contradiction of** xSy:	$GB^{t+1} = \langle A^t_{GB},\ R^t_{GB},\ S^t_{GB} \setminus \{(x, y)\} \rangle$

Table 7.1: Update of the Gameboard

Note the asymmetry here: introducing a new argument can only be done via an attack or support assertion, since it is never permitted to contradict an attack (or support) referring to an argument that was not already introduced.

When no agent wants to contribute to the debate any more, we say that the GB is *stable*. The *outcome of the debate* refers to the debate's winners and losers.

[16] We call "non-attack" and "non-support" the relations which an agent does not possess in his own system, therefore he objects to their validity.

1. Agents report their individual view on the issue to the central authority, which assigns (privately) each agent to PRO or CON.
2. The first round starts with only the issue on the Gameboard.
3. $ToPlay \leftarrow CON$
4. While $Winner = \emptyset$ loop
   (a) agents of $ToPlay$ independently propose moves to the authority (*ProposedMoves*);
   (b) the authority filters relevant moves only (*RelevantMoves*);
   a relevant move for PRO (resp. CON) increases (resp. decreases) the valuation of the issue on the GB;
   (c) if $RelevantMoves = \emptyset$
   then $Winner \leftarrow \overline{ToPlay}$
   else $PickedMove \leftarrow select(RelevantMoves)$
   (d) update the GB with *PickedMove*
   (e) if $sign(v_{GB}^{t+1}) \neq sign(v_{GB}^{t})$ then $ToPlay \leftarrow \overline{ToPlay}$

Table 7.2: Protocol π_0 (category-based, $k = 2$)

We now give the details of our two illustrative protocols. As far as notation is concerned, if $X = PRO$ (resp. $X = CON$), then let $\overline{X} = CON$ (resp. $\overline{X} = PRO$). As mentioned before, π_0 is a two-sided category-based protocol. It is described in Table 7.2. On the other hand, π_1 is a cluster-based protocol with n agents playing with the objective to make the value of the issue matching their own valuation. It is described in Table 7.3.

An important issue of the protocols is the turn-taking. When should the turn be given to another group, and which one should be chosen? It is important to notice the difference in turn-taking between π_0 and the two-sided acceptability-based protocol proposed in [BM11]. In this latter protocol, the notion of relevance coincides with the condition for turn-taking: each relevant move passes the turn (by definition) to the other side. In π_0 this is not the case (and neither in π_1). More specifically, in π_0, a move may be relevant, thus helping a group change the sign of the issue's valuation, but not sufficient to switch the turn by itself. In that case, the group will be given the possibility to play additional moves, until it succeeds to switch the turn. If this does not happen, and the group has no more relevant moves, then it loses the debate.[17]

Furthermore, we clearly see another type of design choice that can be made in these protocols. In π_0 relevant moves correspond to moves that update the Gameboard in a way that is in line with the preferences of the agents. However, once several relevant moves are proposed, how shall the authority *select* which relevant move to pick? Initially, we make a simple choice: in π_0 the central authority picks the first (or a random) move from *RelevantMoves*. Then, in π_1, we make a slightly more elaborate choice: the function *select* chooses the move which changes *the most* the value of the issue on the Gameboard.

Types of protocol properties

We now present some types of properties that these protocols may, or may not, satisfy. We remind that the merged system is an aggregated system whose conclusions are considered *rational*. Therefore, the outcome of a debate resulting from a specific sequence of moves, obeying one of these protocols, will typically be compared with the merged outcome. We would like to know, for example, if these two outcomes are always identical, or if they may differ, under some circumstances. The basic types of properties of these protocols are the following:

[17] Observe also that, at the end of a two-sided protocol, the group winning the debate is $\overline{ToPlay}$.

1. While $Winner = \emptyset$ loop
   (a) agents independently propose moves to the authority (*ProposedMoves*);
   (b) the authority filters relevant moves only (*RelevantMoves*);
   an agent's relevant move makes the valuation of the issue on the GB more preferable to that agent;
   (c) if $RelevantMoves = \emptyset$
   then $Winner \leftarrow argmin_i |v_i(d) - v_{GB}(d)|$
   else $PickedMove \leftarrow select(RelevantMoves)$
   (d) update the GB with *PickedMove*.

Table 7.3: Protocol π_1 (cluster-based, $k = n$)

- **Termination:**
  Is it certain that the debate will always terminate after a finite number of moves by the agents? For these protocols, termination is trivially guaranteed, as we assume:
  1. A finite number of participants $|N|$.
  2. Private argumentation systems with a finite number of arguments, and therefore a finite number of attacks and supports.
  3. No agent can repeat a move during the debate.

  The two extreme cases, as far as the number of moves in a debate is concerned, are the following: (i) The debate consists of 0 moves, if at the beginning of the debate every agent is PRO, thus satisfied with the issue's valuation. (ii) The debate consists of $|N| \times 2 \times |A|^2$ moves, if all $|N|$ agents play two moves on every pair of arguments.[18] In the latter case, the stable GB at the end of the debate is obligatorily identical to the merged system.

- **Non-determinism:**
  Does the debate's outcome depend on the agents' (or the central authority's) choice of moves? In the following we see that non-determinism holds, in the general case. We argue that a positive effect is that the agents' strategic considerations during the debate can play an important role on the outcome.

- **Convergence of the debate to the $MBAF_N$:**
  Is it guaranteed, possible, or impossible to obtain, at the end of the debate, a result identical to the result of the $MBAF_N$? Here, we may refer to one of the two following comparisons:
  1. The debate's stable GB with the $MBAF_N$.
  2. The debate's outcome with the merged system's outcome.

  Of course, if the debate's stable GB is identical to the $MBAF_N$, then the outcomes are also identical. On the other hand, if the outcomes are identical (the winners coincide in the debate and in the $MBAF_N$), it does not follow that the two systems are identical.

- **Agreement of the majority (resp. unanimity) with the outcome of the debate (and of the $MBAF_N$):**
  Is it guaranteed, possible, or impossible that the majority (resp. unanimity) agrees with the outcome of the debate (and of $MBAF_N$)?

[18] For any pair of arguments $(a, b) \in A$, an agent can play a maximum of two moves in a debate: one move stating his opinion on the existence of the attack aRb, and one move stating his opinion on the existence of the support aSb. Since the agents are assumed truthful, and their private systems are assumed static, no agent can vote both positively and negatively on the same relation.

The following examples will illustrate the behavior of the protocols π_0 and π_1.

7.3.4 Illustrative examples and properties of π_0, π_1

We now provide some examples which illustrate important properties of the protocols π_0 and π_1. In what follows, we only focus on argumentation systems which do not contain cycles of attacks and supports. Therefore, we do not have any private systems, or GBs, containing arguments $a_1, a_2, \ldots, a_n$ such that: $a_1 R a_2$ (or $a_1 S a_2$), ..., $a_{n-1} R a_n$ (or $a_{n-1} S a_n$), and $a_n R a_1$ (or $a_n S a_1$). In acyclic systems the calculation of arguments' valuation is more straightforward, compared to systems with cycles.

Protocol π_0

We recall that π_0 is a two-sided category-based protocol, with a threshold arbitrarily fixed at 0. Recall that double arrows represent the attack relations, whereas simple arrows represent the support relations. In the following, we shall refer to the set of private agent systems as *the debate's initial configuration.*

Example 35 (cont.) *Let $N = \{a_1, a_2, a_3\}$ be a set of three agents whose systems are illustrated in Figure 7.11, together with their merged framework.*

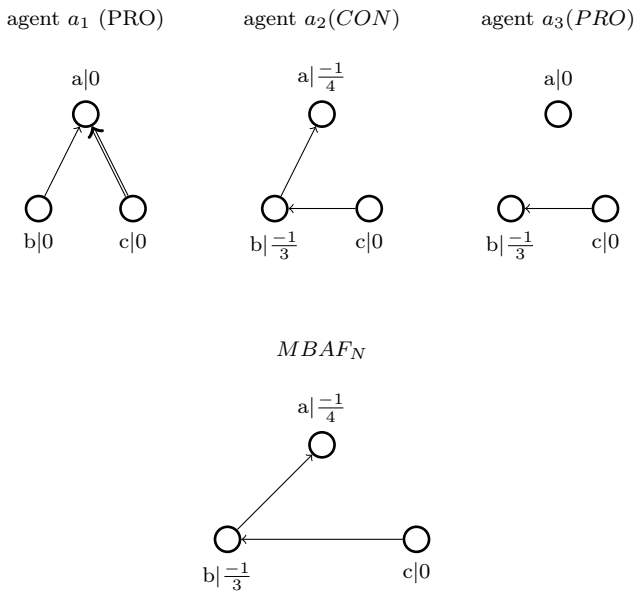


Figure 7.11: The BAFs of three agents and their corresponding merged framework.

Let argument a be the issue of the debate. As we can see from the valuations of argument a in the three systems, the agents are partitioned into the groups $PRO = \{a_1, a_3\}$, and $CON = \{a_2\}$. We remind that the debate starts with just the issue on the Gameboard, so $GB^0 = \langle\{a\}, \emptyset, \emptyset\rangle$. Let us see, in Figure 7.12, a sequence of moves which is allowed by the protocol π_0. Above each snapshot of the GB, the corresponding timestep is indicated, as well as the agent who has played the last move.

Let us comment the agents' moves in the previous run of the protocol. The debate starts with just argument a being on the GB and with CON having the token. At $t = 1$, a_2 plays for CON and attacks argument a with b. As the valuation of a is now negative, the token passes to PRO. One permitted and relevant move (by agent a_3) is to attack argument b with c, while another one

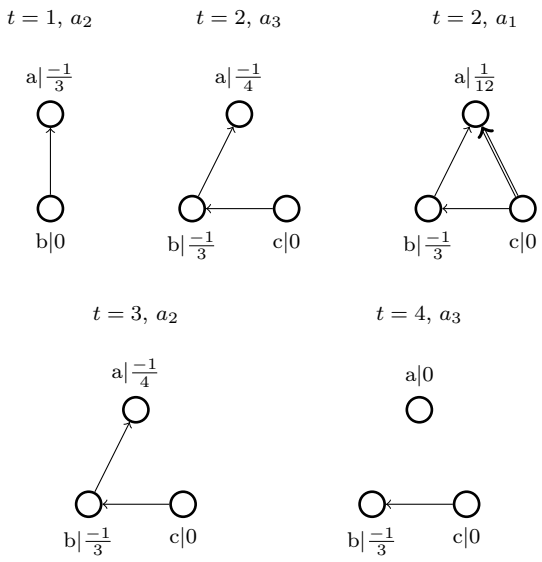


Figure 7.12: A run of the π_0 protocol where the PRO agents win.

(by agent a_1) is to support a with c. We assume that the central authority picks the first move (the addition of cRb). This move increases the valuation of a, but it still remains negative. Next, agent a_1 plays another move for PRO and adds the support relation cSa. The valuation of a becomes now positive, and the token is passed to CON. Agent a_2 removes the support relation cSa, therefore turning a's valuation negative. Finally, PRO agent a_3 removes the attack bRa, leading to $v(a) = 0$. The agents of the CON group (agent a_2) cannot propose any new moves. The final GB is stable, and the protocol stops with $v(a) = 0$, so PRO agents a_1, a_3 are the winners.

Let us now analyze some key points of the previous debate. The first interesting thing to observe here, is that CON agents cannot win a debate starting with the above initial configuration. This happens because PRO agents (thanks to agent a_3) can ensure that the attack bRa will stay off the GB. Therefore, the initial configuration leads to a deterministic outcome, as far as the winning group is concerned. As a side-note, the final, stable GB is not deterministically obtained in every possible debate, because the insertion of the attack cRb is not obligatory. Simply removing bRa will suffice for a PRO win. As a result, there are two possible stable GBs that can be obtained, starting with this initial configuration.

Let us now compare the debate stable GB with the $MBAF_N$. We see that, not only the above stable GB is different structurally from the $MBAF_N$, but also the outcome differs in these two systems. As we said above, we leave it to the reader to check that in this example, it is impossible to reach a GB where the sign of the issue's valuation is negative (like the sign of the issue in $MBAF_N$). But, why does this happen? The reason is that agent a_1 has no reason to insert the attack relation bRa, which appears in $MBAF_N$. As studied in a different context in [RL08], as well as in [BM11], this can be seen as a strategic manipulation by withholding an argument or an attack between arguments.

Finally, in this example the majority's opinion on the issue's valuation (positive sign), agrees with the debate's outcome, but disagrees with the merged outcome (negative sign).

Let us now provide another example of application of the π_0 protocol, starting with a different set of BAFs.

Example 37 *Let $N = \{a_1, a_2, a_3\}$ be a set of three agents whose systems are illustrated in Figure 7.13, together with their merged framework.*

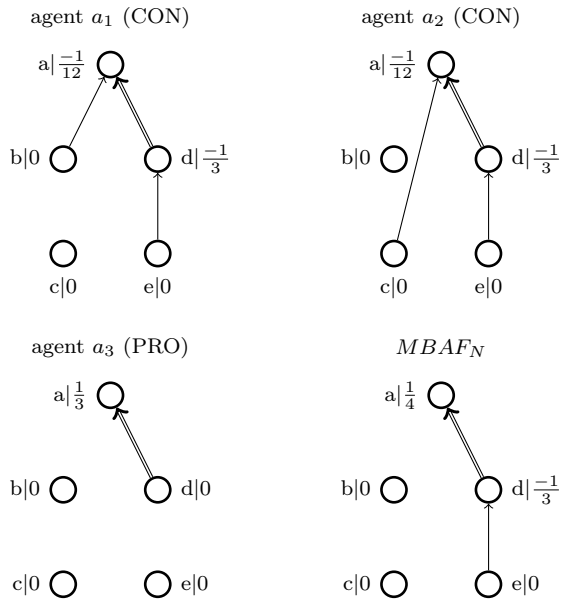


Figure 7.13: Three BAFs and their corresponding merged framework.

Let the issue of the dialogue be the argument a. As seen above, it holds that $PRO = \{a_3\}$, $CON = \{a_1, a_2\}$. A sequence of moves allowed by this protocol is illustrated in Figure 7.14:

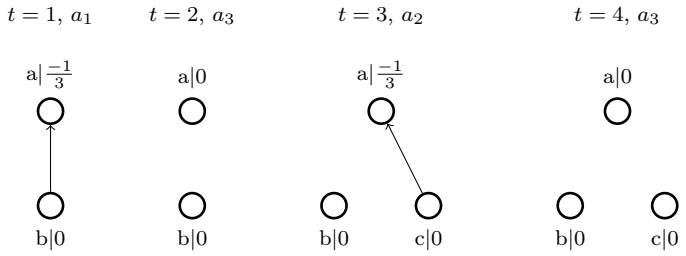


Figure 7.14: A sequence of moves allowed by π_0, where PRO agents win.

Let us comment the agents' moves. At $t = 1$, a_1 plays for CON and attacks the argument a with b. As the valuation of a is now negative, the token is given to PRO. Agent a_3 proposes two permitted (and relevant) moves to the central authority: The removal of bRa and the assertion of dSa. Assume that the authority picks the first one, and removes bRa. The valuation of a is turned positive, and the token is given to CON. Agent a_2 attacks a with c, then a_3 removes this attack (as above he has another permitted and relevant move, to assert dSa), and the token goes again to CON. CON agents have no relevant moves left, so the GB is stable, and the protocol stops with $v(a) = 0$. The PRO agents win the debate.

Let us now analyze some key points of the debate. Similarly to the previous example, this debate's result is deterministic, given the initial configuration. The PRO group (containing only agent a_3) will always win, by simply removing the attacks which the CON agents will assert (it can also assert dSa). Once again, the final GB is not deterministically obtained, since PRO agents could have added, at some point, the support dSa. In that case the CON agents would have attacked it with eRd, but they would have been unable to remove it.

Also, in contrast to the previous example, here the outcome of the debate coincides with the

merged outcome. Finally, the majority disagrees with the debate's (and with the merged system's) outcome.

In the two previous examples the debate's result was deterministic, given the initial set of private BAFs. However, the following example shows that this is not always the case. We will see that, in the general case, the debate's result depends on the moves chosen by the agents (and by the central authority).

Example 38 *Let $N = \{a_1, a_2, a_3\}$ be a set of three agents having the argumentation systems illustrated in Figure 7.15. Let the issue of the debate be the argument a. We have $PRO = \{a_1\}$, $CON = \{a_2, a_3\}$.*

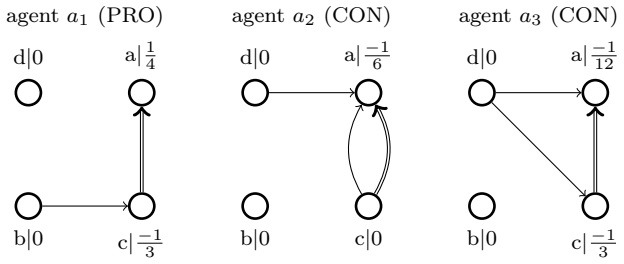


Figure 7.15: Three private BAFs. The outcome of this debate is non-deterministic.

The sequence of moves in Figure 7.16 allows agents in CON to win.

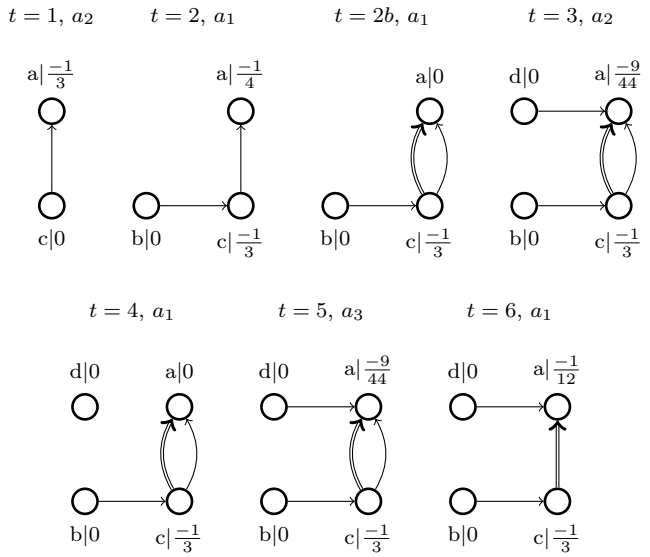


Figure 7.16: A first sequence of moves allowed by π_0, where the CON agents win.

At $t = 1$, a_2 plays for CON and attacks the argument a with c. Then, a_1 attacks the argument c with b. This move increases the valuation of a, but it remains negative. So, a_1 plays another move for PRO and adds a support relation between c and a. The valuation of a is now 0, so the token is given to CON. a_2 attacks a with d, a_1 removes this attack, but a_3 puts it back at $t = 5$. Finally, a_1 removes the attack between c and a, but it is not sufficient to turn the valuation of a non-negative. a_1 has no relevant moves left, so he cannot raise the valuation of a any more. The GB is stable, the protocol stops with $v(a) = \frac{-1}{12}$, and CON agents win.

Now let us see in Figure 7.17 a second sequence of moves, this time allowing PRO to win:

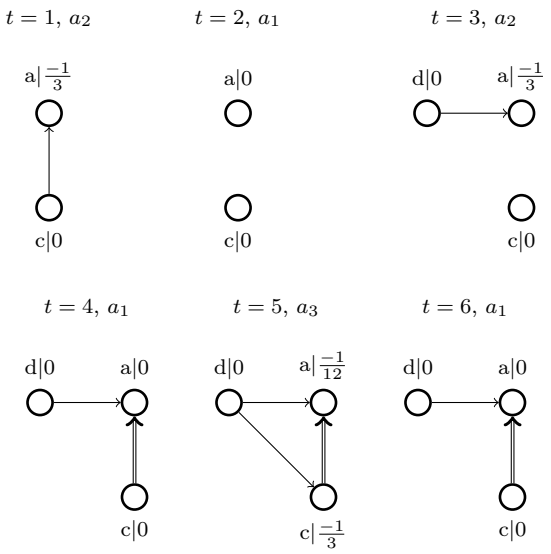


Figure 7.17: A second run of the π_0 protocol where the PRO agents win.

At $t = 1$, a_2 plays for CON and attacks the argument a with c, but a_1 removes this attack. At $t = 3$, a_2 attacks a with d. Then, a_1 adds a support relation between c and a, thus raising a's valuation to 0. Next, a_3 attacks c with d, but finally, a_1 removes this attack, again raising a's valuation to 0. Agents in CON have no relevant moves left, the GB is stable, the protocol stops with $v(a) = 0$, and the PRO agents win.

What is the main difference between the two previous debates? In the first one, PRO agents introduce at some point the attack *bRc*, in order to raise the issue's valuation. Unfortunately for them, this turns out to be a "mistake", as although they indeed raise the issue's valuation, this attack will later affect the support relation *cSa*, thus leading to a situation where the PRO group loses. In contrast, in the second move sequence, PRO agents do not introduce *bRc*, and they are able to win the debate.

Therefore, we see that the debate's result, given an initial configuration, is not always deterministic. The fact that the introduction of the attack *bRc* may raise the issue's valuation in some cases, while it may drop it in others, is closely related to the notion of **switch** presented in [BM11].

The last example shows that, roughly, a debate using π_0 may be non-deterministic if there exists an attack or support which can be used to both increase and decrease the issue's valuation, in different circumstances. In the last example, such a relation was the attack *bRc*.

Now, we state a necessary condition in order for a debate configuration to have a non-deterministic winner, when the π_0 protocol is used. Roughly, it says that there must be at least one move which is able to increase the issue's valuation (in at least one case), but also to decrease the issue's valuation (in at least one other case).

Property 22 *Let $N = \{a_1, \ldots, a_n\}$ be a set of agents, where $\forall i \in N$ $BAF_i = \{A, R_i, S_i\}$ is agent i's private bipolar argumentation framework. In order for the application of π_0 to have a non-deterministic winner, the following is a necessary condition: it is possible to construct two systems, $BAF_x = \langle A, R_x, S_x \rangle$ with $R_x \subseteq \bigcup_{i=1}^{n} R_i$, $S_x \subseteq \bigcup_{i=1}^{n} S_i$ and $BAF_y = \langle A, R_y, S_y \rangle$ with $R_y \subseteq \bigcup_{i=1}^{n} R_i$, $S_y \subseteq \bigcup_{i=1}^{n} S_i$, such that there exists a relation (attack or support) such that: (1)*

removing it from BAF_x increases the issue's valuation; and (2) removing it from BAF_y decreases the issue's valuation.

Proof 22 *Let us assume that the above condition is not verified. Then, we will prove that the application of π_0 has a deterministic winner.*

Given that the above condition is not verified, all relations (attacks and supports) which can be inserted in a debate (following π_0) can be partitioned in two sets: (1) Set_A which contains all the relations which, whenever inserted, increase the issue's valuation. (ii) Set_B which contains all the relations which, whenever inserted, decrease the issue's valuation. Therefore, following π_0, PRO will only insert relations from Set_A, while CON will only insert relations from Set_B.

Now, we define the bipolar framework GB^* which is formed of the following relations (and their corresponding arguments): All relations in Set_A, such that the number of PRO having them (in their private systems) is greater than the number of CON not *having them, and also all relations in Set_B, such that the number of CON having them is greater than the number of PRO* not *having them.*

We will prove that whenever the issue's valuation in GB^* is greater than (or equal to) zero, then the debate's winner (following π_0) is PRO; and whenever the issue's valuation in GB^* is less than zero, then the debate's winner (following π_0) is CON. In other words, we will prove that the debate's winner (following π_0) is always deterministic.

Let us assume the opposite: the issue's valuation in GB^* is greater than (or equal to) zero (resp. less than zero), and there exists a debate which is won by CON (resp. PRO). This would mean that, at the end of that debate, PRO (resp. CON) have no relevant moves left (so they lose). This is impossible: PRO (resp. CON) must obligatorily have at least one relevant move left to play, as otherwise it would have been impossible for the issue's valuation in GB^* to be greater than (or equal to) zero (resp. less than zero). Therefore, that debate which follows π_0 has not yet finished. As a result, such a debate may only finish with PRO (resp. CON) being the winner, so the winner of such a debate (following π_0) is always deterministic. We have thus proved that the stated condition is necessary, in order for a debate following π_0 to have a non-deterministic winner.

Let us recapitulate on category-based protocols, based on the previous property and on the illustrative examples of the π_0 protocol:

- The debate's outcome is non-deterministic, in the general case. Therefore, there is room for strategical considerations by the agents, who may try to identify "good" and "bad" moves.
- We have provided a necessary condition for the debate's winner to be non-deterministic, when π_0 is used.
- The debate's outcome does not coincide with the merged outcome, in the general case.
- The debate's outcome does not coincide with the majority's opinion, in the general case.

Now, let us turn our attention to the cluster-based π_1 protocol.

Protocol π_1

In this protocol, the agents want to turn the valuation of the issue on the GB, as close as possible to the valuation on their private system. We remind that each cluster contains a single agent, and that the agents can play anytime (there are no PRO and CON groups). The central authority picks and plays on the GB the move which amends the most the valuation of the issue.

Let us see a first illustrative example.

Example 35 (cont.) *Let us consider the set of debating agents in Figure 7.18, where previously the π_0 protocol has been applied. Now we shall see an application of the π_1 protocol instead.*

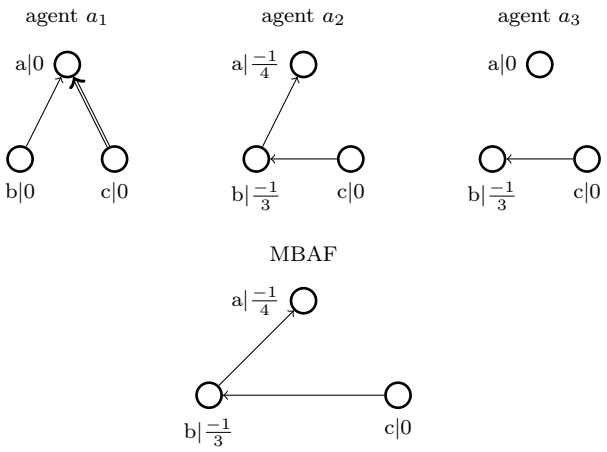


Figure 7.18: A previously seen set of three BAFs and their merged framework.

As before, the issue of the dialogue is the argument a. At the beginning of the debate, when only argument a is on the GB, agents a_1 and a_3 are perfectly happy with the situation, because their valuations coincide with the GB's valuation. The only agent who has an incentive to play a move is a_2. A sequence of moves allowed by this protocol is presented in Figure 7.19.

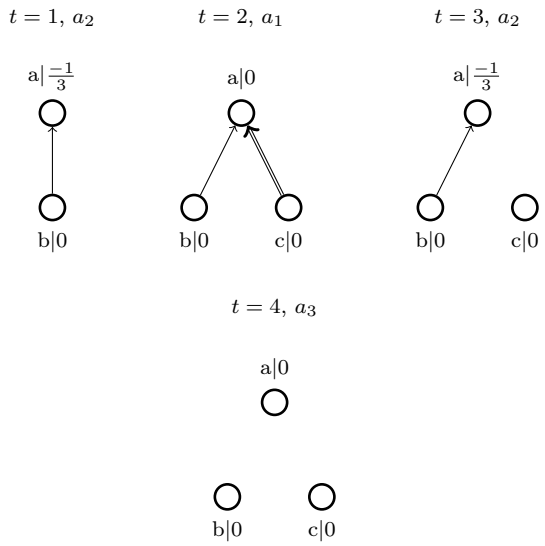


Figure 7.19: A run of the π_1 protocol where the agents a_1 and a_3 finally win.

We shall first analyze the moves played by the agents and the central authority. At $t = 1$, agents a_1 and a_3 are perfectly happy with the valuation of argument a being $v(a) = 0$. But, agent a_2 proposes to add the attack bRa, in order to turn the valuation of argument a on the GB closer to his valuation. The central authority executes this move. At $t = 2$, each agent proposes a move: (i) a_1 proposes to add the support relation cSa, to set the valuation of a back to 0, (ii) a_3 proposes to remove the attack between b and a for the same purpose, whereas (iii) a_2 proposes to attack b with c to obtain exactly his own argumentation framework. As the latter move has a smaller effect on the valuation of a, the central authority will not choose it, and it will choose one of the two former moves instead. Assume that the central authority chooses to add the support relation cSa,

proposed by a_1. At $t = 3$, agent a_2 is the only one who will propose a move, and his most relevant one is to remove the support cSa. Then, at $t = 4$, agent a_1 cannot propose any move (as he has already added the support cSa). The two remaining relevant moves are: for a_2 to add cRb, and for a_3 to remove bRa. As the latter move has the biggest effect on the valuation of a, it is picked by the central authority. No more moves can be proposed by the agents. The GB is stable and the protocol stops with $v(a) = 0$. Therefore, the agents a_1 and a_3 are the winners of the debate, as the distance of the issue's valuation in the GB and in their private systems is 0 (the smallest). Moreover, we can verify that all the possible move sequences lead to the same outcome, $v(a) = 0$, where agents a_1 and a_3 win. Therefore, the debate's result is deterministic, for this specific initial configuration.

What about the relation between the merged outcome and the debate's outcome? This example shows that, in the general case, they do not coincide (similarly to category-based protocols). The merged framework indicates a_2 as the "winner" (the valuation of the issue on the merged framework is identical to the valuation of a_2), but in the previous debate agents a_1, a_3 were the winners.

But not every initial set of BAFs leads to a deterministic outcome, as we shall now see. Also, interestingly, the necessary condition, stated in Property 22, in order to have non-deterministic winners, does not apply in the case of protocol π_1. In other words, we may have non-deterministic winners, even if there exists no move which can both increase and decrease the issue's valuation (in different cases). Let us provide an interesting example illustrating this.

Example 39 *Let $N = \{a_1, a_2, a_3\}$ be a set of three agents having the frameworks illustrated in Figure 7.20.*

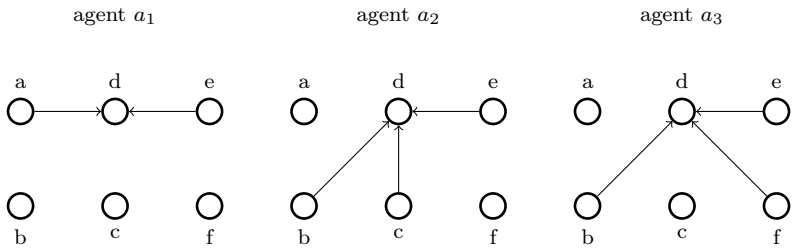


Figure 7.20: Three BAFs which can lead to different debate outcomes.

Let the issue of the debate be the argument d. In this example the exact valuations of the arguments will not play a significant role, therefore we shall omit them. We can see that the attacks in the agents' systems cannot "influence" the issue's valuation both positively and negatively, but only negatively. Regardless of this fact, and in contrast to what Property 22 states for protocol π_0, here the debate's winners are non-deterministic. Let us see why.

First, we provide in Figure 7.21 a sequence of moves where agent a_1 is the winner. Let us see what happened in that sequence of moves. Initially, the agents are not satisfied with the state of the debate, so they propose the insertion of some attacks against d. We assume that, in $t = 1$, agent a_1 proposed the addition of eRd and that this move was chosen by the central authority. In $t = 2$, the agents again propose the addition of attacks against d, as this way the issue's valuation gets closer to their private valuations. We assume that the central authority picks the addition of bRd, which was proposed by a_2. At this point, notice that agent a_1 is perfectly satisfied with the GB, while agents a_2 and a_3 are less satisfied with it, as their private valuation still differs. In $t = 3$, $t = 4$, $t = 5$ and $t = 6$, agent a_2 adds eRd, but a_1 responds by removing it. Then the same happens when agent a_3 adds fRd; again a_1 is able to remove it. At the end of timestep $t = 6$, no agent has a relevant move left, therefore the debate finishes, and the winner is agent a_1 whose private valuation is identical to the GB's valuation. Notice that agent a_1 has avoided to assert the attack aRd.

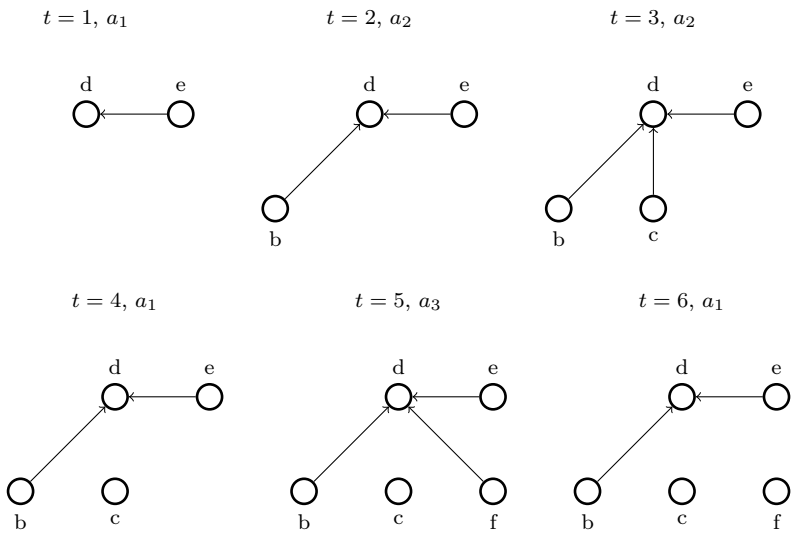


Figure 7.21: A first sequence of moves allowed by π_1, where a_1 wins.

Next, we provide in Figure 7.22 a second sequence of moves, where agents a_2 and a_3 win.

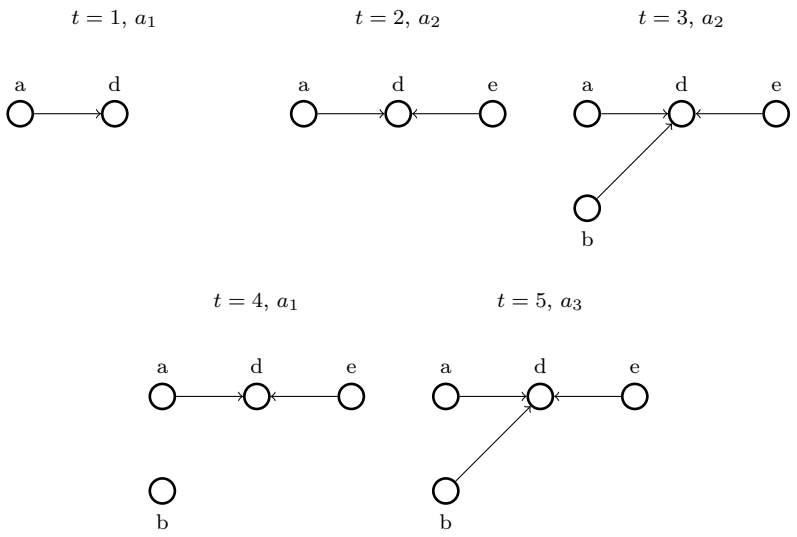


Figure 7.22: A second sequence of moves allowed by π_1, where a_1 now loses (a_2 and a_3 win).

The difference, compared to the first move sequence, is that at $t = 1$, agent a_1 asserted the attack aRd. This move backfired later in the debate, because agents a_2 and a_3 were able to assert (and keep on the GB) the attacks eRd and bRd. As a result, the debate ended with the issue's valuation coinciding with the valuation of agents a_2 and a_3, who are the winners.

It can be argued that the attack *aRd*, in the above example, is some kind of switch (in the sense of [BM11]) as its assertion by a_1 will either help him win the debate, or it will backfire. We shall not elaborate more on this, but this subject merits further study.

As a final note, it seems that the strict rule of selection of the most relevant move (used in π_1), instead of the rule of selection of a random move (used in π_0), usually leads to deterministic

outcomes. For example, notice that in the configuration of example 38, if the protocol π_0 had used the "most relevant move" selection rule, then that would have prevented PRO from adding the attack bRc. Instead PRO would have inserted the support cSa. This would have ensured that PRO wins the debate, always avoiding to play a move which would later "backfire".

Now we recapitulate on cluster-based protocols, based on the previous examples of application of π_1. The conclusions are very similar to those concerning category-based protocols.

- The debate's outcome is non-deterministic, in the general case, so there is room for strategical considerations by the agents.
- The debate's outcome does not coincide with the merged outcome, in the general case.
- The debate's outcome does not coincide with the majority's opinion, in the general case.

7.3.5 Conclusion on bipolar protocols

In argumentative debates among humans, gradual valuation of arguments can be used when just a few possible values for the issue (e.g. accepted and rejected) are considered insufficient. In these cases, the agents may want to know *how much* the issue is accepted. In our study, we have assumed that every agent has a private bipolar framework and that the Gameboard is also a bipolar framework.

We realized that, in order for a debate of this type to be focused, in the sense of [Pra05], we must define protocols where the agents influence the issue's valuation, with every move they make, in specific ways. But what are the main possibilities? We answered this question by defining and using attraction values which guide the agents during a debate. Also, we addressed the problem of group formation in such debates. We analyzed two possible motivations for group formation, leading to two types of protocols: category-based protocols and cluster-based protocols.

Category-based protocols are useful when the agents' opinions on the issue can be predefined. They assume some intervals where the agents want the issue's valuation to fall into. We defined a specific type of category based protocol, π_0, which uses only two categories (and partitions the agents into PRO and CON). We proved that debates following it always terminate, and that their result may depend on strategical considerations of the agents, two properties which may be desired in real debates. We also showed that the majority does not necessarily win the debate. Depending on the case, the debate's central authority can judge whether this is a desired property or an inconvenience.

Finally, we defined cluster-based protocols when no predefined categories exist. These protocols may be more useful in cases where the debating agents want a specific valuation for the issue: they are not satisfied by its valuation simply belonging in a predefined interval (category), but they continue changing it towards a specific value. We defined a specific type of cluster based protocol, π_1, and we showed that it satisfies similar properties to π_0.

7.4 Protocol Evaluation

Finally, we undertake a general evaluation of the three debate protocols defined in this work, and of the closely related protocol defined in [BM11]. We hold that such an evaluation can be valuable to the designer of a debating platform, since, if there are several available choices (protocols), then the designer's decision should be aided by a rigorous analysis of their strong and weak points. As explained in Section 6.4, the authors of [MPW02] and [AdSC13] propose a number of criteria for the evaluation of the quality of interactions. In the following we use these criteria in order to evaluate the debates proposed in this work, and in [BM11]. We start with the desiderata in [MPW02]:

- **Stated dialogue purpose:** The purpose of these debates is quite clear: a debate must decide which is the "correct" acceptability status of the issue. In other words, which group's viewpoint over the acceptability of the issue will be verified.

- **Diversity of individual purposes:** First, in the protocol in [BM11] and in our protocol based on target sets, there are two groups of agents, PRO and CON. PRO (resp. CON) agents want to win the debate by making the issue, at the end of the debate, accepted (resp. rejected). Second, in our protocols using numerical argument valuation, there are two cases: in category-based protocols, the agents try to make the issue's valuation end up in "their" category, while in cluster-based protocols, the agents try to make the issue's valuation end up (as close as possible) to "their" cluster. Therefore, for given protocol type, the individual purposes of the agents are rather similar. The fact that all the protocols focus on a single issue restricts the diversity of individual purposes.

- **Inclusiveness:** All the debates are open to properly qualified agents who are willing to participate. In our work, we proposed a way to take into consideration the agents' expertise in relevant topics, and we underline that these debates are, *a priori*, designed to facilitate the participation of numerous agents.

- **Transparency:** For every debate type, the debate's rules as well as the structure of the Gameboard are transparent to all the agents.

- **Fairness:** No debate protocol gives, to the best of our understanding, the advantage to some particular agents. An exception is our protocol focusing on target sets, where big experts in the topics under discussion have an edge over non-experts. Apart from the protocols, it remains to be shown whether there exist other elements (e.g. the structure of the agents' private systems) which sometimes increase the chances of winning of some agents (e.g. PRO or CON).

- **Clarity of argumentation theory:** In all the debates, the argument acceptability semantics, as well as the dialogue rules have been clearly defined.

- **Separation of syntax and semantics:** This is also achieved through the use of abstract argumentation and different types of well-established acceptability semantics.

- **Rule-consistency:** The debate protocols ensure that there can be no deadlocks and termination issues, as they prohibit move repetitions.

- **Encouragement of resolution:** Resolution of debates is facilitated by the fact that all the protocols allow only relevant moves. The definition of relevant move differs for one protocol type to another.

- **Discouragement of disruption:** Disruption is not a problem in these debates. Firstly, the agents cannot retract or repeat moves in order to create problems in the dialogue procedure. Secondly, the protocols allow only relevant moves, so it is difficult for agents to make a debate lose its focus on the issue.

- **Enablement of self-transformation:** This is a criterion that these debates do not satisfy. Dynamic private systems, which can evolve during the debate, are not considered. This is left as an important topic of future research.

- **System simplicity:** The available locutions (argument and attack assertions, votes on attacks) as well as the debates' rules are, arguably, quite simple.

- **Computational simplicity:** Deciding argument acceptability is computationally efficient, both in the case of numerical argument valuation in presence of attacks and supports (we remind that we do not consider cases where the graphs contain cycles), as well as in the case of grounded semantics. As far as the protocol focusing on target sets in concerned, more work must be done on identifying the complexity of target set computation, as it can lead to an overhead in the computations made by the central authority of the debate.

Next, we evaluate our debates and the debates in [BM11], based on the criteria proposed in [AdSC13] for persuasion dialogues.

1. **Finiteness:**
   It has been proven that these debates' protocols ensure termination.

2. **Non-determinism:**
   We have provided examples, for all these debates, showing that in the general case, the outcome is non-deterministic. Therefore, the agents can act strategically, in order to win a debate. On the other hand, there always exist some initial configurations (sets of agents with their private systems and expertise) which lead to deterministic outcomes.

3. **Consistency:**
   In debates using extension-based acceptability semantics, the result is consistent, in the sense that every extension is conflict-free. In debates using numerical argument valuation, it is possible to have some kind of inconsistency: e.g. arguments a and b both have positive valuations, and bRa. This is indeed possible, since another argument c could be supporting a, being the reason why a has a positive valuation.

4. **Natural-attacks-allowance:**
   For all these debates, this criterion is not satisfied. The agents are able to assert attacks, without stating whether they are, for example, rebuts or undercuts, because we abstract from the arguments' content. Similarly, agents can vote negatively on attacks (and thus remove them), without offering any explanation for their vote. If we extend our work by using structured arguments, then we should think of possible ways to satisfy this criterion.

5. **Dissimulation:**
   In these debates dissimulation is satisfied, as the agents are allowed to hide arguments, as well as their opinions on attacks, in order to achieve their goals. Such types of strategic moves are essential, as they can change the outcome of a debate.

6. **Non-triviality:**
   We consider that non-triviality is satisfied, in the sense that every relevant move (whether it is an argument and attack assertion, or a vote on some attack) provides some non-trivial information to the debate.

Finally, in [TG10] the authors propose three criteria for the evaluation of an argumentative game's complexity. According to these criteria, our debates could be characterized as follows:

- **Agent type: Indicator**
  These debates have a single issue, so every agent focuses on the acceptability status of the same argument. This is the simplest case with respect to this criterion.

- **Awareness: No awareness**
  The agents are assumed to not know anything about the other agents' beliefs, preferences or expertise. Again, this is the simplest case with respect to this criterion.

- **Game protocol: Dialectical**
  In all these debates, there are (more or less) elaborate definitions of turn-taking, and the debates consist of multiple turns. Unlike the above, this is the most complicated choice with respect to this criterion.

In this chapter we have defined, studied and evaluated a diverse set of debate protocols: one protocol which is based on extension-based argument evaluation (and makes extensive use of target sets), and a couple of protocols which are based on numerical argument valuation. All these protocols are centered around a single (same for all agents) issue. They ensure that the participants are focused, and that their contributions are always relevant to that issue. Additionally, we have

defined a number of agent strategies, and we have proposed some strategy evaluation criteria, so that agents can decide which strategy to adopt. We believe that our analysis of debate protocols and strategies can motivate debate administrators and users to adopt them, and to define several variations, for use in real-life debate settings.

Chapter 8

Conclusion

Multi-agent argumentative debates face several issues. For example, they may lead to unclear conclusions, they may lack focus, and their participants may not be well coordinated by the debate's central authority. Also the profiles of the participants are not usually taken into account. The contribution of this thesis is multiple:

- We have proposed representations of debates which can be used in practical settings (e.g. in debate platforms) where the agents' expertise plays an important role.
- We have studied the controversy of a debate's conclusions, in order to provide decision support to the mediator of a debate.
- We have studied specific ways in which the agents can pursue their goals in debates, by modifying the structure of an argumentation system.
- We have defined different debate protocols whose properties were analyzed, and also different agent strategies which were experimentally evaluated.

We have started from the observation that, in real debates, some users' opinions usually count more than some others'. For example, the opinion of users with verified expertise on specific topics is considered more important than the opinion of non-experts. But this depends on the topic under discussion, as generally users do not have the same expertise in every topic.

In order to **model a debate among experts**, we consider a possible set of topics characterizing the agents' expertise. Arguments put forward in the discussion have been also characterized by those topics. In order to solve the problem of disagreement among agents on the attacks between arguments, we have proposed a method which aggregates the agents' votes based on our proposed model of expertise. The debate is represented in the form of a single argumentation system (called Gameboard) whose attacks have evaluation vectors.

Next, motivated by the idea that sometimes a debate's conclusions do not convince its users, we analyzed the problem of **controversial debates**. This analysis can be used for decision support by a debate's mediator who wants to increase the users' confidence on the debate's conclusions. We argued that the conclusions of a debate cannot always be taken for granted and we analysed the controversy of a Gameboard based on three criteria:

- *The stability of the attacks:* The stability of an attack depends both on its weight, as well as on its max-weight. We partitioned the attacks of a Gameboard into three sets, depending on the certainty about their status. *Beyond any doubt* attacks certainly belong to the system (or they certainly do not), *strong* attacks cannot change signs with a single additional vote, while weak attacks may change sign after a single additional vote.

- *The persistence of the arguments:* This notion is directly connected to the notion of attack stability. The persistent arguments are those whose acceptability status would remain the same, no matter the changes on the Gameboard's weak attacks.
- *The decidability of the arguments:* We consider that the more undecided arguments there are on the Gameboard, the less clear the debate is.

Afterwards, we studied how an additional expert can come to the debate, cast his votes on the Gameboard's attacks, and influence its stability. This may happen in two ways: (1) by increasing and decreasing the absolute values of the attacks' weights, he may change the attacks' stability, and the arguments' persistence; (2) by changing the signs of the attacks' weights, he may change the acceptability status of the arguments (and turn decided arguments into undecided, and vice-versa).

Given that our task is to provide decision support to a debate's mediator, we asked the following question: how can we choose an agent (whose expertise is known, but not his opinions on the attacks) who will stabilize a given debate? In order to provide an answer we had to define several *dominance relations* over experts. Roughly, an expert dominates another one if he is able to bring about systems which are more stable than the systems the other can bring about. We have defined several types of dominance relations, based on the three criteria of controversy, and also different versions of these dominance relations (e.g. optimistic dominance compares the "best scenarios", while pessimistic dominance compares the "worst scenarios"). Then, we have provided a number of theoretical results which highlight the relations between the different dominance relations. These results may help a mediator choose the best expert to stabilize a given debate.

Then, we focused on **how the agents can modify a Gameboard** in order to achieve their goals in a debate. We defined an argumentation system with modifiable attacks (ASMA), as a system containing some "fixed" and some "debated" attacks. Fixed attacks cannot be put into question, while debated attacks can be added or removed from the system. Given an argumentative goal (e.g. the acceptance under some semantics of an argument), we studied how the initial system may be structurally modified in order for that goal to be achieved.

A type of change which we found particularly interesting, was *minimal* change achieving a goal. The reason is that relying on minimal change achieving a goal, is intuitively the simplest and most economical way to achieve that goal. As far as the study of (minimal) change is concerned, our contribution was the following: we first proved some properties of (minimal) change achieving an argumentative goal. Then, we defined and implemented in Maude a term rewriting procedure which, roughly, does the following: it takes as input an encoded initial ASMA, and it computes all the possible minimal changes (called *target sets*) for some particular types of goals. Moreover, we studied the meta-dynamics of ASMAs, focusing on the evolution of target sets. More specifically, given an ASMA and a goal, we studied how the initial target sets evolve, after different types of changes on the system: (1) changes which are not part of any target set, (2) changes which are part of some target set. These theoretical results have been used in order to show that focusing on target sets is preferable, in order to achieve a goal.

One of the essential elements of a debate is its protocol. We have defined multilateral argumentative debate protocols, where the agents argue over the acceptability of an issue. First, inspired from our work on target sets, we considered debates where the agents focus on minimal change achieving a goal. In these protocols, extension-based argument acceptability was used (specifically grounded semantics). Second, we considered bipolar systems (with both attack and support relations), using numerical valuation of arguments. These different types of argumentation systems naturally lead to quite different protocols, which were analyzed and then evaluated, based on a number of criteria proposed in the literature. Let us remind the types of protocols we have defined:

- **A method for quick convergence to a persistent issue:**
  We have proposed a method to search for additional opinions when the issue of a debate is

non-persistent. It is based on binary decision diagrams, and it finds the ordering of questions on (weak) attacks which leads, as soon as possible, to the issue becoming persistent.

- **Protocols for bipolar systems:**
  We have defined two types of protocols: *category-based* protocols (where agent groups can be defined a priori, based on exogenously given information) and *cluster-based* protocols (where agent groups depend on the agents' valuations of the issue). Following [Pra05], we wanted our debates to be both flexible (thus allowing a number of possible moves, at every point, to the participants), but also coherent (focusing on the issue). Thus, we introduced different definitions of *move relevance*, for the two types of protocols. Next, we studied whether the protocols satisfy some properties. A very important property is termination and it is always ensured in our setting. Another important property is non-determinism of the debate's outcome. We showed that our debates are non-deterministic, in the general case, so there is room for strategic considerations by the agents. In some illustrative examples, we identified some potentially "bad" and "good" moves which can alter the outcome of a debate. Moreover, we studied whether the result of a debate always agrees with the majority's opinion, or with the agents' unanimous opinion.

- **Protocol based on target sets:**
  We also defined a debate protocol based on target sets. The motivating idea was that, at any given point of the debate, the agents who are currently losing would prefer to: (1) change the outcome as quickly as possible, (2) do this in a way that their task does not become more difficult. In this respect, focusing on target sets is a reasonable choice for the agents, as we have formally proved that playing *outside* target sets, turns them bigger. We also defined a number of agent strategies, of increasing complexity, as well as some heuristics for move selection. The strategies and the heuristics where evaluated in an experimental setting, where more than $72,000$ debates were generated and run. We proposed a way to randomly generate initial debate configurations, and then we tested how every strategy (and heuristics) fared against the others, based on four evaluation criteria: (i) probability to win the debate, (ii) debate length, (iii) rationality of the outcome (coincidence with the outcome of the merged system), and (iv) agent happiness (coincidence with the majority's opinion).

We believe that our analysis of abstract argumentation dynamics, and of potential ways to coordinate debates where participants have different expertise and beliefs, can be of practical interest to both a debate's central authority (designer, administrator, or mediator), as well as to the users of a debate. Firstly, a central authority can use our work in order to better coordinate a debate (by choosing from a wide range of protocols), and to draw its conclusions (as well as to evaluate their stability). Secondly, simple users can profit from a quite straightforward way of debating, from ways to reason about the potential evolutions of a debate, and from a set of strategies which can help them achieve their personal goals in a debate.

8.1 Directions of Future Research

There are many possible ways in which this work can be extended and applied. First, we will consider some amendments and extensions. Second, we will identify two main directions in which we believe this work could be proven useful.

8.1.1 Amendments and extensions

We now provide some possible amendments and extensions of our work, which we believe have both a theoretical and a practical interest.

Extensions on the aggregation of expert opinions

As far as potential extensions on the aggregation of expert opinions are concerned, we could do the following:

- We could propose alternative methods of aggregating the experts' opinions. In our work, a simple sum has been used in order to aggregate the experts' opinions on attacks. But, is a simple sum always sufficient to aggregate the available information, or sometimes more elaborate choices are preferable? For example, how should the aggregation be done when experts in one topic agree among themselves, while at the same time, they disagree with experts in another topic? In that case, is a single aggregated system sufficient? Or perhaps more than one aggregations should be considered? In that case, how could we define the debate's outcome?
- We could study some complexity issues regarding the computation of a debate's controversy (e.g. the computation of the persistent arguments), as well as the computation of the dominance relations over experts, based on their ability to stabilize a debate.
- We could study how humans evaluate the controversy of debates, and compare with the results of our approach. Also, we could use our approach in real debates, and evaluate how helpful such a support is to the users.

Extensions on minimal change

As far as potential extensions on minimal change are concerned, we could do the following:

- We could consider multi-issue scenarios, focusing on the status of many arguments, at the same time, and define a rewriting procedure which is capable of computing target sets for multi-issue scenarios.
- Given the representation of a debate in the form of Gameboard, we could provide visual help to the users who want to achieve their argumentative goals. For example, we could provide them a graphical explanation of why a particular change would help them achieve their goal. Finally, we could evaluate how helpful this visualization is to human users.
- We could study the nature of target sets in different systems (e.g. systems retrieved from on-line discussions, randomly generated systems), and compare the target sets' numbers, their sizes, and how much they "overlap".

Extensions on debate protocols and strategies

Finally, as far as potential extensions on debate protocols and strategies are concerned, we could do the following:

- In a bipolar setting, we could define variations of the π_0 and π_1 protocols, by changing: the turn-taking, the definition of relevant move, the winning condition. Then, we could study the properties of these protocols, and search for dominant strategies.
- In a setting where acceptability is extension-based (focusing or not on target sets), we could consider more acceptability semantics and perform new experiments. Also, we could study how different types of private systems (e.g. graphs with cycles, trees with small or big branching factor) affect the debate's outcome.
- We could consider the possibility that agents have different sets of arguments. For example, some arguments could be known to some agents, but unknown to others. Of course, in that case some choices must be made, for example an agent who does not know about argument a, how will he respond to another agent who puts forward the attack (a, b)?

- We could study the case of agents having dynamic private argumentation systems. In that case, an agent's private system could change several times during a debate, and as a result, his group (if there are any groups defined) could also change several times.
- We could consider the possibility of lying agents. In our examples and experiments we have assumed that all the agents are truthful, in the sense that they never vote on an attack in a way which contradicts their beliefs. This is naturally a restricting assumption, as in many settings agents lie, if this can be helpful.
- We could define strategies which do not only take into account the situation on the Gameboard, but also a model of the opponents' beliefs, and they try to predict the opponents' next moves.
- We could study how our debate protocols could be modified and applied, in a setting where the exchanged arguments have specific structure and content.

8.1.2 Contribution to some broader research directions

Finally, we indicate a couple of research directions in which we believe that our work can open new perspectives. We focus on argumentation among human agents, but many ideas can be applied in settings with software agents, or in hybrid systems, containing both types of agents.

Mediating a debate among human agents

The way humans argue over the Internet faces various problems, as we have already explained. If we focus on the arguments' content, humans rarely express themselves in a formal way, but for their convenience (or motivated by strategical considerations) they employ enthymemes [Wal08], whose excessive use can easily lead to confusion. Another major issue is the repetition of the same information (in the form of arguments) by several agents, but in slightly different ways. We argue that a debate platform should coordinate the gradual refinement of the Gameboard, by demanding for example explanations over confusing enthymemes. Also, it should be able to identify and remove redundant information (different variations of the same argument). If the agents disagree on the representation of some crucial information (argument), then a solution could be to split the Gameboard in two or more Gameboards. The platform could provide support to the agents throughout the debate, by indicating the major points of disagreement, and by evaluating the issue (conceptually more difficult if the Gameboard is split). Moreover, a platform could classify the agents with respect to their beliefs, or their ways of arguing. Also, it could provide feedback to the agents indicating, for example, how contradicting their opinions are, and proposing some opinions they might want to reconsider, in order to be in-line with their "profile", or simply in order not to self-contradict. Moreover, such a platform may consider the use of more complex relations over arguments like those proposed in [BW10, CMDM06], as humans do not always use the standard binary attack and support relations. Finally, it would be useful to analyse available debates over the Internet, in order to understand the behaviour of real agents, and not just work on conveniently constructed examples. The goal would be to better understand how humans debate, and to provide a mediation which improves these debates' quality.

Enriching agent-based social simulation

In the field of social-based agent simulation [GC95], real agents and their social behaviours are modelled. Then, simulations of agent interactions are performed, and the agents' behaviour, both individually, but also collectively, is observed. Finally, the reasons leading to the observed behaviour are studied. If these reasons are successfully identified, then good predictions about the real-world social system can be made in the future. Given that humen are reasoning entities, who often use arguments, it may be helpful in many cases to consider that every agent is equipped with a private argumentation system, or with a knowledge base from which he is able to construct

arguments. Modeling the agents' reasoning mechanisms may be an important addition to this type of simulations, because it can lead to the understanding of how collective opinions are formed in a social context.

Bibliography

[ABC+06] Leila Amgoud, Lianne Bodenstaff, Martin Caminada, Peter McBurney, Simon Parsons, Henry Prakken, Jelle van Veenen, and G. A. W. Vreeswijk. Final review and report on formal argumentation system. deliverable d2. 6. Technical report, ASPIC IST-FP6-002307, 2006. 25

[AC02] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002. 25, 33

[ACLSL08] Leila Amgoud, Claudette Cayrol, Marie-Christine Lagasquie-Schiex, and Pierre Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23:1062–1093, 2008. 11, 17, 31, 32, 144

[ADM08] Leila Amgoud, Yannis Dimopoulos, and Pavlos Moraitis. Making decisions through preference-based argumentation. In *Proceedings of the Eleventh International Conference on the Principles of Knowledge Representation and Reasoning (KR'08)*, pages 113–123, 2008. 11, 17, 26, 34

[AdSC13] Leila Amgoud and Florence Dupin de Saint-Cyr. An axiomatic approach for persuasion dialogs. In *Proceedings of the Twenty-fifth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'13)*, pages 618–625, 2013. 105, 121, 122, 161, 163

[AGM85] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(02):510–530, 1985. 58

[AM02] Leila Amgoud and Nicolas Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proceedings of the Ninth International Workshop on Non-Monotonic Reasoning (NMR'02)*, pages 399–407, 2002. 12, 119

[AMP00] Leila Amgoud, Nicolas Maudet, and Simon Parsons. Modelling dialogues using argumentation. In *Proceedings of the Fourth International Conference on MultiAgent Systems (AAMAS'00)*, pages 31–38, 2000. 25

[Aus75] John L. Austin. *How to do things with words*, volume 1955. Oxford University Press, 1975. 104

[Bau12] Ringo Baumann. What does it take to enforce an argument? minimal change in abstract argumentation. In *Proceedings of the Nineteenth European Conference on Artificial Intelligence (ECAI'12)*, pages 127–132, 2012. 58

[BB10] Ringo Baumann and Gerhard Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In *Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10)*, pages 75–86, 2010. 58

[BC03] Trevor Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003. 25, 26, 34, 127

[BCdSCLS11] Pierre Bisquert, Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Change in argumentation systems: exploring the interest of removing an argument. In *Scalable Uncertainty Management*, pages 275–288. Springer, 2011. 54

[BCdSCLS12] Pierre Bisquert, Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Characterizing change in argumentation by using duality between addition and removal. Technical report, Tech. rep., IRIT, UPS, Toulouse, France, 2012. 54

[BCG11] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(04):365–410, 2011. 19

[BCPR12] Richard Booth, Martin Caminada, Mikołaj Podlaszewski, and Iyad Rahwan. Quantifying disagreement in argument-based reasoning. In *Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS'12)*, pages 493–500, 2012. 51, 52

[BG09] Pietro Baroni and Massimiliano Giacomin. Semantics of abstract argument systems. In *Argumentation in Artificial Intelligence*, pages 25–44. Springer, 2009. 18, 47

[BGK+14] Richard Booth, Dov Gabbay, Souhila Kaci, Tjitze Rienstra, and Leendert van der Torre. Abduction and dialogical proof in argumentation and logic programming. In *Proceedings of the Fifteenth International Workshop on Non-Monotonic Reasoning (NMR'14)*, 2014. 59

[BGP+11] Guido Boella, Dov Gabbay, Alan Perotti, Leendert van der Torre, and Serena Villata. Conditional labelling for abstract argumentation. In *Proceedings of the Workshop on Theory and Applications of Formal Argumentation (TAFA'11)*, pages 232–248, 2011. 55, 56, 57, 58, 61, 63

[BGvdTV10] Guido Boella, Dov Gabbay, Leendert van der Torre, and Serena Villata. Support in abstract argumentation. In *Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10)*, volume 216, pages 111–122, 2010. 32

[BGW05] Howard Barringer, Dov Gabbay, and John Woods. Temporal dynamics of support and attack networks: From argumentation to zoology. In *Mechanizing Mathematical Reasoning*, pages 59–98. Springer, 2005. 18, 32

[BH01] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1):203–235, 2001. 49

[BKvdT09a] Guido Boella, Souhila Kaci, and Leendert van der Torre. Dynamics in argumentation with single extensions: Abstraction principles and the grounded extension. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09)*, pages 107–118, 2009. 54, 55

[BKvdT09b] Guido Boella, Souhila Kaci, and Leendert van der Torre. Dynamics in argumentation with single extensions: Attack refinement and the grounded extension. In *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'09)*, pages 1213–1214, 2009. 54

[BLZ04] Emilia Bellucci, Arno R. Lodder, and John Zeleznikow. Integrating artificial intelligence, argumentation and game theory to develop an online dispute resolution environment. In *Proceedings of the Sixteenth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'04)*, pages 749–754, 2004. 106

[BM11] Elise Bonzon and Nicolas Maudet. On the outcomes of multiparty persuasion. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11)*, pages 47–54, 2011. 12, 38, 39, 40, 104, 111, 112, 113, 114, 115, 117, 125, 135, 143, 147, 150, 153, 156, 160, 161, 162, 163

[BW10] Gerhard Brewka and Stefan Woltran. Abstract dialectical frameworks. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR'10)*, pages 102–111, 2010. 17, 33, 56, 169

[CA09] Dan Cartwright and Katie Atkinson. Using computational argumentation to support e-participation. *Intelligent Systems, IEEE*, 24(5):42–52, 2009. 127

[Cam06] Martin Caminada. On the issue of reinstatement in argumentation. In *Proceedings of the Tenth European Conference on Logics in Artificial Intelligence (JELIA'06)*, pages 111–123. Springer, 2006. 18, 21, 36, 50, 88, 91, 108

[CDE+99] Manuel Clavel, Fransisco Durán, Steven Eker, Patrick Lincoln, Narciso Martí-Oliet, José Meseguer, and Jose F. Quesada. The maude system. In *Proceedings of the Tenth International Conference on Rewriting Techniques and Applications (RTA'99)*, pages 240–243, 1999. 72

[CdSCLS08] Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Revision of an argumentation system. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, pages 124–134, 2008. 52

[CdSCLS10] Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38(1):49–84, 2010. 52, 53, 54

[CLS05a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, pages 366–377, 2005. 31, 32

[CLS05b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research (JAIR)*, 23:245–297, 2005. 18, 23, 24, 32

[CLS05c] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, pages 378–389, 2005. 17, 32

[CLS11] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Weighted argumentation systems: A tool for merging argumentation systems. In *Proceedings of the Twenty-third IEEE International Conference on Tools with Artificial Intelligence (ICTAI'11)*, pages 629–632, 2011. 35

[CMDK+07] Sylvie Coste-Marquis, Caroline Devred, Sébastien Konieczny, Marie-Christine Lagasquie-Schiex, and Pierre Marquis. On the merging of dung's argumentation systems. *Artificial Intelligence*, 171:740–753, 2007. 12, 34, 35, 37, 126, 132, 145, 146

[CMDM06] Sylvie Coste-Marquis, Caroline Devred, and Pierre Marquis. Constrained argumentation frameworks. In *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR'06)*, pages 112–122, 2006. 17, 33, 169

[CMKMM14] Sylvie Coste-Marquis, Sébastien Konieczny, Jean-Guy Mailly, and Pierre Marquis. On the revision of argumentation systems: Minimal change of arguments statuses. In *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'14)*, volume 14, pages 52–61, 2014. 58

[CMKMO12] Sylvie Coste-Marquis, Sébastien Konieczny, Pierre Marquis, and Mohand Akli Ouali. Weighted attacks in argumentation frameworks. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'12)*, pages 593–597, 2012. 31

[CP11] Martin Caminada and Gabriella Pigozzi. On judgment aggregation in abstract argumentation. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 22(1):64–102, 2011. 12, 36

[CPP11] Martin Caminada, Gabriella Pigozzi, and Mikolaj Podlaszewski. Manipulation in group argument evaluation. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11) - Volume 3*, pages 1127–1128. International Foundation for Autonomous Agents and Multiagent Systems, 2011. 36

[CV12] Elena Cabrio and Serena Villata. Combining textual entailment and argumentation theory for supporting online debates interactions. In *Proceedings of the Fiftieth Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 208–212. Association for Computational Linguistics, 2012. 8

[Dav11] Martin Davies. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher education*, 62(3):279–301, 2011. 8

[DDLN10] Caroline Devred, Sylvie Doutre, Claire Lefèvre, and Pascal Nicolas. Dialectical proofs for constrained argumentation. In *Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10)*, pages 159–170, 2010. 50

[DGPS01] Didier Dubois, Michel Grabish, Henri Prade, and Philippe Smets. Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: The assessment of the value of a candidate. *International Journal of Intelligent Systems*, 16(11):1245–1272, 2001. 41

[DHM+11] Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, 2011. 11, 25, 28, 31, 34

[DHP14] Sylvie Doutre, Andreas Herzig, and Laurent Perrussel. A dynamic logic framework for abstract argumentation. In *Proceedings of the Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning (KR'14)*, 2014. 58

[DMA09] Yannis Dimopoulos, Pavlos Moraitis, and Leila Amgoud. Extending argumentation to make good decisions. In *Algorithmic Decision Theory*, pages 225–236. Springer, 2009. 26, 27

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995. 11, 18, 19, 20, 21, 62, 63, 143, 144

[DV04] Frank Dignum and Gerard Vreeswijk. Towards a testbed for multi-party dialogues. In *Advances in Agent Communication*, pages 212–230. Springer, 2004. 7

[EML13] Sinan Eğilmez, Joao Martins, and Joao Leite. Extending social abstract argumentation with votes on attacks. In *Proceedings of the Second Workshop on Theory and Applications of Formal Argumentation (TAFA'13)*, pages 16–31, 2013. 28, 30, 31, 36

[ET12] Valentinos Evripidou and Francesca Toni. Argumentation and voting for an intelligent user empowering business directory on the web. In *Proceedings of the Sixth International Conference on Web Reasoning and Rule Systems (RR'12)*. Springer, 2012. 32, 33

[ET14] Valentinos Evripidou and Francesca Toni. Quaestio-it. com: a social intelligent debating platform. *Journal of Decision Systems*, 23(3):333–349, 2014. 8

[FFMM94] Tim Finin, Richard Fritzson, Don McKay, and Robin McEntire. Kqml as an agent communication language. In *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*, pages 456–463, 1994. 103

[FIP08] TCC FIPA. Fipa communicative act library specification. *Foundation for Intelligent Physical Agents*, 2008. 103

[Gab09] Dov Gabbay. Fibring argumentation frames. *Studia Logica*, 93(2-3):231–295, 2009. 51

[GC95] Nigel Gilbert and Rosaria Conte. *Artificial Societies: the computer simulation of social life*. Taylor & Francis, Inc., 1995. 169

[GR12] Dov Gabbay and Odinaldo Rodrigues. A numerical approach to the merging of argumentation networks. In *Proceedings of the Thirteenth Workshop on Computational Logic in Multi-Agent Systems (CLIMA'12)*, pages 195–212. Springer, 2012. 35, 36

[GS04] Alejandro J. Garcia and Guillermo R. Simari. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(1-2):95–138, 2004. 25

[HSM+13] Christos Hadjinikolis, Yiannis Siantos, Sanjay Modgil, Elizabeth Black, and Peter McBurney. Opponent modelling in persuasion dialogues. In *Proceedings of the Twenty-third international Joint Conference on Artificial Intelligence (IJCAI'13)*, pages 164–170, 2013. 119, 121

[Hun07] Anthony Hunter. Real arguments are approximate arguments. In *Proceedings of the Twenty-second Conference on Artificial Intelligence (AAAI'07)*, volume 7, pages 66–71, 2007. 34

[Hun12] Anthony Hunter. Some foundations for probabilistic abstract argumentation. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA'12)*, pages 117–128, 2012. 28

[Hun13] Anthony Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013. 28, 29

[JV99] Hadassa Jakobovits and Dirk Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9(2):215–261, 1999. 21

[KBM+13] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, Alan Perotti, Leendert van der Torre, and Serena Villata. Rewriting rules for the computation of goal-oriented changes in an argumentation system. In *Proceedings of the Fourteenth Workshop on Computational Logic in Multi-Agent Systems (CLIMA'13)*, pages 51–68, 2013. 61

[KBMM11] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, and Pavlos Moraitis. Regulating multiparty persuasion with bipolar arguments: Discussion and examples. In *Proceedings of the Journées Francophones sur les Modèles Formels d'Interactions (MFI'11)*, pages 119–129, 2011. 125

[KBMM12] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, and Pavlos Moraitis. Picking the right expert to make a debate uncontroversial. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA'12)*, pages 486–497, 2012. 28, 37, 61

[KBMM14a] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, and Pavlos Moraitis. Empirical evaluation of strategies for multiparty argumentative debates. In *Proceedings of the Fifteenth International Workshop on Computational Logic in Multi-Agent Systems (CLIMA'14)*, pages 105–122, 2014. 125

[KBMM14b] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, and Pavlos Moraitis. On the use of target sets for move selection in multi-agent debates. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence (ECAI'14)*, pages 1047–1048, 2014. 61

[KL11] Souhila Kaci and Christophe Labreuche. Arguing with valued preference relations. In *Proceedings of the Eleventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'11)*, pages 62–73, 2011. 31

[KM03] Antonis Kakas and Pavlos Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*, pages 883–890. ACM, 2003. 17, 25, 29

[KM06] Antonis Kakas and Pavlos Moraitis. Adaptive agent negotiation via argumentation. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06)*, pages 384–391. ACM, 2006. 17

[KMD94] Antonis Kakas, Paolo Mancarella, and Phan Minh Dung. The acceptability semantics for logic programs. In *Proceedings of the International Conference on Logic Programming (ICLP)*, volume 94, pages 504–519, 1994. 18

[KMM05] Antonis Kakas, Nicolas Maudet, and Pavlos Moraitis. Layered strategies and protocols for argumentation-based agent interaction. In *Argumentation in Multi-Agent Systems*, pages 64–77. Springer, 2005. 119

[KPB+13] Dionysios Kontarinis, Alan Perotti, Elise Bonzon, Nicolas Maudet, Leendert van der Torre, and Serena Villata. Using rewriting rules to compute successful modifications of an argumentation system. In *Proceedings of the Septièmes Journées de l'Intelligence Artificielle Fondamentale (JIAF'13)*, pages 180–189, 2013. 61

[KR70] Werner Kunz and Horst W. J. Rittel. *Issues as elements of information systems*, volume 131. Institute of Urban and Regional Development, University of California Berkeley, California, 1970. 8

[KT99] Antonis Kakas and Francesca Toni. Computing argumentation in logic programming. *Journal of Logic and Computation*, 9(4):515–562, 1999. 50

[KW82] David M. Kreps and Robert Wilson. Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253–279, 1982. 11

[LM11] João Leite and João Martins. Social abstract argumentation. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 2287–2292, 2011. 28, 30, 31, 36

[LON12] Hengfei Li, Nir Oren, and Timothy J. Norman. Probabilistic argumentation frameworks. In *Proceedings of the Workshop on Theory and Applications of Formal Argumentation (TAFA'12)*, pages 1–16, 2012. 29

[MC09] Sanjay Modgil and Martin Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in artificial intelligence*, pages 105–129. Springer, 2009. 48, 49, 51, 52, 56, 61

[MGS08] Diego C. Martınez, Alejandro J. Garcıa, and Guillermo R. Simari. An abstract argumentation framework with varied-strength attacks. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, pages 135–144, 2008. 27, 28, 29

[Mod09] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9):901–934, 2009. 17, 26, 28, 34

[Mod13] Sanjay Modgil. Revisiting abstract argumentation frameworks. In *Proceedings of the Second Workshop on Theory and Applications of Formal Argumentation (TAFA'13)*, 2013. 34

[MP09] Peter McBurney and Simon Parsons. Dialogue games for agent argumentation. In *Argumentation in Artificial Intelligence*, pages 261–280. Springer, 2009. 103, 104, 114

[MPW02] Peter McBurney, Simon Parsons, and Michael Wooldridge. Desiderata for agent argumentation protocols. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, pages 402–409, 2002. 105, 120, 161

[MT08] Paul-Amaury Matt and Francesca Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Proceedings of the Eleventh European Conference on Logics in Artificial Intelligence (JELIA?08)*, pages 285–297. Springer, 2008. 24

[Nii86] H. Penny Nii. The blackboard model of problem solving and the evolution of blackboard architectures. *AI magazine*, 7(2):38, 1986. 40

[NK09] Victor Noël and Antonis Kakas. Gorgias-c: Extending argumentation with constraint solving. In *Logic Programming and Nonmonotonic Reasoning*, pages 535–541. Springer, 2009. 25

[OR94] Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994. 110, 111

[OW11] Emilia Oikarinen and Stefan Woltran. Characterizing strong equivalence for argumentation frameworks. *Artificial intelligence*, 175(14):1985–2009, 2011. 48, 55

[PMSW07] Simon Parsons, Peter McBurney, Elizabeth Sklar, and Michael Wooldridge. On the relevance of utterances in formal inter-agent dialogues. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*, page 240. ACM, 2007. 11

[POT69] Chaim Perelman and Lucie Olbrechts-Tyteca. *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press, 1969. 8

[Pra05] Henry Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15(6):1009–1040, 2005. 10, 11, 12, 17, 39, 40, 104, 105, 106, 107, 108, 109, 111, 112, 148, 161, 167

[Pra06] Henry Prakken. Formal systems for persuasion dialogue. *The Knowledge Engineering Review*, 21(02):163–188, 2006. 17, 39, 40, 104, 105, 106, 111

[Pra10] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010. 25

[PS97] Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of applied non-classical logics*, 7(1-2):25–75, 1997. 25

[PTS+11] Simon Parsons, Yuqing Tang, Elizabeth Sklar, Peter McBurney, and Kai Cai. Argumentation-based reasoning in agents with varying degrees of trust. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11)*, pages 879–886, 2011. 38

[RL08] Iyad Rahwan and Kate Larson. Pareto optimality in abstract argumentation. In *Proceedings of the Twenty-third Conference on Artificial Intelligence (AAAI'08)*, pages 150–155, 2008. 153

[RL09] Iyad Rahwan and Kate Larson. Argumentation and game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009. 104, 110, 132

[RPRS08] Régis Riveret, Henry Prakken, Antonino Rotolo, and Giovanni Sartor. Heuristics in argumentation: A game theory investigation. In *Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08)*, pages 324–335, 2008. 119, 121

[RR04] Chris Reed and Glenn Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979, 2004. 8

[RT10] Iyad Rahwan and Fernando Tohmé. Collective argument evaluation as judgement aggregation. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10)*, pages 417–424, 2010. 36

[RTO13] Tjitze Rienstra, Matthias Thimm, and Nir Oren. Opponent models with uncertainty for strategic argumentation. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI'13)*, pages 332–338, 2013. 119, 121

[Rya84] Eugene E. Ryan. *Aristotle's theory of rhetorical argumentation*. Cambridge University Press, 1984. 17

[SCG13] Zaher Salah, Frans Coenen, and Davide Grossi. Extracting debate graphs from parliamentary transcripts: A study directed at uk house of commons debates. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, pages 121–130. ACM, 2013. 8

[Sea69] John R. Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge University Press, 1969. 104

[SLPM10] Oliver Scheuer, Frank Loll, Niels Pinkwart, and Bruce M. McLaren. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1):43–102, 2010. 10

[TBS08] Fernando A. Tohmé, Gustavo A. Bodanza, and Guillermo R. Simari. Aggregation of attack relations: a social-choice theoretical analysis of defeasibility criteria. In *Foundations of Information and Knowledge Systems*, pages 8–23. Springer, 2008. 35

[TG10] Matthias Thimm and Alejandro J. García. Classification and strategical issues of argumentation games on structured argumentation frameworks. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10)*, pages 1247–1254, 2010. 105, 121, 163

[TT11] Francesca Toni and Paolo Torroni. Bottom-up argumentation. In *Proceedings of the Workshop on Theory and Applications of Formal Argumentation (TAFA'11)*, pages 249–262, 2011. 38

[VBvdT11] Serena Villata, Guido Boella, and Leendert van der Torre. Attack semantics for abstract argumentation. In *Proceedings of the Twenty-second international Joint Conference on Artificial Intelligence (IJCAI'11)*, pages 406–413, 2011. 22, 63, 64

[Wal08] Douglas Walton. The three bases for the enthymeme: A dialogical theory. *Journal of Applied Logic*, 6(3):361–379, 2008. 34, 126, 169

[WK95] Douglas Walton and Erik Krabbe. Commitment in dialogue. *Basic Concepts of Interpersonal Reasoning*, 35, 1995. 8, 11, 13, 17, 103, 105

[Woo09] Michael Wooldridge. *An introduction to multiagent systems*. John Wiley & Sons, 2009. 7

[WRM08] Douglas Walton, Chris Reed, and Fabrizio Macagno. *Argumentation schemes*. Cambridge University Press, 2008. 127

Appendix A

Table of Basic Notations

Notation	Meaning
$\langle A, R\rangle$	an abstract argumentation system
A	a set of arguments
R	an attack relation over arguments
D	a defeat relation over arguments
$Pref$	a preference relation over arguments
Adm (resp. $Pref$, $Comp$, $Stab$, Gr)	admissible (resp. preferred, complete, stable, grounded) semantics
$\mathcal{E}_S(AS)$	the set of extensions of system AS, under semantics S
$E_{Gr}(AS)$	the grounded extension of AS
$lab(a)$	the label of argument a
in, out, $undec$	three possible labels of arguments
1, 0, ?	three possible labels of attacks
$\langle A, R, R+, R-\rangle$	an abstract argumentation system with modifiable attacks (ASMA)
R^+	the set of addable attacks of an ASMA
R^-	the set of removable attacks of an ASMA
R^{fix}	the set of fixed attacks of an ASMA
R^{fixN}	the set of fixed non-attacks of an ASMA
R^{deb}	the set of debated attacks of an ASMA
Att	the set $R \cup R^+$ of an ASMA
$S_\exists(d)$	the goal of making argument d credulously accepted under semantics S
$S_\forall(d)$	the goal of making argument d skeptically accepted under semantics S
$\neg S_\exists(d)$	the goal of making argument d not credulously accepted under semantics S
$\neg S_\forall(d)$	the goal of making argument d not skeptically accepted under semantics S
$\mathbb{M}$	the set of all successful moves (for some goal)
$\mathbb{T}$	the set of all target sets (for some goal)
$\langle A, R, Eval\rangle$	a Gameboard (type of Weighted Argumentation System)
$Eval$	a set of evaluation vectors
$\vec{v}(a, b)$	the evaluation vector of attack (a, b): $\vec{v}(a, b) = \langle w(a, b), mw(a, b)\rangle$
$w(a, b)$	the weight of attack (a, b) (numerical)
$mw(a, b)$	the max-weight of attack (a, b) (numerical)
$w(a)$	the weight of argument a (numerical)
$v(a)$	the valuation of argument a (numerical, or part of a totally ordered set)
$val(a)$	the value of argument a (non-numerical)
GB_{cp}	the counterpart (non-weighted) system of Gameboard GB
$R^{\mathbb{W}}_{wk}$ (resp. $R^{\mathbb{W}}_{str}$, $R^{\mathbb{W}}_{bd}$)	the set of weak (resp. strong, beyond doubt) attacks of system $\mathbb{W}$
$A^{\mathbb{W}}_{pers}$ (resp. $A^{\mathbb{W}}_{\overline{pers}}$)	the set of persistent (resp. non-persistent) arguments of system $\mathbb{W}$
$A^{\mathbb{W}}_{dec}$ (resp. $A^{\mathbb{W}}_{undec}$)	the set of decided (resp. undecided) arguments of system $\mathbb{W}$
$\langle A, R, S\rangle$	a bipolar argumentation framework (BAF)
S	a support relation over arguments

Notation	**Meaning**
T	a set of topics
$top(a)$	the set of topics of argument a
$top(a, b)$	the set of topics of attack (a, b)
N	a set of agents
$exp(i)$	the set of topics of expertise of agent i
imp_i(a,b)	the impact of agent i's vote on attack (a, b) (numerical)

Appendix B

Maude's listing

```
mod RP_PROCEDURE is
protecting QID .
******************** SORTS AND SUBSORTS
sorts Attack Argument Sign Atom Conjunct .
subsort Atom < Conjunct . subsort Qid < Attack .
******************** CONSTANTS
ops top btm : -> Atom [ctor] .
ops + - 1 ? 0 * ** # : -> Sign [ctor] .
ops d : -> Argument [ctor] .
******************** VARIABLES
vars X Y : Attack .
vars S T : Sign .
var At : Atom .
******************** OPERATORS
op ___ : Attack Sign Sign -> Atom [ctor] .
op PRO_ : Argument -> Atom [ctor] .
op CON_ : Argument -> Atom [ctor] .
op _hits_ : Attack Attack -> Atom [ctor] .
op _hitsArg_ : Attack Argument -> Atom [ctor] .
op isNotHit_ : Attack -> Atom [ctor] .
op isNotHitArg_ : Argument -> Atom [ctor] .
op _and_ : Conjunct Conjunct -> Conjunct [ctor assoc comm] .
******************** EQUATIONS - SIMPLIFICATION RULES
eq (X S T) and (X S T) = (X S T) .   *** Eq. 1
eq (X S **) and (X S *) = (X S *) .  *** Eq. 2
eq (X S *) and (X S #) = (X S #) .   *** Eq. 3
eq (X 0 S) and (X 1 T) = btm .   *** Eq. 4
eq (X 0 S) and (X ? T) = btm .   *** Eq. 5
eq (X ? S) and (X 1 T) = btm .   *** Eq. 6
eq (X 0 *) and isNotHit(X) = btm .   *** Eq. 7
eq (X ? S) and isNotHit(X) = btm .   *** Eq. 8
eq PRO(d) and isNotHitArg(d) = top .   *** Eq. 9
eq CON(d) and isNotHitArg(d) = btm .   *** Eq. 10
eq At and btm = btm .     *** Eq. 11
******************** REWRITING RULES - EXPANSION RULES
--------- Expansion rules for (X 1 *) atoms (rules 1, 2 and 3) ---------
*** RULE 1: The attack Y is on the system.
crl [expand_X1*_with_YO*] : (X 1 *) and (Y hits X) =>
(X 1 *) and (Y 0 *) if not (substr(string(Y),0,1) == "+") .
*** RULE 2: The attack Y is removable.
crl [expand_X1*_with_Y-#_YO#] : (X 1 *) and (Y hits X) =>
(X 1 *) and (Y - #) and (Y 0 #) if (substr(string(Y),0,1) == "-") .
*** RULE 3: The attack Y is addable.
crl [expand_X1*_with_YO#] : (X 1 *) and (Y hits X) =>
(X 1 *) and (Y 0 #) if (substr(string(Y),0,1) == "+") .
--------- Expansion rules for (X 0 *) atoms (rules 4 and 5) ---------
*** RULE 4: The attack Y is on the system.
crl [expand_X0*_with_Y1*] : (X 0 *) and (Y hits X) =>
```

```
(X 0 #) and (Y 1 *) if not (substr(string(Y),0,1) == "+") .
*** RULE 5: The attack Y is addable.
crl [expand_X0*_with_Y+#_Y1*] : (X 0 *) and (Y hits X) =>
(X 0 #) and (Y + #) and (Y 1 *) if (substr(string(Y),0,1) == "+") .
--------- Expansion rules for (X ? **),(X ? *) atoms (rules 6-12) ---------
*** RULE 6: Sign **, the attack Y is on the system.
crl [expand_X?**_with_Y?**] : (X ? **) and (Y hits X) =>
(X ? *) and (Y ? **) if not (substr(string(Y),0,1) == "+") .
*** RULE 7: Sign **, the attack Y is addable.
crl [expand_X?**_with_Y+#_Y?**] : (X ? **) and (Y hits X) =>
(X ? *) and (Y + #) and (Y ? **) if (substr(string(Y),0,1) == "+") .
*** RULE 8: Sign *, the attack Y is on the system.
crl [expand_X?*_with_Y?**] : (X ? *) and (Y hits X) =>
(X ? *) and (Y ? **) if not (substr(string(Y),0,1) == "+") .
*** RULE 9: Sign *, the attack Y is on the system.
crl [expand_X?*_with_Y0*] : (X ? *) and (Y hits X) =>
(X ? *) and (Y 0 *) if not (substr(string(Y),0,1) == "+") .
*** RULE 10: Sign *, the attack Y is removable.
crl [expand_X?*_with_Y-#_Y0#] : (X ? *) and (Y hits X) =>
(X ? *) and (Y - #) and (Y 0 #) if (substr(string(Y),0,1) == "-") .
*** RULE 11: Sign *, the attack Y is addable.
crl [expand_X?*_with_Y0#] : (X ? *) and (Y hits X) =>
(X ? *) and (Y 0 #) if (substr(string(Y),0,1) == "+") .
*** RULE 12: Sign *, the attack Y is addable.
crl [expand_X?*_with_Y?#] : (X ? *) and (Y hits X) =>
(X ? *) and (Y ? #) if (substr(string(Y),0,1) == "+") .
--------- Expansion rules for PRO, CON atoms (rules 13-19) ---------
*** RULE 13: PRO, and the attack Y is on the system.
crl [expand_PRO_with_Y0*] : PRO(d) and (Y hitsArg d) =>
PRO(d) and (Y 0 *) if not (substr(string(Y),0,1) == "+") .
*** RULE 14: PRO, and the attack Y is removable.
crl [expand_PRO_with_Y-#_Y0#] : PRO(d) and (Y hitsArg d) =>
PRO(d) and (Y - #) and (Y 0 #) if (substr(string(Y),0,1) == "-") .
*** RULE 15: PRO, and the attack Y is addable.
crl [expand_PRO_with_Y0#] : PRO(d) and (Y hitsArg d) =>
PRO(d) and (Y 0 #) if (substr(string(Y),0,1) == "+") .
*** RULE 16: CON, and the attack Y is on the system.
crl [expand_CON_with_Y1*] : CON(d) and (Y hitsArg d) =>
(Y 1 *) if not (substr(string(Y),0,1) == "+") .
*** RULE 17: CON, and the attack Y is addable.
crl [expand_CON_with_Y+#_Y1*] : CON(d) and (Y hitsArg d) =>
(Y + #) and (Y 1 *) if (substr(string(Y),0,1) == "+") .
*** RULE 18: CON, and the attack Y is on the system.
crl [expand_CON_with_Y?**] : CON(d) and (Y hitsArg d) =>
(Y ? **) if not (substr(string(Y),0,1) == "+") .
*** RULE 19: CON, and the attack Y is addable.
crl [expand_CON_with_Y+#_Y?**] : CON(d) and (Y hitsArg d) =>
(Y + #) and (Y ? **) if (substr(string(Y),0,1) == "+") .
endm
```