
# Collective reasoning on multi-agent debates: A coherent approach

**Jordi Ganzer Ripoll**

Supervisors: Prof. Simon Parsons

Dr. Natalia Criado

This dissertation is submitted for the degree of

Doctor of Philosophy

Department of Informatics

King's College London

April 2021

# Abstract

Currently, the Internet and its virtual platforms are the primary forms of communication in our lives. From international to local communities, citizens search for and demand better ways to express their opinion and decide collectively about the world they live in. However, current collective decision making methods have yet to improve to achieve their potential.

Inspired by e-participation systems, that is, online processes involving government and citizens, this dissertation explores multi-agent debates and collective reasoning. We present three novel approaches to represent a multi-agent debate —the Target oriented discussion framework, the Relational model and the Abstract multi-agent debate— and we use them to study collective reasoning methods. The use of dependencies within a debate and coherence, a notion to capture opinion consistency, play a key role throughout this research.

The Target oriented discussion framework structures an argumentation-based debate allowing both positive and negative relationships between the arguments and making it possible for participants to express their opinions about the arguments. In particular, it addresses the problem of how participants can reach an agreement about a single issue being discussed. Several new methods to reach a collective decision are assessed by means of social choice properties. Further to the analysis, a computational assessment shows their applicability in real scenarios.

The Relational model overcomes drawbacks of existing approaches by leaving aside arguments and attack and defence notions to arrange a more general representation of a multi-agent debate. This model clearly distinguishes between different features composing a debate while offering more expressiveness to participants. A family of new opinion aggregation functions is defined, and an exhaustive analysis of their performance regarding their social choice properties is provided. Additionally, a computa-

tional analysis demonstrates that collective opinions can be computed efficiently for real-sized debates.

Finally, the Abstract multi-agent debate model extends the notion of a multi-agent debate allowing it to be an abstraction for different approaches. After proving its capability to represent other debate models, we introduce an approach to analyse the quality of a multi-agent debate.

# Acknowledgements

First of all, my highest gratitude to my supervisors (official and unofficial) throughout this PhD project: Simon Parsons, Natalia Criado, Maite Lopez-Sanchez and Juan Antonio Rodriguez-Aguilar.

Thanks, Jar and Maite, for welcoming me to start my research career with you, promoting me to go forward, providing me with this amazing PhD opportunity,... In short, for helping in any way you could until this point.

Thank you, Simon, for trusting me, a completely unfamiliar face, to course a PhD under your wing. Still amazes me the luck I had to get this opportunity to work with such a great supervisor, researcher, professor,... Above all, such a great person.

Thank you, Natalia, for joining this project and being my second supervisor, providing your excellent point of view (sometimes challenging for me, but always very appreciated), and helping in any way you could.

Simon and Natalia, I am sure I have been quite challenging sometimes, but I want you to know how grateful I am that you always did your best to support, guide, and always find a way to help me with any problem I had.

I can assure you that, in the company of you all, I have never felt helpless. It has been my great pleasure to learn so much and work alongside you all. Any student would be fortunate to just have one of you as a supervisor, imagine my luck for having you all four at my side. One more time, my deepest and sincere gratitude to you all.

I wish to express my deepest gratitude to my family and friends Borja, Patrícia, Júlia and Aina. To my parents and the rest of my family for always supporting me throughout my studies and all my life. You always encouraged and helped me to pursue any path I needed to. To all of you, I want to thank you for simply being there for me and always encouraging me to carry on. I do not forget how often you made an effort to listen to

me about my work even though you do not have the background to understand half of it. I really appreciate it. It is crystal clear to me that, without your emotional support and your company, this journey would have been much more difficult.

In particular, I want to give my special gratitude to my best friend, Borja. Being such a great mathematician, your insight often inspired me to look at my work in new ways and find new paths to follow. But, aside from your help on this project, I want to thank you especially for being such a great friend. By only discussing with me or working on our unusual ideas, many times you gave me a very needed place where I could relax my mind.

To all of you, along with the Department of Informatics and the King's College London, for giving me the opportunity and means to pursue this PhD, my most sincere gratitude.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Nowadays, we are facing increasing challenges that threaten our world as we know it. Climate change consequences are increasing, the Covid-19 pandemic and its subsequent economic and social crises are destabilising many countries, new technologies are coming to stay and change our everyday lives, etc. We will face many challenges within the next decades that can severely impact our world. In the meantime, political institutions are being overwhelmed by the global situation, unable to produce satisfactory solutions for the society they serve.

Simultaneously, social networks are gaining presence and importance in our lives. More and more people use social networks to voice their opinion and hear others, searching for the opportunity to change something about the world surrounding them. Indeed, it is already happening. With daily-increasing frequency, politicians feel the pressure directed towards them in social networks and stay alert for the appearing tendencies that might affect them in the future. However, societies still do not have specialised tools for taking advantage of the potential available on the Internet.

The Internet is an environment that, with the right tools, can serve to strengthen our democracies and improve our society. The Internet provides us with the means to connect citizens and the political institutions that represent them, to enable the voice of the people to be heard more often by those who are allowed to decide for them, instead of only at elections.

On a local scale: neighbourhoods, communities or associations, with the consensus of their members, aim for new policies and actions that improve their surroundings and lives. Such mechanisms need the ability to put together the preferences of their

members, a task which is not always feasible or recommendable in every situation. It may be physically impossible for a large number of members to gather in one place, or even it may be made inadvisable by other circumstances, such as the current Covid-19 pandemic that prevents people from meeting. In such contexts, online participation systems or other methods for collective reasoning appear to be the perfect way to reach their goals. Even at a larger scale: city halls, country governments or even bigger institutions can benefit from having online platforms to gather the collective thoughts of their represented citizens or even make it possible for them to decide on some issues. These tools are a means to obtain a more precise picture of the needs and preferences of the society, which may lead to policies better received by the society. Moreover, it would allow citizens to be more involved in building their society, thus strengthening the connection between voters and representatives.

It is our best interest to research these topics and develop new tools to decide collectively. Appropriate collective decision mechanisms can serve to make citizens more visible to their governments and achieve more consensual decisions affecting their lives. Furthermore, such collective decision mechanisms can engage citizens in politics while providing them with more information about the decisions that affect them.

Social networks are starting to perform such tasks though still disorganised and very susceptible to the tendencies that skew the genuine opinion of society. That is why they have inspired new technologies and research lines aiming for improved methods to ease the communication —e-governance, participatory democracy, or more specifically, e-participation systems [Fung and Wright 2001]. These systems are designed to allow citizens to propose, discuss, and even decide policies through online platforms.

Significant examples of deployed e-participation systems are those used by the local governments in Barcelona [Decidim Barcelona], Reykjavik [Better Reykjavík], Madrid [Decide Madrid], Helsinki [City of Helsinki], and the one used by the French government [Parlement & Citoyens]. In these systems, citizens can carry out structured discussions around diverse topics. For example, the systems deployed by Barcelona, Reykjavik, Madrid, and Helsinki are for proposals about local issues, while the French government system, Parlement et Citoyens, is for discussing potential national legislation. Some other systems are not tied to a specific institution, such as [Consider.it] and [Appgree] whose main focus is scalability — making the systems fit for use by large

numbers of participants.

From an academic perspective, several research areas have explored how to address the topic of collective decision-making. *Social choice theory* [Arrow 1963], studies how to find a collective outcome from individual inputs such as votes, preferences, etc. [List 2018]. For instance, given a set of alternatives and a set of agents who possess preference relations over the alternatives, social choice theory focuses on how to yield a collective decision that appropriately reflects the agents' individual preferences. A related approach, *judgement aggregation*, whose focus is on finding consistent collective judgements on a set of propositions based on the group of individual judgements on them [List and Pettit 2002, Endriss and Moulin 2016]. A different approach that focuses on resolving conflicts in opinions is *computational argumentation* [Rahwan and Simari 2009]. Given a set of arguments and a set of attack relations (conflicts) between the arguments, argumentation is concerned with identifying those arguments that a rational agent might accept. Mixing judgement aggregation or social choice theory with argumentation, we find several proposals that structure debates using argumentation frameworks, such as in [Leite and Martins 2011, Awad et al. 2015]. These allow participants to put forward arguments, relations between arguments, and opinions about which of these arguments and relationships hold for them. The systems then produce an output intended to reflect the collective opinion of the participants in the debate.

Within this context, our research contributes further to the area of collective reasoning. This research focuses on collective reasoning processes regarding many subjects by developing in formal terms: new models to represent a multi-agent debate, novel concepts that help to analyse formally the information contained in a discussion, and several methods to tackle the problem of obtaining a collective decision for an issue under discussion.

## 1.1 Motivation

Currently, the Internet and its virtual platforms are the primary forms of communication among people around the globe. Nowadays, the Covid-19 pandemic forced the Internet to be the main form of communication. From international to local communities, virtual platforms are used to connect lives and put together the interests and thoughts of their

users.

Many settings and approaches can be used to provide a deliberation process depending on the particular requirements of the platform, the actions allowed in the discussions, the methods to gather information, etc. This research builds upon a theoretical approach that can be studied in general terms and applied to many real applications. For such reason, the debate models introduced in this work aim at abstractly representing different kinds of multi-agent debates that are sufficiently general to enable the computational implementation of solutions to participatory systems. The term "agent" generally refers to an autonomous, rational and self-interested party, such as individuals, groups of individuals (represented together), companies, countries, etc., but, for the sake of clarity in the topic of this research, the term "agent" hereafter refers to a person that can interact with others in a debate and provide new inputs, i.e. a participant. This way, we use the expression "multi-agent debate" to refer to a debate multiple participants hold.

The debate models provided in this work aim to represent important features of a debate that have not been considered in existing approaches tackling the same topic, such as in [Leite and Martins 2011, Klein 2012, Awad et al. 2015] and others (see Chapter 2). An argumentation approach to organise a debate (as in [Dung 1995]) can restrict the participants' expressiveness when we consider the complexity of human thinking and deliberative processes. New and different forms of organisation might be helpful to produce a more accurate picture of a collective. In addition to this topic, the participants' opinions in a debate might not always be in terms of good or bad, so we should consider new forms of opinion.

In any deliberation process, we expect participants to express their opinions on the topic under discussion freely. However, these opinions might not always be reasonable, either by the participant's bounded rationality or by the debate specifications, which may hinder their expressiveness. In this matter, a study of the rationality or consistency of the participants' opinions has been widely used to characterise the correctness of an opinion and, if necessary, to provide corrections in the debate ([Dung 1995, Leite and Martins 2011, Rago and Toni 2017], see Chapter 2). This research also considers this matter to provide a new notion named *coherence*, which is a less restrictive characterisation compared to the rationality that other approaches use [Dung 1995, Caminada

2006, Awad et al. 2015]. Instead of identifying an opinion as not consistent when an element (whether an argument, a statement and so on) is contradicted by one related —i.e. one claim negates another—, it considers the consistency of its average support from related elements.

Concerning collective decision processes, online participation systems usually involve a considerable amount of people discussing an issue to make a decision. In such cases, manually extracting the collective decision (as opposed to automatically) can be difficult, if not impossible. Besides, we can consider many methods to extract a collective decision (majority rule, unanimous decision, etc.), though not all of them would be deemed acceptable by the participants in the debate. The study on this matter — social choice theory, judgement aggregation, see Chapter 2— considers many factors that influence the aggregation to extract a collective decision. This research addresses this topic by providing several aggregation functions to reach a collective decision and analyses its features regarding desirable properties for the collective decision. Particularly, the focus is on exploiting the connections among the several pieces of information and the participants' opinions on them. Although this work offers some insight into how we may build a multi-agent debate, the focus is on processing the information obtained, not the progress of an open debate. For this reason, strategic manipulation —i.e. the forms in which a group might act strategically to manipulate the results—, due to its significant relationship with the characteristics of how a debate is built (i.e. visibility of the information in the debate, the restrictions that apply to the participants, etc.), is not in the scope of this research.

In addition to the collective decision, a multi-agent debate is a complex process involving many factors that can affect the quality of the participation process, i.e. the participants' opinions might be negatively affected by the debate construction. Either by misunderstanding how the debate should work, limitations when presenting the information, or other causes, a debate may have issues restricting the expressiveness of the participants and, consequently, using optimally any collective reasoning method applied to it. For such reason, a quality assessment of a debate can be a helpful tool to improve its performance. On this matter, this research offers a novel approach for such analysis that can detect structural problems in a multi-agent debate.

## 1.2 Research questions

This thesis comprises three main parts undertaking different studies on two particular models and a general model, respectively. The first two models, the *Target oriented discussion framework* (TODF) and the *Relational model* (RM), in different manners, aim at representing a multi-agent debate in which to study collective reasoning. The third model, the *Abstract multi-agent debate* (AMAD), generalises several multi-agent debate models.

Many issues arise in the study of collective reasoning in a deliberation process. This work answers the research questions introduced below with their justification.

**RQ-1** ***Can we find new models to represent a multi-agent debate?***

We can undertake a multi-agent deliberation process from many perspectives and consider many features for it. There are several ways we can structure the information, different actions that the participants can perform to participate, the availability and accessibility of the information on a platform, and many more. To answer the question, this thesis proposes three different and novel models to represent a multi-agent debate. Although each one represents a debate in distinct forms, all three share some common features and perspectives.

Since the main objective of this research is to study collective reasoning in a debate, the formalisation in all three models aims to *capture a whole debate*, i.e. the resulting debate once all the agents have concluded their participation and all the information (arguments, opinions, etc.) has been collected to be processed. Consequently, any study about the process or ethical codes while creating a debate is beyond this thesis's scope.

To represent a debate, all three models, by means of relationships, *capture the connections among inter-related pieces of information* forming a directed graph. In addition, *the participants' opinions is attached to the information structuring the debate* in the form of evaluations. The opinions, though in different forms in each model, are included in the debate to represent each participant's point of view.

Furthermore, we *the participant's opinions to be expressed freely* in the discussion. Therefore, no rationality or consistency constraints are applied to the

participants' opinions. Even though consistency is a valuable attribute for an opinion, which is why the notion of coherence is created, assuming rational interactions in a human context can be unrealistic. Human thinking might sometimes be unreasonable or emotional but not negligible for these reasons.

With the above considerations in mind, additional features determine the formal representation of a multi-agent debate model. The specific formalisation of each model correlates to a particular point of view of a debate.

The TODF is a novel framework designed to represent a discussion, aiming to decide on a single topic, in which participants express their opinions over arguments that relate to other arguments. Thus, the TODF extends the abstract argumentation framework [Dung 1995][1], by structuring a debate using *abstract arguments*, *attack and defence relationships*[2] and a *labelling system* as in [Awad et al. 2015][3] to represent the opinions over the arguments. In addition, a *target argument* is distinguished from the rest to represent the discussion's topic and final aim of the discussion.

The RM aims to relax some of the constraints of the TODF by not only leaving aside abstract arguments but also allowing participants to express more accurately their opinions on the information. The RM aims to describe a debate as a collective reasoning process where participants can provide their point of view. The RM structures the information in the debate by means of elemental *statements*, which do not contain any kind of reasoning, that relate to other statements by reasoning steps represented by *directed relationships*. In addition, the participants' opinions are two-sided, thus able to represent two different types of opinions. They are issued over the statements and the relationships by means of two continuous real-valued functions, which provide more nuanced and expressive opinions. Similarly to the TODF, the RM also represents a discussion aimed at resolving a set of topics. Thus, it distinguishes a set of statements as

[1]The abstract argumentation framework from Dung [1995] only uses an attack relationship.

[2]Similar to bipolar argumentation[Cayrol and Lagasquie-Schiex 2005, Cayrol and Lagasquie-Schiex 2005, Amgoud et al. 2008], though, here, the "support" relationship is called defence and we provide the framework with more features.

[3]We employ the labellings only as symbolic representation, whereas Awad et al. [2015] *do* use the semantic approach used in [Caminada 2006].

*targets* of the debate.

Finally, the purpose of AMAD is to abstract the elemental characteristics of a multi-agent debate to generalise different types of specific debate models. In particular, AMAD can generalise both the TODF and RM. This way, any analysis performed on AMAD can provide general results for any particular models that AMAD extends. Explicitly, AMAD captures the graph-like structure of a debate by means of *abstract nodes and relationships* and the participants' opinions on the structure using *functions to describe several types of opinions*. Differently from any particular model, though, AMAD does not impose any behaviour or predefined semantic interpretation on its components.

**RQ-2** ***Can we find a more flexible notion of consistency for an opinion?***

Usually, an opinion about some topic is considered to be rational or consistent when its parts act together harmoniously, i.e. when the opinions support each other and do not contradict. An excellent example of this notion is the extensions created by Dung [1995], on which an argument cannot be accepted if it has a counterargument already accepted. Notice that the previous example is very strict. Only one accepted argument attacking is needed to create an inconsistency, regardless of the other arguments that might be related.

In this dissertation, we aim to create a more flexible definition that, for example, would allow an argument and its counterargument to be both accepted (when other conditions are satisfied). This concept will be captured by *coherence*, which in this dissertation generalises the classical notion of rationality, widely used in the literature (e.g. [Dung 1995, Baroni et al. 2011, Awad et al. 2015]).

We will define a notion of coherence within the three models presented in this dissertation. We will understand a coherent opinion to be a collection of opinions that do not contradict excessively when considering their connections. In basic terms, a coherent opinion will guarantee that an opinion is in line with its directly related opinions. As an example contrasting the previous with Dung's rationality, an accepted argument, attacked by three other arguments, will be coherent if two of its attackers are not accepted, even when the third attacker is accepted (i.e. the accepted attackers are not the majority).

Our notion of coherence will prove very useful when assessing the performance of the aggregation functions and analysing the quality of a debate.

**RQ-3** ***Can we use dependencies to aggregate a collective opinion?***

As pointed out, the TODF and RM aim to set a debate from where to obtain a collective decision, called the *collective opinion*. The collective opinion is aimed to be the best representative of the plural opinions given by the agents that took part in the discussion. Thus, the main goal of this research is to find means to aggregate the agents' opinions into a single opinion that represents them all.

Departing from other work [Caminada and Pigozzi 2011, Awad et al. 2015, Chen and Endriss 2019], the opinion aggregation functions proposed in this work aspire to maximise the use of the relationships within the structure of a debate, called *dependencies*. This way, the collective opinion output from an aggregation function will consider the influence that the opinions cause on other opinions via the existing relationships in the debate. In both the TODF and the RM, several aggregation functions defined explore different ways to exploit these dependencies in a wide range of aggregation operators.

**RQ-4** ***Can we assess the aggregation functions considering the dependencies in a debate?***

A deliberation process's collective opinion should represent the participants' opinions faithfully. For this reason, many behaviours and features have been identified as beneficial for an aggregation process. In this matter, Social choice theory [List and Pettit 2002] has classified several desirable properties for an aggregation function to have. Thus, this research borrows these social choice properties and proposes new ones to assess the proposed aggregation functions.

The social choice properties, adapted to apply in the TODF or the RM, can characterise behaviours in the aggregation process, such as the ability to consider equally important the opinions of the different participants or the capacity to recognise and respect a unanimous opinion from the participants. The novel properties are designed to identify additional behaviours that take into account aggregation functions exploiting dependencies. As an example, the Endorsed unanimity property characterises how an aggregation function behaves when the

agents' unanimity is on the related opinions of an argument instead of the argument itself.

**RQ-5** ***Does considering dependencies benefit the aggregation?***

In the TODF and RM models, each aggregation function produces a collective opinion exploiting the dependencies in the debate differently. Then, we use the above-mentioned social choice properties to exhaustively analyse each aggregation function and discover its strengths and weaknesses. Complementary, to better grasp the implications of the aggregation problem, we provide a comprehensive comparison of the aggregation function in terms of the properties they fulfil. This comparison helps us determine the trade-offs for choosing one method over another and, therefore, select the aggregation function that offers the best trade-off. Moreover, the analysis of the opinion aggregation functions also provides a study of their computational complexity.

**RQ-6** ***Can we assess the quality of a debate?***

Finally, this dissertation explores a different line of research, an assessment of the quality of a debate. The quality of a deliberative process can relate to many features [Friess and Eilders 2015], such as diversity of opinion, the rationality of the participants, the relevance of the participants' opinions, etc. In fact, studying the performance of an aggregation function for a debate is an example of quality analysis in a debate. However, this last line of research does not aim at producing new aggregation functions nor expanding the study on them. The analysis offered to answer this question focuses on detecting possible situations in a debate that may suffer structural problems, i.e., issues relating to their structure.

AMAD, whose abstraction can be used to analyse many specific debate models at once, is the model where to study this issue. A generalised analysis, called Systematic incoherence analysis, uses the notion of coherence in AMAD to identify places with possible structural problems in a debate.

## 1.3 Contributions

Having presented the different goals we tackle in this project, we list below the specific contributions of this research. The list of contributions is divided into three categories: Formalisation of multi-agent debates in Section 1.3.1, Coherence and quality in Section 1.3.2 and Collective opinion in Section 1.3.3.

### 1.3.1 Formalisation of multi-agent debates

The three new models, presented in independent parts of this dissertation, can represent a multi-agent debate in different forms. We list below the particular features distinguishing each model.

**1 –** *The Target oriented discussion framework* (TODF).

- A *defence relationship* as the counterpart of the attack relationship.
- *Target of the debate*. A single argument is set to be the root and goal of the discussion.
- *No cyclic relationships*. Supporting the target oriented structure, this feature allows the debate to point directly or indirectly towards the target argument.

**2 –** *The Relational model* (RM).

- *Clear distinction between debate structure and opinion*. The structural information, i.e. the statements and relationships that organise the debate, is completely distinguished from the participants' opinions, which are issued on the elements of the structure.
- *Reasoning-based discussion*. RM leaves behind an argumentation scheme to structure the information in the debate. Instead, it uses *statements* without reasoning connected via *reasoning relationships*.
- *Targets of the debate*. Similarly to TODF, RM characterises a set of statements to be the root and goal of the discussion.
- *Different types of opinions*. Given two structural objects organising the debate, statements and relationships, two types of opinions are related to each of them.

- *Continuous opinions*. The functions representing the opinion on the structural objects are real-valued and continuous.

**3 –** *The Abstract multi-agent debate* (AMAD).

- *Distinction between structure and opinion*. Similar to RM, AMAD distinguishes between the information organising the debate, the structure, and the participants' opinions issued on its components.
- *Semantic-free components*. Neither the structure nor the opinions intend to have a specific semantic interpretation of each component. This characteristic allows AMAD to represent many types of debates sharing its elemental features —for example, the nodes of the structure can be understood as arguments, statements or any other type of object that can relate to others through relationships.
- *Undefined opinion functions*. To enable the generalisation of AMAD, the opinion functions can be defined using either discrete or real values.

### 1.3.2 Coherence and quality

Two novel approaches serve to analyse the features of a debate and can be used to improve both the construction of a debate and its resulting collective aggregation.

**4 –** *The notion of coherence*. Coherence captures an original approach to understanding consistency in the opinions. This notion is defined for TODF and RM to characterise a more flexible notion of rationality than the widely used approach defined by Dung [1995]. Furthermore, the generalised coherence defined for AMAD proves to be a basic characterisation for consistency.

**5 –** *The Systematic incoherence analysis*. Using the notion of coherence over the AMAD model, a novel method to assess the quality of a multi-agent debate is defined. The Systematic incoherence analysis, based on analysing the participants' opinions, detects problematic parts of a debate suffering from structural organisation issues.

### 1.3.3 Collective opinion

Finally, several contributions concern the final stage of a deliberation process to obtain a collective opinion.

**6 –** *New social choice properties*. We define several new properties to assess the quality of opinion aggregation functions:

- for TODF: Collective Coherence, Endorsed Unanimity and Familiar Monotonicity; and
- for RM: $\epsilon$-Collective coherence, Sided Unanimity, Weak Unanimity, Endorsed Unanimity and Familiar Monotonicity.

**7 –** *Novel aggregation functions* for computing collective decisions for both TODF and RM.

Part I defines the following aggregation functions for TODF: the Opinion first function, the Support first function and the Balanced function.

Part II, concerning RM, introduces two families of opinion aggregation functions: $\alpha$-Balanced and the $\alpha$-Recursive families, which explore in different manners the dependencies exploited to form a collective opinion.

## 1.4 Publications

The following list of publications is related to the Target oriented discussion framework.

**(i)** J. Ganzer-Ripoll, M. López-Sánchez, and J. A. Rodriguez-Aguilar. A multi-agent argumentation framework to support collective reasoning. In *International Workshop on Conflict Resolution in Decision Making*, pages 100–117. Springer, 2016.

**(ii)** J. Ganzer-Ripoll, M. López-Sánchez, and J. A. Rodriguez-Aguilar. A Target-Oriented Discussion Framework to Support Collective Decision Making. In N. Criado Pacheco, C. Carrascosa, N. Osman, and V. Julián Inglada, editors, *Multi-Agent Systems and Agreement Technologies*, pages 481–489, Cham, 2017. Springer International Publishing.

**(iii)** J. Ganzer-Ripoll, N. Criado, M. Lopez-Sanchez, S. Parsons, and J. A. Rodriguez-Aguilar. Combining social choice theory and argumentation: Enabling collective decision making. *Group Decision and Negotiation*, 28(1):127–173, 2019.

**(iv)** M. Serramià, J. Ganzer, M. López-Sánchez, J. A. Rodríguez-Aguilar, N. Criado, S. Parsons, P. Escobar, and M. Fernández. Citizen Support Aggregation Methods for Participatory Platforms. In *Frontiers in Artificial Intelligence and Applications: Artificial Intelligence Research and Development*, volume 319, pages 9–18. IOS PRESS, 2019

From the previous publication list: items (i), (ii) and (iii) explore the theoretical research on TODF and (iv) presents a comparison between TODF and the Proposal argument map (PAM) model when applied to real data from [Decidim Barcelona]. Part I develops the contents of these publications.

The article below summarises all the findings relating to RM, presented in part II:

**(vi)** J. Ganzer, N. Criado, M. Lopez-Sanchez, S. Parsons, and J. A. Rodriguez-Aguilar. A model to support collective reasoning: Formalization, analysis and computational assessment. `https://arxiv.org/abs/2007.06850`, 2020.

Finally, we list below the online code base for implementations of, respectively, the aggregation methods for TODF and RM:

- J. Ganzer, N. Criado, M. Lopez-Sanchez, S. Parsons, and J. A. Rodriguez-Aguilar. Collective decision making library. `https://bitbucket.org/jariiia/argumentation-for-collective-decision-making/src/master/`, 2017.
- J. Ganzer, N. Criado, M. Lopez-Sanchez, S. Parsons, and J. A. Rodriguez-Aguilar. Relational model library. `https://bitbucket.org/jariiia/workspace/projects/DRF`, 2020.

## 1.5 Structure of the thesis

As already mentioned, this research explores several topics that have been divided into three main parts. In line with this, we present below how the dissertation is structured.

Chapter 2, *Literature Review*, surveys the existing literature regarding the collective reasoning area relating to this project and points out where each contribution of this project relates within the literature.

Part I, *Target Oriented Discussion Framework*, presents all the research relating to the TODF: Chapter 3 introduces and formally defines the TODF model, presenting its features and contributions, and describes the coherence concept for an agent's opinion; Chapter 4 defines several aggregation functions analysing their social choice properties and computability characteristics; Chapter 5 summarises the contributions of TODF and highlights the issues that motivated the second part.

Part II, *Relational model*, explains the RM: Chapter 6 motivates and formally defines the RM and introduces the new notion of coherence in RM; Chapter 7 defines a complete set of aggregation functions over the RM and analyses its properties and computability results; Chapter 8 concludes this part by reviewing the research and contributions of this part of our research related to RM.

Part III, *Abstract multi-agent debate*, introduces the AMAD and the quality analysis on it: Chapter 9 presents the AMAD and the main definitions that will be used afterwards; Chapter 10 shows the applicability of AMAD and formally introduces the systematic incoherence analysis, a method to identify structural problems in a debate. Chapter 11 summarises the progress of part III.

To conclude this dissertation, Chapter 12 summarises the main contributions of this research and presents future lines of work that could follow from it.

# Chapter 2

# Literature review

This dissertation tackles, in different forms, the features and procedures relating to a multi-agent debate. This chapter identifies several broad research areas that connect to this research.

Section 2.1 offers background knowledge that connects closely to this research, whereas Section 2.2 conducts a literature review to identify connections with other works and weaknesses that we address in this thesis.

## 2.1 Background

Next, we introduce in Section 2.1.1 some background concepts from [Dung 1995] and [Caminada 2006]. These two articles define the basic concepts of abstract argumentation required to understand [Awad et al. 2015], reviewed in Section 2.1.2.

### 2.1.1 Abstract argumentation

We reproduce basic definitions from [Dung 1995] to provide a basic understanding of the study of acceptable sets of arguments in an argumentation debate. This approach regards a debate as a set of abstract arguments that can formally relate to other arguments by means of an attack relationship. It is formalised in the following manner:

**Definition 2.1.1** (Argumentation framework)**.** An *argumentation framework* is a pair $AF = \langle AR, Att \rangle$, where $AR$ is the set of arguments and $Att \subset AR \times AR$ is a binary relationship representing the attacks between arguments.

In his argumentation framework, Dung defines several notions regarding different characterisations of the arguments relating to the attacks of other arguments.

**Definition 2.1.2** (Argumentation concepts)**.**

(1) A set of arguments $S \subset AR$ is said do be *conflict-free* if there are no arguments $a, b \in S$ such that $a\ Att\ b$.

(2) An argument $a \in AR$ is *acceptable* with respect to a set $S \subset AR$ iff for each argument $b \in AR$: if $b\ Att\ a$ then $b$ is attacked by $S$, i.e. exists $c \in S$ such that $c\ Att\ b$.

(3) A conflict-free set $S$ is *admissible* iff each argument in $S$ is acceptable with respect to $S$.

Using this characterisation Dung [1995] is able to provide several semantics for the argumentation framework. It defines a diverse classification of consistent sets called *extensions*, formally defined below.

**Definition 2.1.3** (Semantics)**.**

(1) A *preferred extension* in an argumentation framework $AF$ is a maximal admissible set of $AF$, with respect to the set inclusion.

(2) A conflict-free set $S$ is called a *stable extension* iff $S$ attacks each argument which does not belong to $S$.

(3) An admissible set $S$ is a *complete extension* iff each argument acceptable with respect to $S$ belongs to $S$.

(4) A *grounded extension* of $AF$, is the least complete extension with respect ot the set inclusion.

Among several properties, the following proposition shows how the extensions relate to each other in different ways.

**Proposition 2.1.1.**

(1) Each preferred extension is a stable extension.

(2) For each admissible set $S \subset AF$, there exists a preferred extension $E \subset AF$ such that $S \subset E$.

(3) $S$ is a stable extension iff $S = \{a \in AR \mid a \text{ is not attacked by } S\}$.

(4) Each preferred extension is a complete extension.

(5) The complete extensions form a complete semi-lattice with respect to set inclusion.

(6) Every well-founded $AF$ has only one complete extension, which is also grounded, stable and preferred —an $AF$ is well-founded if there is no infinite ordered sequence of arguments where each attacks the next in line.

These basic definitions and propositions are then studied in [Caminada 2006] using a different approach that changes the view from consistent sets of arguments to that of using a labelling system to identify the acceptable and rejected arguments in the framework, as formally defined below.

**Definition 2.1.4** ($AF$-labelling)**.** Let $AF = \langle AR, Att \rangle$ be an argumentation framework. An *$AF$-labelling* is a total function $L : AR \longrightarrow \{\texttt{in}, \texttt{undec}, \texttt{out}\}$ mapping each argument $a \in AR$ to a label `in` (accepted), `undec` (undecided) or `out` (rejected).

Using this labelling system Caminada [2006] defines the notion of *reinstatement labelling* to make then an equivalence between different types of reinstatement labellings and the different extensions defined in [Dung 1995].

**Definition 2.1.5** (Reinstatement labelling)**.** Let $L$ be an $AF$-labelling. $L$ is a reinstatement labelling iff it satisfies the following:

- For all $a \in AR$, $L(a) = \texttt{out}$ iff exists $b \in AR$ such that $b \; Att \; a$ and $L(b) = \texttt{in}$.
- For all $a \in AR$, $L(a) = \texttt{in}$ iff for all $b \in AR$ such that if $b \; Att \; a$ then $L(b) = \texttt{out}$.

The set of accepted arguments related to a labelling $L$ is $\texttt{Lab2Ext}(L) = \{a \in AR \mid L(a) = \texttt{in}\}$. The labelling associated to a set of accepted arguments $S$ is $\texttt{Ext2Lab}(S) = \{(a, \texttt{in}) \mid a \in S\} \cup \{(a, \texttt{out}) \mid \exists a' \in S \; : \; a' \; Att \; a\}$

The next proposition is a concise summary of many results from [Caminada 2006] showing how different types of reinstatement labellings relate to the different extensions defined in [Dung 1995].

**Proposition 2.1.2.** Let $\langle Arg, Att \rangle$ be an argumentation framework and $L$ a reinstatement labelling.

1. If $\texttt{undec}(L) = \{a \in Arg \mid \mathcal{L}(a) = \texttt{undec}\} = \emptyset$ then $\texttt{Lab2Ext}(L)$ is a stable extension. If $S$ is a stable extension then $L = \texttt{Ext2Lab}(S)$ is a labelling such that $\texttt{undec}(L) = \emptyset$.

2. If $\texttt{in}(L) = \{a \in Arg \mid \mathcal{L}(a) = \texttt{in}\}$ is maximal, $\texttt{out}(L) = \{a \in Arg \mid \mathcal{L}(a) = \texttt{out}\}$ is maximal or $\texttt{undec}(L)$ is minimal then $\texttt{Lab2Ext}(L)$ is a preferred extension. If $S$ is a preferred extension then $L = \texttt{Ext2Lab}(S)$ is a labelling such that $\texttt{in}(L)$ and $\texttt{out}(L)$ are maximal.

3. If $\texttt{undec}(L)$ is maximal, $\texttt{in}(L)$ is minimal or $\texttt{out}(L)$ is minimal then $\texttt{Lab2Ext}(L)$ is a grounded extension. If $S$ is a grounded extension then $L = \texttt{Ext2Lab}(S)$ is a labelling such that $\texttt{in}(L)$ and $\texttt{out}(L)$ are minimal and $\texttt{undec}(L)$ is maximal.

The next section provides a brief survey of the work by Awad et al. [2015].

### 2.1.2 Judgement aggregation using abstract argumentation

The work in [Awad et al. 2015] builds upon the abstract argumentation framework and the labelling system, provided by Dung [1995] and Caminada [2006] respectively, to set the grounds for a multi-agent debate and the subsequent judgement aggregation study on it. Awad et al. [2015] consider a collection of labellings from the participants to be issued on a shared set of arguments connected via the attack relationship, i.e. an argumentation framework. The problem to solve then is how to compute a set of labels for the arguments that reflects the opinions of all the participants and which desirable social choice properties the aggregation of opinions satisfies.

Next, we introduce those definitions from [Awad et al. 2015] that relate more to the research in this dissertation. Given that previous Section 2.1.1 introduced the argumentation framework, in definition 2.1.1, and the AF-labellings, or as it is called in [Awad et al. 2015] argument labelling, in definition 2.1.4, we introduce directly the next concepts:

**Definition 2.1.6** (Labelling aggregation problem [Awad et al. 2015])**.** Let $Ag = \{1, \ldots, n\}$ be a finite non-empty set of agents and $AF = \langle AR, Att \rangle$ be an argumentation

framework. A *labelling aggregation problem* is a pair $\mathcal{LAP} = \langle AF, Ag \rangle$.

Thus, a labelling aggregation problem is a framework representing the debate together with the agents involved in it. The agents' opinions on the framework form the labelling profile:

**Definition 2.1.7** (Labelling profile [Awad et al. 2015])**.** Let $\mathcal{LAP} = \langle AF, Ag \rangle$ be a labelling aggregation problem. The collection $\mathcal{L} = (L_1, \ldots, L_n) \in \mathbf{L}(AF)^n$ denotes a *labelling profile*, where $\mathbf{L}(AF)$ is the class of labellings over an $AF$. $\mathcal{L}(a) = (L_1(a), \ldots, L_n(a)$ is the labelling profile of the argument $a \in AR$.

Finally, the function in charge of combining the opinions into one single labelling is the aggregation function.

**Definition 2.1.8** (Aggregation function [Awad et al. 2015])**.** Let $\mathcal{LAP} = \langle AF, Ag \rangle$ be a labelling aggregation problem. An *aggregation function* for $\mathcal{LAP}$ is a function $F : \mathbf{L}(AF)^n \longrightarrow \mathbf{L}(AF)$.

Once the aggregation problem and the method to solve it are defined, they describe the social choice properties that will serve to assess the aggregation functions. Next, we present a list of social choice properties from [Awad et al. 2015][1]:

- Universal domain. $F$ can take as input any profile of $\mathbf{L}(AF)^n$.
- Collective rationality. The outcome of the aggregation, $F(\mathcal{L})$, is a complete labelling.
- Anonymity. The result of $F$ does not depend on the order of the agents, i.e., for any $\mathcal{L} = (L_1, \ldots, L_n)$, if $\mathcal{L}' = (L_{\sigma(1)}, \ldots, L_{\sigma(n)})$ where $\sigma$ is a permutation then $F(\mathcal{L}) = F(\mathcal{L}')$. A consequence of satisfying this property is non-dictatorship, an additional property to state the fact that no single agent can unilaterally decide on an outcome labelling.
- Unanimity. If all agents label equally one argument, then the aggregation function must output that label.

[1]Some of these are only mentioned but not formally defined due to not being of particular relevance to the present research.

- Independence. On any argument, the result of the aggregation function only depends on the values given to that argument. Awad et al. also define two more restrictive notions relating to the independence of the aggregation, called Weak and Strong systematicity.

- Supportiveness. The outcome of the aggregation on one argument must be one of the labels issued by some agent.

- Monotonicity. If some agents switch their label on an argument to the label resulting from the aggregation function on that argument, then the aggregated label remains the same. A variation of monotonicity, called `in`/`out`-montonicity, considers only the case when restricting the labels to be `in` or `out`.

The first, and unique, aggregation function Awad et al. [2015] study is the majority function, which in they call the Argument-wise plurality rule:

**Definition 2.1.9** (Argument-wise plurality rule [Awad et al. 2015])**.** Given any argument $a \in AR$ and any profile $\mathcal{L}$ it holds that $[M(\mathcal{L})](a) = l_a \in \{\texttt{in}, \texttt{undec}, \texttt{out}\}$ iff

$$|\{i \in Ag \mid L_i(a) = l_a\}| > \max_{l'_a \neq l_a} |\{i \in Ag \mid L_i(a) = l'(a)\}|$$

As can be easily observed, $M$ is not defined for those profiles involving ties on an argument, therefore it does not satisfy the Universal domain property. After a more detailed analysis in [Awad et al. 2015], the social choice properties fulfilled by $M$ are anonymity, unanimity, independence, supportiveness and monotonicity.

Additional results are presented in [Awad et al. 2015], however, these results are not closely related to the present research, and we do not review them here.

## 2.2 Related work

First, we review the deliberative processes in general terms to set some basic concepts that relate to a multi-agent debate. Among many, the following two definitions serve to establish what *deliberation* is: "the reasoned and well-informed discussion of opinions by the people involved, under conditions of equality and respect" [List 2018], and "an informed discussion between individuals about issues that concern them" [Friess and

Eilders 2015]. Thus, a deliberation process between multiple people can be regarded as a multi-agent debate.

Deliberation has been studied from social, philosophical and political perspectives [Freeman 2000, Friess and Eilders 2015, Landemore and Page 2015, List 2018], but these studies are not closely related to the research presented here. However, a basic conceptualisation of deliberation on these terms will help classify this section's contents better.

As established in [List 2018], there is a distinction between *deliberative procedures* —the settings in which participation can take place —and *deliberative behaviours* — the ways in which people actually discuss. However, for the purpose of this chapter, it will prove more useful to adopt the classification proposed in [Friess and Eilders 2015], where the study of deliberation in participation systems is divided into three main levels, *input, throughput* and *output*:

- *Institutional **input** level*, which relates to the conditions of the deliberation, i.e., how to design an environment for a deliberation process. A few features relating to this level can be the actions that agents can take to participate, the possibility of a moderation team to manage the participation, and the form with which the arguments are issued.

- *Communicative **throughput** level* refers to the question "how should people communicate?". That is aimed at deciding rules and codes of behaviour for the people to interact in a discussion. Some of the questions relating to this level are "Do we require rationality in the participants' arguments?", "Are participants being respectful to each other?" or "Do participants feel equally important as the other participants?".

- *Productive **output** level*, as its name indicates, is concerned with the result of the deliberation, for example, paying attention to the individual emotions of the participant after deliberating or the properties of the result obtained from it.

Each of the above-mentioned levels can be further characterised with several subcategories, but we do not need these further distinctions in this work.

Section 2.2.1 reviews the most important participation systems that have been produced for online discussions.Clearly, a participation system must address the three deliberation levels presented above. However, although the main goal of this research is to study the output level of the deliberation, aiming at the collective output of a debate, Section 2.2.1 focuses on the input level, or more precisely, the design of a discussion among multiple participants. The section presents several real participation systems to review how they structure and construct a debate and compare them to the design of the models introduced in this project.

Section 2.2.2 surveys the research on computational argumentation, which relates closely to the present research in several ways. In this area, although many references help us to distinguish among different argumentation approaches and the research in this work, especially in terms of its input level features (e.g. design, objects, procedures, etc.), we survey aspects of the throughput level, namely, opinion and rationality of the participants.

Third, Section 2.2.3, reviews the different approaches to produce collective decisions from multi-agent participation systems and how they relate to part I and II. Here the output level is of importance. From aggregation methods to social choice properties, the approaches from the literature for obtaining an outcome from a discussion and its desirable features are contrasted with the collective reasoning approaches studied in this research.

Finally, in Section 2.2.4, we present different quality measures for debates and how this research contributes to the area using an alternative approach. Section 2.2.4 exclusively focuses on the throughput level, relating to how deliberation is supported and the problems that might affect the development of a debate.

### 2.2.1 Tools for online discussion

As mentioned in the previous chapter, this research was inspired by the work on online participatory tools such as [Decidim Barcelona] (see Figure 2.1), [Better Reykjavík] and [Parlement & Citoyens].

In these platforms, the participants can contribute to structured discussions around some topic, typically a policy proposal. These particular sites permit deliberation and vary slightly in their input characteristics. They have in common that they allow partici-

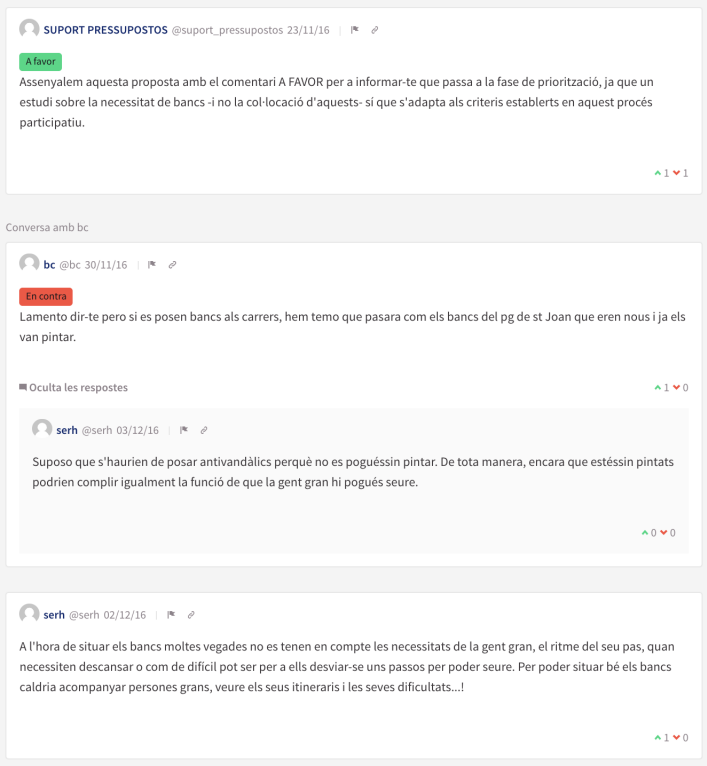



Figure 2.1: Screenshot from [Decidim Barcelona] website showing a fragment of a discussion on putting up benches in public spaces for older people (in Catalan language).

pants to offer arguments for and against a proposal and vote/support for it in the context of a specific public institution such as a city council, etc. However, some of the citizen participation tools behind some of these sites, such as [Decidim Barcelona], or [Consul] for Madrid, have been used in other cities and organisations. Thus, for example, [City of Helsinki] applies an instance of Decidim and [New York City Participatory Budgeting] is based on [Consul]. Other participatory tools have also proliferated outside the context of public institutions. For instance, we find [Consider.it], and [Appgree] whose main focus is scalability —making the systems fit for use by large numbers of participants; Loomio [Jackson and Kuehn 2016], where participants can both comment on proposals, albeit in an unorganised way, and also vote on them; or [Kialo], which organises debates in a structured way. The different approaches offer different institutional input procedures depending on the main view and goal for the participation system.

There is also a long-standing line of work which develops tools to map the structure of arguments on some topic. This line of work draws from a range of sources, summarised in [Shum 2003], and exemplified by [Suthers et al. 1995, Carr 2003, Van Gelder

2003, Reed and Rowe 2004]. The focus of the present research is on drawing the relationships between arguments as a means of helping people to follow and understand better a discussion, and as is pointed out in [Rinner 2006, Benn and Macintosh 2011, Iandoli et al. 2014], the resulting graphical representation of the overall debate can be used to support group decision-making. However, these approaches deal only with *graphical representation* of the arguments —there is no attempt to compute a summary of the discussion nor to create a different form of deliberation. In contrast, our research aims to use the debate as the input to a computational process to obtain an aggregated view of the collective deliberation rather than providing support for the debate itself.

There are other approaches that allow for structured argument-based discussions and aim to compute the outcome of the discussion. One notable body of research here is Klein's work on the Deliberatorium [Klein 2012, Klein and Convertino 2015] which allows for the presentation of arguments and their interactions and aggregates the opinions. However, there is no analysis of the properties of the aggregation with respect to social choice principles.

In contrast to the works discussed so far, where participants in the debate have the task of structuring their arguments into the correct format, Cabrio and Villata [2013] considers extracting arguments from natural language texts and constructing a formal argumentation representation from them. Such a representation can then be summarised as discussed in [Rajendran et al. 2016]. Baroni et al. [2015] discuss how this kind of approach can be combined with the approaches to provide graphical representations for arguments presented above.

The models developed in our research aim to be an abstract formalisation of a debate that can provide the theoretical foundations for a participatory system or represent the information from an already existing participatory system. For example, the TODF model has been applied to real data from debates in [Decidim Barcelona] to perform a comparison experiment with the PAM model [Rodriguez-Aguilar et al. 2016, Serramià et al. 2019]. Each model personalises its input features to handle and represent a multi-agent debate in its own way. We highlight the weaknesses of the works mentioned above and their connections to this research.

- Creating a participatory system or the procedure to manage one, i.e. to control

the functioning of the deliberation, is not in the scope of this research.

- The existing participatory systems are very limited in terms of expressiveness. For example, some of the previous approaches do not allow to issue rebuttals or counterarguments of existing arguments [Klein 2012], only on a proposal, or limit the opinion to be only "agree or disagree" [Consul]. Though similar in some cases, the debate models in this research also organise the information in a structural way, and we aim to increase the expressiveness of the participants —for example, allowing multiple levels of arguments in the discussion or capturing a wider range of opinions.

- Some of the aforementioned related work only aim at capturing or representing a debate, i.e. allowing comments from the users or their opinions on the debate [Decidim Barcelona, Consul,Rinner 2006], but they do not provide means to obtain a collective opinion.

### 2.2.2 Computational argumentation

Computational argumentation [Rahwan and Simari 2009] has a long history within artificial intelligence, going back at least as far as [Fox et al. 1980, McGuire et al. 1981].

First, historically, argumentation theory is concerned with the internal structure of arguments —what arguments are constructed from and how this construction takes place. Early examples of work exploring this problem include those cited above, along with [Loui 1987, Fox et al. 1993, Krause et al. 1995, Parsons 1997], and [Prakken and Sartor 1997]. This line of work has reached its current endpoint with *structured argumentation* systems like logic-based argumentation [Besnard and Hunter 2001], assumption-based argumentation [Dung et al. 2006] and structured argumentation systems such as DeLP [García and Simari 2004], and ASPIC+ [Modgil and Prakken 2013].

Secondly comes a line of work on *abstract argumentation*, begun by Dung [1995], which focuses much less on the internal structure of arguments, and, instead, is mainly concerned with the relationships among arguments. This has led to a large body of work expanding on [Dung 1995], for example [Vreeswijk 1997, Baroni and Giacomin 2009, Modgil and Caminada 2009]. In [Dung 1995], the focus is solely on "attack" relations, where arguments are in conflict, and subsequent work has expanded the scope

to consider "support" relations as well and define bipolar argumentation frameworks [Cayrol and Lagasquie-Schiex 2005, Cayrol and Lagasquie-Schiex 2005, Amgoud et al. 2008]. The work on bipolar argumentation [Cayrol and Lagasquie-Schiex 2005, Amgoud et al. 2008] is concerned with establishing a consistent subset of arguments rather than achieving an aggregation of opinions from several agents.

There is another way to broadly classify work on argumentation into two groups. The first group sees argumentation as a mechanism for extracting consistent points of view from an inconsistent knowledge base. This first approach is again exemplified by [Dung 1995, Caminada 2006]. The other line of work deals with how arguments combine, or accrue, in favour of or against some conclusion. This distinction cuts across the structured/abstract distinction with, for example Baroni and Giacomin [2009] being concerned with consistency in abstract argumentation, and Modgil and Prakken [2013] dealing with consistency in structured argumentation. On the other hand, in [Verheij 1995, Besnard and Hunter 2001, Prakken 2005] they discuss accrual in structured argumentation, while Cayrol and Lagasquie-Schiex [2005] look at accrual in abstract argumentation.

Dung [1995] uses argumentation as a mechanism for a single entity to come to a conclusion. However, as has been pointed out in [Sycara 1990, Walton and Krabbe 1995] and others, argumentation is also a natural mechanism for multiple entities to reach consensus on some topic. As a result, argumentation has been used [Amgoud et al. 2000, McBurney and Parsons 2009, Coste-Marquis et al. 2007, Leite and Martins 2011, Awad et al. 2015] in multi-agent systems as a mechanism for *rational interaction* [McBurney 2002], for a particular meaning of "rational" in the sense that each stage in the interaction is supported by well-founded reasons. For example the work in [Coste-Marquis et al. 2007], which takes as input different sets of arguments and relationships between them and outputs consistent sets of arguments, thus "merging" the input sets. Similarly, Awad et al. [2015] use argumentation and argument labellings [Caminada 2006] to extract a rational view from a number of conflicting opinions.

We find several approaches relating to opinions and how to include evaluations on argumentation frameworks. Baroni et al. [2011] and Awad et al. [2015] use functions that maps arguments into labels —`in`, `out`, or `undec` to represent accepted, rejected or undecided, respectively—, Bench-Capon [2003] and Rodriguez-Aguilar et al. [2016]

use a real-valued function to evaluate the arguments, or [Joseph and Prakken 2009, Dunne et al. 2011] in which they use weights only on relationships, not in the arguments or the nodes connected by the relationships.

Taking into account real-valued opinions in a multi-agent context also creates the need for further comparison and differentiation with work on accrual argumentation. This approach is closely related to social argumentation [Leite and Martins 2011, Rodriguez-Aguilar et al. 2016, Rago and Toni 2017], and previous work on collective argumentation[2] or [Caminada and Pigozzi 2011, Awad et al. 2015] which will be discussed further in next section.

Finally, rationality or consistency is a central feature regarding the throughput and output level of a deliberation procedure. In this research, coherence replaces the commonly used notion of "rationality" to determine the consistency of an opinion [Dung 1995, Baroni et al. 2011]. Joseph and Prakken [2009] also use a notion of coherence in an argumentation framework though they directly define a coherent framework to manage single coherent-based agents and use it to study deliberation about norms.

From all these different approaches and features, next, we list their specific connections to this research.

- Though the similarity of the TODF with bipolar argumentation —which has a "defence" relationship similar to the "support relationship"— the purpose in the TODF is to achieve a collective opinion using aggregation function instead of establishing a consistent subset of arguments [Cayrol and Lagasquie-Schiex 2005, Amgoud et al. 2008][3].

- The use of arguments and attack relationships to represent the information of a discussion [Dung 1995, Leite and Martins 2011, Awad et al. 2015] limits the type of information represented in the debate and the type of opinions that can be issued on it. Simple facts and reasonings, which are implicit in an argument [Besnard and Hunter 2001], cannot be represented differently and therefore can-

[2]Note that this research, and [Caminada and Pigozzi 2011, Awad et al. 2015] has little commonality with the "collective argumentation" of [Bochman 2003], which is concerned with argumentation in which relationships exist between sets of arguments.

[3]We do not use the term "support" for this positive relation between arguments to stress the difference between our work and bipolar argumentation frameworks.

not be evaluated differently. Though the RM and AMAD may share many formal features, neither uses arguments as elements in their formalisation. The RM represents separately the reasonings, represented by relationships, from the elemental sentences they connect, represented as statements. In another way, AMAD uses generic concepts, namely, nodes and relationships, that do not represent anything specific.

- By merging opinion and structural information in a discussion, such as the attack relationship in [Dung 1995] that encodes both the connection and the subjective evaluation of "attack", the expressiveness in a debate is limited. Some participants may agree with an argument but not with it attacking another argument. However, they cannot express this different evaluation since both elements are merged into the shared structure. An approach separating clearly between structural and subjective elements can address this problem. In our research, the RM and AMAD clearly distinguish between structure and opinion.

- Though used in the TODF model in our research, the argument labellings to represent the opinions in a debate [Baroni et al. 2011, Awad et al. 2015] are a very limited form to capture an opinion. In contrast, using continuous real-valued methods to represent the opinions allows for more expressiveness from the participants and, therefore, for a more accurate representation of reality. This approach is followed in the RM.

- Work such as [Dung 1995, Bench-Capon 2003, Caminada 2006, Bench-Capon 2003, Dunne et al. 2011] only aim to represent a single entity predicament and to study on it different consistent semantics. In our research, though we have methods to assess consistency for single agents, we aim to represent multi-agent debates, i.e., discussions with multiple entities.

- The classical notion of rationality [Dung 1995] is not sufficient to characterise the consistency of an opinion because it represents an overly strict consistency. To overcome this problem, we will introduce a novel notion of consistency that aims to characterise consistency in terms of related support (i.e. majority of opinions agreeing with a claim) rather than in avoiding contradictory claims.

- Similar to [Coste-Marquis et al. 2007, Awad et al. 2015, Rodriguez-Aguilar et al. 2016], the TODF and the RM will serve to perform collective aggregation on them, i.e. find functions to extract a collective opinion, if possible coherent, from the individual inputs of the agents.

### 2.2.3 Aggregating opinions

Social choice theory studies how to derive a collective verdict or judgement from a collection of opinions or preference values from members of a group or a society [Gaertner 2009]. Therefore, social choice theory fits entirely into the output level of consideration for a deliberation analysis [Friess and Eilders 2015]. Its main purpose is to study and provide collective decision methods regarding personal and collective satisfaction relative to the outcome [List 2018]. With this aim, social choice theory has extensively explored many ways of aggregating agents' individual preferences [Gaertner 2009]. Since there is a consensus in the literature on several desirable properties that a fair way of aggregating preferences should satisfy (e.g. no single agent can impose their view on the aggregate; if all agents agree, the aggregate must reflect the agreement; and so in [Gaertner 2009]), aggregation functions can be characterised and compared in terms of the desirable properties they satisfy. Notice though, that social choice theory counts on multiple *negative* results, namely impossibility results showing the incompatibility of certain sets of desirable properties (e.g. Arrow's famous impossibility theorem [Arrow and Maskin 2012])[4].

Much of the work in social choice theory has placed little emphasis on the structure of the information over which agents express their preferences. However, there is a growing body of research that takes the subject of the preferences to be arguments in some form or other. Along these lines, the first work was that of [Rahwan and Tohmé 2010], later developed in [Awad et al. 2015], which considered —from one perspective— the very same problem that is tackled here relating the collective reasoning research relating to the TODF and RM, given a topic of discussion and a set of agents expressing individual opinions about the statements made in the discussion, how their opinion can be merged to reach a collectively rational decision? The way that this is tackled in [Awad et al. 2015], using judgement aggregation with abstract argumenta-

[4]Though it is studied in a simple context of deciding one of several alternatives.

tion, as Bodanza et al. [2017] explain, is a version of the "merging" problem mentioned above.

Judgement aggregation is a sub-area of social choice dedicated to the study of how to aggregate the judgements from individuals on a set of propositions in a consistent way [List 2018]. The survey of Awad et al.'s work [Awad et al. 2015] in Section 2.1.2 presents a judgement aggregation work that is closely related to our research, especially with part I.

The same problem of computing a collective labelling was considered in [Caminada and Pigozzi 2011]. As Awad et al. [2017] point out, Awad et al. [2015] and Caminada and Pigozzi [2011] take different approaches. Awad et al. [2015] consider the opinions as votes, combined by taking the plurality for individual arguments, while Caminada and Pigozzi [2011] offer a range of operators that yield a labelling which satisfy the constraints of argumentation semantics[5] while also not disagreeing with the opinions of any participant. Awad et al. [2017] compare the plurality approach with one of the operators in [Caminada and Pigozzi 2011] using human participants.

The recent work in [Chen and Endriss 2019] can be viewed as an extension of the line of work in [Caminada and Pigozzi 2011, Awad et al. 2015]. Like [Caminada and Pigozzi 2011, Awad et al. 2015], Chen and Endriss [2019] propose methods for aggregating a collection of individual argumentation frameworks, each corresponding to a participant in the debate, into a single argumentation framework that appropriately reflects the views of the group as a whole. Chen and Endriss [2019] investigate the properties of the aggregation rules introduced in the paper and employs techniques from social choice theory in the analysis. However, the work of Chen and Endriss [2019] have the aim to analyse aggregation rules in terms of their preservation of *semantic properties* in the argumentation framework, not of social choice properties of the aggregation operators.

Besides using argumentation, in [Endriss and Grandi 2017] they tackle graph aggregation. Given a group of agents and a set of vertices, Endriss and Grandi [2017 consider the aggregation of edges from a set of individual graphs connecting the vertices, each one representing an agent's point of view. Then they consider several aggregation rules that can compute a collective graph, and they study their compatibility with groups of

[5]The constraints here are that the resulting labellings are either admissible or complete [Dung 1995].

social choice properties. In [Endriss and Grandi 2017] they consider aggregation rules that are "collectively rational", i.e. aggregation rules that can output a collective graph preserving some property fulfilled by the individual graphs. Similar to [Chen and Endriss 2019], in [Endriss and Grandi 2017] they focus on the preservation of properties, i.e. the collectively rational aggregation rules.

We notice that from a pure social choice perspective (not a combined argumentation and social choice perspective), it is common in the literature on judgement aggregation and preference aggregation to impose properties on the objects under aggregation so that aggregation operators can guarantee desirable properties. For instance, in the case of distance-based aggregators, the *Kemeny* rule [Endriss and Moulin 2016] only considers consistent judgement sets and hence disregards those which are not. In contrast, premise-based aggregators [Endriss and Moulin 2016] typically make assumptions on the agenda to guarantee consistency and completeness. From the perspective of argumentation, in [Caminada and Pigozzi 2011, Chen and Endriss 2019] they take the approach to start from a set of opinions that are well-formed in an argumentation sense.

A rather different line of work (and one that makes no reference to the work of [Caminada and Pigozzi 2011, Awad et al. 2015, Chen and Endriss 2019]) is that of Rago and Toni [2017]. The QuAD-V framework in [Rago and Toni 2017] allows pro and con arguments (attackers and defenders in our terminology) and agents' votes over arguments (labels). However, the main focus of the work is not on computing a collective opinion but on the agents' contribution with individually rational opinions. With the purpose of computing a collective opinion, Rodriguez-Aguilar et al. [2016] propose a model based on abstract arguments classified as in favour, neutral or against a proposal, and real-valued opinions issued on the arguments. Then, by means of the operator WOWA (Weighted ordered weighted average), they compute a collective opinion for the proposal from the individual opinions given to the arguments.

Regarding the model from which judgement aggregation is performed, [Caminada and Pigozzi 2011, Awad et al. 2015, Rodriguez-Aguilar et al. 2016, Rago and Toni 2017, Chen and Endriss 2019], all deal with abstract arguments and the opinions are expressed about individual arguments but not about the relations between them (these are assumed to be fixed). A different kind of representation is in [Leite and Martins 2011, Rago and Toni 2017], which use real-valued evaluations.

Finally, a main characteristic to consider is *independence* between arguments or statements when computing aggregation. This property is assumed as a fundamental postulate for [Caminada and Pigozzi 2011, Awad et al. 2015, Chen and Endriss 2019]. Conversely, Leite and Martins [2011] do use the dependencies between the related objects of a debate to aggregate the multiple opinions.

The work in [Awad et al. 2015] questions the necessity of assuming independence because of the dependencies between arguments that come already encoded in the form of relationships such as attack. Together with other conditions, independence can imply dictatorship [Lang et al. 2016], and it is considered as not very plausible [Mongin 2008]. This explains why relaxing independence has been subject of much research (e.g. [Mongin 2008, Pigozzi et al. 2008, Dietrich and Mongin 2010, Lang et al. 2016]).

The list below concludes this section, providing connections between our research and the reviewed literature.

- Similarly to the work in [Caminada and Pigozzi 2011, Awad et al. 2015], the aim of the TODF is to compute a collective labelling from the agents' labellings. The RM, though not using labels or arguments, also is used to compute collective opinions from individual inputs.

- In both parts I and II, the different aggregation functions are studied regarding their properties. Differently from [Endriss and Grandi 2017, Chen and Endriss 2019] though, we focus the analysis of the aggregation function in terms of the social choice properties they fulfil. Aside, our research does not tackle preservation of graph properties as in [Endriss and Grandi 2017]. Their central notion of "collectively rational" is related to the preservation of graph properties through aggregation. So, it is not related to the consistency of an opinion, which is a central notion in this dissertation.

  Besides that, the notion of collectively rational from [Endriss and Grandi 2017], that may be confused with a rationality-related concept

  the notion of collectively rational in [Endriss and Grandi 2017] is defined with respect to a *graph* property

- Caminada and Pigozzi [2011] and Chen and Endriss [2019] have an unrealistic assumption about the opinions in a debate. The assumption that a debate starts

from a set of opinions that are well-formed or "legal" in an argumentation sense [Baroni et al. 2011] is an assumption too restrictive for human debates. Human opinions may not be rational in an argumentation-theoretic sense, but not for this reason unqualified for an aggregation process. In this research, we do not assume this restriction.

- Similarly to part I with the TODF, the QuAD-V framework from [Rago and Toni 2017] allows attacks and defences, and the notion of coherence for the TODF coincides with the notion of strict rationality from [Rago and Toni 2017]. However, we do focus on computing a collective opinion.

- Despite the importance of independence as a fundamental property in the judgement aggregation literature [Caminada and Pigozzi 2011, Awad et al. 2015, Chen and Endriss 2019] because of its theoretical value in proving strategy-proofness and strategic manipulation[6], this dissertation regards independence as a too strong property. The contribution of our study on collective reasoning is to introduce several opinion aggregation functions that use the participants' opinions to compute a collective opinion while considering different degrees of *dependencies* between statements or arguments. In fact, when aggregating in the RM, two wide families of aggregation functions explore all possible dependency degrees.

### 2.2.4 Quality measures for a debate

In the literature, quality measures for a debate relate entirely to the communicative throughput level of deliberation [Friess and Eilders 2015], i.e. how the discussion is produced. The key question about quality is how can we ensure a suitable level of quality of the debate that is being constructed? This question is answered by focusing on the measures and methods used so far to analyse a debate and provide an insight of its quality. Undoubtedly, the final goal for ensuring quality standards in a debate is to achieve acceptance of the results by the people involved in the deliberation [Bachtiger et al. 2009, De Vries et al. 2011, Friess and Eilders 2015]. Given the nature of deliberation, quality studies are mostly examined through social and philosophical studies and

[6] If the independence criterion is not satisfied, then the function aggregating judgements is not immune to strategic manipulation [Dietrich and List 2007].

are based on content analysis of a debate.

In such studies, quality is characterised by moral or ethical values that a debate should promote and ensure. Therefore, most works must rely on content analysis and subjective criteria to classify the discourse in debates. One common approach is using coding schemes for each feature under study specifying how to evaluate them rationally, sometimes leading to statistical analysis based on frequency [Steenbergen et al. 2003, Trénel 2004, Stromer-Galley 2007, Manosevitch et al. 2014]. Such approaches, though, rely on human analysis of the linguistic features of deliberation. These approaches differ extremely from the abstract approach taken in this research. The aim here is to model a debate using abstract objects that structure the information, i.e. relationships between objects without text or semantics to interpret. Thus, in this abstract approach, a content analysis methodology to analyse the quality of a debate would be impossible to perform.

Not relying on human examination, nor any other automated content analysis like those using language processing, there are only a few approaches due to the difficulty of evaluating the semantic contents of a debate with quantitative methods objectively. Among them, Gómez et al. [2008] and Gonzalez-Bailon et al. [2010] propose methods to objectively analyse a debate.

Using a graph-like representation of a debate and the participation rates for its arguments, Gonzalez-Bailon et al. [2010] describe four types of debate depending on the measurements of width and depth.

- Type I - High width and high depth. Those debates attracting the attention of a larger number of users and which exhibit a higher frequency in the interaction. They maximise both the width and the depth.

- Type II - High depth and low width. Those that capture high intensity interactions in which only a few participants engage. They exhibit long chains of exchange but only between a few users.

- Type III - Low width and low depth. Where participants are generally not very engaged in a dialogue with other users nor contribute with arguments.

- Type IV - High width and low depth. Debates with more success relating the argument contribution, but agents are not prone to engaging in dialogues that can

form chains.

In [Gómez et al. 2008] there is an application of the well-known $h$-*index* —a measure to characterise the scientific output of a researcher [Hirsch 2005]– to measure the degree of controversy in a discussion. This work defined the $h$-index of a discussion as the maximum number of levels in a discussion, i.e. at most $h$ levels, in which there are at least $h$ comments. Looking at it the other way around, in a discussion with index $h$, for any given $h+1$ levels of the discussion, in at least one of them, there are fewer than $h+1$ comments.

Combining both methods from [Gonzalez-Bailon et al. 2010] and [Gómez et al. 2008], Aragón [2019] presents an analysis about the platform effects when changing the deliberative format of [Menéame.net] website, a political discussion forum. The [Menéame.net] platform changed the presentation and organisation with which the arguments were issued, from lists of arguments to nested relationships between arguments. Aragón [2019] was able to determine that the change of presentation in the platform promoted a general increase of the depths in the deliberation.

At this point, the difference between their view of a debate and the view in this work must be stressed. In [Gómez et al. 2008, Gonzalez-Bailon et al. 2010, Aragón 2019], they considered a debate to be formed by only comments/arguments related to others, without considering different types of relationships nor allowing the participants' opinions on them. In this work, instead, there is the possibility of having different types of relationships, and furthermore, the participants are allowed to express their opinion on the debate structure[7]. These differences are key for defining the quality analysis method for a debate introduced in part III. The method defined there, called Systematic incoherence analysis, takes advantage of the additional features of AMAD, namely different types of relationships and participants' opinions, to use a coherence-based method to point out structural problems within the debate. More specifically, the analysis determines nodes of the debate suffering from an excessive number of incoherent opinions, thus indicating a structural issue. Therefore, it does not rely on content or statistical analysis, such as [Steenbergen et al. 2003, Trénel 2004, Stromer-Galley 2007, Manosevitch et al. 2014], nor does it rely on just a structural analysis like [Gonzalez-Bailon et al. 2010, Aragón 2019].

[7] In different ways depending on which model is considered.

### 2.2.5 Summary of the related work

In the previous sections, we surveyed the work that connects to the research presented in this dissertation. In what follows, we summarise our main observations:

- In Section 2.2.1 we surveyed the existing approaches of online participation platforms and their forms of representing a debate, such as [Decidim Barcelona, Consul] or [Klein 2012]. Compared to the approaches reviewed, our research aim to represent a multi-agent debate, not to manage its creation, to allow more expressiveness to the participants and, more importantly, to obtain a collective opinion.

- Section 2.2.2 reviews argumentation-based research. We pay special attention to [Dung 1995, Caminada 2006] and [Coste-Marquis et al. 2007], which use abstract argumentation as a tool for rational interaction. This connects with our aim in this research to asses consistency, though, unlike [Dung 1995, Caminada 2006], we aim at aggregating opinions instead of searching for consistent sets of arguments.

- Section 2.2.3 reviews social choice theory, the area that studies how to obtain a collective view from a collection of individual preferences or opinions from a group of participants. From this field, we find [Caminada and Pigozzi 2011, Chen and Endriss 2019] and, more importantly, [Awad et al. 2015], that inspired our research to use aggregation functions as a mechanism to compute a collective outcome from a debate and use social choice properties to assess them. Differently, much work like [Arrow et al. 2002, Endriss and Moulin 2016] focus on the possibility and impossibility results regarding sets of social choice properties.

  On another hand, as to the assessment of opinions, in [Caminada and Pigozzi 2011, Awad et al. 2015] they use a notion of consistency (from [Dung 1995]) more restrictive than ours. Along this topic, in [Rago and Toni 2017] we find a notion of consistency similar to our notion of coherence. However, they do not use this notion to study the aggregation of opinions.

- Finally, in Section 2.2.4 we survey the research relating to quality measures for a debate. In this area, aside from social or philosophical approaches, there are few approaches to assess the quality of a debate. Tackling this study objectively, we find [Gómez et al. 2008], and [Gonzalez-Bailon et al. 2010]. However, contrarily

to our approach, they focus on the structure of the debate and do not consider the participants' opinions as we do.

# Part I

# Target oriented discussion framework

# Chapter 3

# Modelling debates through arguments and labellings

The content of this part is organised as follows. This chapter introduces and formally develops the *Target Oriented Discussion Framework (TODF)* model and its related concepts. Chapter 4 presents the aggregation problem, the definition of the aggregation functions, and a formal and computation analysis of these functions. Finally, Chapter 5 discusses the work developed in Part I highlighting its contributions and the conclusions that can be extracted from it.

In this chapter, Section 3.1 introduces the TODF and the coherence notion. Section 3.2 formalises the structure of the TODF to represent a debate composed of arguments, attack and defence relationships and a target argument. Section 3.3 defines the labelling system to represent the participants' opinions over the arguments. Section 3.4 explains and defines the notion of coherence that will be used to assess the consistency of the participants' opinions. Finally, a summary of the contributions of this chapter is detailed in Section 3.5.

## 3.1 Introduction of the Target oriented discussion framework

The model explored in this part, the *Target Oriented Discussion Framework (TODF)*, is inspired by the recent developments of e-participation platforms, such as Decidim

[Decidim Barcelona] and Better Reykjavík [Better Reykjavík], which aim to involve citizens into their policy making. The purpose of the TODF is to formalise a multi-agent debate aimed at deciding on one proposal so that the representation of the debate can be processed to extract the collective opinion. In other words, TODF is a formal model to computationally analyse a multi-agent debate on the acceptance of a proposal to establish what the general opinion is.

The studied scenario is somewhat more general than the systems already existing. Current e-participation systems are limited to either providing arguments lists or a forum-like setting where arguments are structured in a tree-like format. In contrast, the TODF is envisaged to represent a more general discussion allowing a diverse range of studies on it, either on the debate itself by analysing the participants' opinions or on the extraction of a collective opinion by proposing and assessing several aggregation functions.

We provide the contributions of this chapter in the list below.

- *Attack and defence relationships*. The model extends the classical Abstract argumentation framework ([Dung 1995]) used by other work such as [Coste-Marquis et al. 2007, Leite and Martins 2011, Awad et al. 2015] as a basic form of representation of a multi-agent debate. An additional type of relationship between arguments is defined, the defence relationship, to allow the debate participants to support an argument by putting forward a new argument defending it, not only by attacking an attacker. This kind of relationship has been already considered in bipolar argumentation [Amgoud and Cayrol 2002], there called "support", although they focus on the semantics of structured discussions.

- *Target oriented debate*. The TODF, as its name suggests, aims at resolving the acceptance or rejection of a single issue or proposal, called the target, while other approaches [Leite and Martins 2011, Awad et al. 2015] do not restrict the debate to resolve any particular issue. Therefore, the TODF characterises the debate for such a purpose by making the framework evolve around one single topic. Furthermore, this feature will be beneficial when deploying aggregation functions based on dependencies between the arguments in the discussion.

- *A more flexible notion of rationality: coherence*. The notion of rationality in

abstract argumentation, stemming from Dung's original work [Dung 1995], has been widely used to determine the acceptance of arguments in a discussion. Contrastingly, we provide TODF with a notion of coherence, a relaxed version of rationality, that allows us to assess the consistency of an opinion in a more natural and less restrictive way. Additionally, we also use this notion to characterise the outcome of the aggregation functions.

To simplify the formal development of the TODF, we provide below a protocol that a discussion could follow to build a debate in the form of a TODF and obtain a collective decision.

1. One agent puts forward the target of the discussion.

   While any agent can start a discussion by putting forward a target, only one target is allowed per discussion.

2. Any agent may then put forward an argument in favour of, or against, the target or any argument that has already been put forward.

   This process continues until no agent has any further arguments to put forward.

3. Agents then express their opinions about whether the arguments that have been put forward hold or whether those arguments do not hold.

   Agents are not required to have an opinion about whether every argument holds or not — they may or may not express an opinion about any given argument — but any agent can express an opinion about any argument. Each agent, however, may only express one opinion about any given argument.

4. The agents' opinions are then merged to establish a consensus about the status of each argument and, consequently, about the final status of the target as well.

## 3.2 Debate representation: the TODF

Following the approach from [Awad et al. 2015], the TODF model is based on an argumentation setting to represent the structure of a debate by extending the argumentation framework (AF) from Dung 1995 and reviewed in Chapter 2. The AF used in [Awad

et al. 2015] is an *abstract structure* formed by *arguments* and one type of directed relationship, the *attack relationship* to represent the influence of arguments against other arguments. In this context, we may understand the arguments to be abstract objects that represent the comments exchanged in a debate by the participants, such as explanations, reasoning or sentences written to support their claims. Additionally, there are relationships to represent the link existing between the arguments. In what follows, the term *attack* expresses the existence of an "against" relationship between two arguments, as is common in the argumentation literature. The term *defence* expresses the existence of a "for" relationship between two arguments, the opposite of an attack.

The first feature to extend the AF is an additional relationship between arguments, the *defence relationship*, thus allowing the possibility of an argument defending another argument. This way, it is possible for the participants not only to issue rebuttals or counterarguments but also to support the claims of others. The TODF aims to be a more faithful representation of a real debate than the AF, which seemed to be a limited representation. The *discussion framework* is the basic structure capturing these relationships, leading to the final framework representing a debate.

**Definition 3.2.1** (Discussion framework)**.** A *discussion framework* is a triple $DF = \langle \mathcal{A}, \mapsto, \Vdash \rangle$, where $\mathcal{A}$ is a finite set of arguments, and $\mapsto \subseteq \mathcal{A} \times \mathcal{A}$ and $\Vdash \subseteq \mathcal{A} \times \mathcal{A}$ stand for attack and defence relationships that are disjoint, namely $\mapsto \cap \Vdash = \emptyset$. We say that an argument $b \in \mathcal{A}$ attacks another argument $a \in \mathcal{A}$ if, and only if, $b \mapsto a$, and that $b$ defends $a$ if, and only if, $b \Vdash a$.

A discussion framework can be depicted as a graph whose nodes stand for arguments and whose edges represent attack or defence relationships between arguments. Figure 3.1 shows the graphical depiction that will be used to represent the attack and defence relationships.

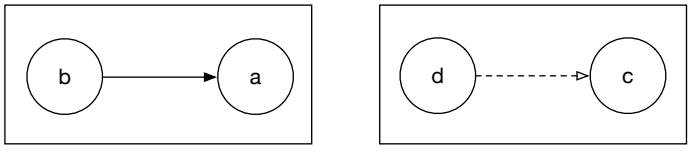



Figure 3.1: Attack (left) and Defence (right) relationships.

Next, we introduce the concept of *descendant* to capture the indirect relationship between two arguments through a sequence of attack and defence relationships.

**Definition 3.2.2** (Descendants)**.** Let $DF = \langle \mathcal{A}, \mapsto, \Vdash \rangle$ be a discussion framework and $a \in \mathcal{A}$ one of its arguments. We say that an argument $b \in \mathcal{A}$ is a *descendant* of $a$ if there is a finite subset of arguments $\{c_1, \cdots, c_r\} \subseteq \mathcal{A}$ such that $b = c_1$, $c_1 R_1 c_2$, $\cdots, c_{r-1} R_{r-1} c_r$, $c_r = a$ and $R_i \in \{\mapsto, \Vdash\}$ for all $1 \leq i < r$.

Given the notion of descendant, next, we formalise the *target oriented discussion framework*.

Aiming the TODF to be a more accurate representation of a typical debate, an additional feature is added to form its structure: the topic under discussion. The model establishes one particular argument, the *target*, as the goal of the discussion (e.g., a norm or proposal), by adding properties to the structure (the target cannot attack nor defend, all the arguments indirectly support or attack the target). The target is the endpoint of all the possible chains of arguments and relationships, thus acting as the root of the discussion. In this way, the TODF is defined as a structure composed of a set of arguments that interrelate between them via attack or defence relationships to discuss a target argument representing the topic to decide.

**Definition 3.2.3** (Target oriented discussion framework)**.** A *target oriented discussion framework*, $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$, is a structure such that $\langle \mathcal{A}, \mapsto, \Vdash \rangle$ is a discussion framework, $\tau \in \mathcal{A}$ is the *target*, and satisfies the following properties:

(i) for every argument $a \in \mathcal{A}$, $a$ is not a descendant of itself,

(ii) for all $a \in \mathcal{A} \setminus \{\tau\}$, $a$ is a descendant of $\tau$.

**Observation 3.2.1.** The previous definitions allow us to identify some properties to characterise the TODF further:

1. *Acyclic structure*. Property (i) ensures that there are no cycles in the TODF, i.e. for any $a \in \mathcal{A}$, we cannot find a chain of relationships (both attack and defence) starting and finishing at $a$. The next two properties follow directly from this fact.

2. *No reflexivity.* No argument can either attack or defend itself. Formally, $\forall a \in \mathcal{A}$, $a \not\mapsto a$ and $a \not\Vdash a$.

3. *No reciprocity.* If an argument $a$ attacks another argument $b$, then $a$ cannot be attacked or defended by $b$, namely $\forall a, b \in \mathcal{A}$, if $a \mapsto b$ then $b \not\mapsto a$ and $b \not\Vdash a$.

Analogously, if an argument $a$ defends another argument $b$, $a$ cannot be defended or attacked by $b$, namely $\forall a, b \in \mathcal{A}$, if $a \Vdash b$ then $b \nVdash a$ and $b \not\mapsto a$.

4. *No target contribution.* The target neither attacks nor defends any other argument, namely for all $a \in \mathcal{A}$, $\tau \not\mapsto a$ and $\tau \nVdash a$. This property distinguishes the special role of the target as the centre of discussion to which attacks and supports are directly or indirectly pointed.

At this point, we introduce a simple example to understand some of the concepts seen until now. We will develop this example throughout this part to illustrate the concepts of each chapter.

**Example 3.2.1** (Formalisation of the neighbours' debate)**.** Suppose Alan, Bart, and Cathy are neighbours, and they aim to reach an agreement on the following norm ($N$):

"Neighbours should take fixed turns at 6 a.m. to clean leaves from the street".

Thus, they discuss the norm by posing the following three different arguments:

$a_1$ ="The schedule is too rigid";

$a_2$ ="6 a.m. is too early"; and

$a_3$ ="Fair task distribution".

Notice that: arguments $a_1$ and $a_2$ are against $N$ whereas $a_3$ is for it; and $a_2$ is in favour of $a_1$, since someone that wakes up later would prefer to change the schedule.

Figure 3.2 depicts the neighbours' $TODF$. The nodes in the graph represent the set of arguments $\mathcal{A} = \{N, a_1, a_2, a_3\}$, where $N$ is the street cleaning norm, and $a_1, a_2, a_3$ are the rest of arguments, respectively. Thus, $N$, the norm under discussion, is taken to be the target $\tau$ in the $TODF$. As to edges, they represent both the attack and defence relationships: $a_1 \mapsto N$, $a_2 \mapsto N$ and $a_2 \Vdash a_1$, $a_3 \Vdash N$, respectively.

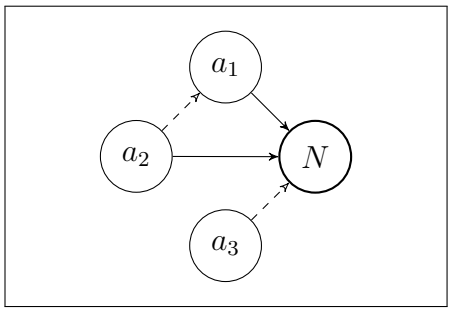



Figure 3.2: Associated $TODF$ graph of neighbours' street cleaning discussion.

Considering the previous definitions and observations, the next proposition follows.

**Proposition 3.2.1.** Let $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ be a target oriented discussion framework and $E = \mapsto \cup \Vdash$. The graph associated to a $TODF$, $G_{TODF} = \langle \mathcal{A}, E \rangle$, is a directed acyclic graph, where $\mathcal{A}$ is the set of nodes and $E$ the edge relationship.

*Proof.* Since $E$ is the union of both attack and defence relationships, that both are directed relationships, the direction of the edges in $G$ is maintained, making it directed. And, by property (i) from definition 3.2.3, there is no path of edges between an argument and itself, which makes the graph $G$ acyclic. □

## 3.3 Introducing agents' opinions: argument labellings

Now that the $TODF$ has been formalised, we introduce the agent's opinions via *argument labellings*. Complementary to the debate structure, the model enables the participants to express their opinions throughout the discussion. Not only the participants are expected to issue arguments, but also participants are expected to give their opinions on the arguments presented by others. In terms of the four-step protocol given above, this corresponds to step 3. Similarly to work in [Awad et al. 2015], a labelling system represents the participants' opinions on each argument.

Here we consider that each agent's opinion corresponds to a *labelling* in the sense of [Caminada 2006, Caminada and Gabbay 2009, Baroni et al. 2011]. That is, a labelling is an assertion about some or all of the arguments in the discussion framework being in one of three states: `in`, meaning that they are accepted by the agent expressing the opinion; `out`, meaning that they are not accepted by the agent expressing the opinion; or `undec` meaning that the agent has not provided an opinion as to whether they are `in` or `out`. Our research dual-purpose the undecided label `undec`, besides expressing "undecided about an argument", it also represents the lack of opinion when a participant has not given one. This way, given the scenario where not all the arguments have been evaluated by each agent, we can still process the debate and finally obtain a collective decision.

This feature is especially relevant in large-scale debates. As seen in [Klein 2012], participants usually give their opinion about those arguments of interest, but we cannot expect them to provide their opinions about all arguments posed within the context of a discussion. Thus, the `undec` label allows working with labels for all the arguments,

even those on which the agents have not provided their opinion. Consequently, from now on, we consider that an argument labelling has every argument attached to a label.

**Definition 3.3.1** (Argument labelling)**.** Let $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ be a target oriented discussion framework. An *argument labelling* for a $TODF$ is a function $L : \mathcal{A} \longrightarrow \{\texttt{in}, \texttt{out}, \texttt{undec}\}$ that maps each argument of $\mathcal{A}$ to one of the labels `in` (accepted), `out` (rejected), or `undec` (undecided). The set of all the possible labellings over a $TODF$ will be noted as $\mathbf{L}(TODF)$.

Considering that we want to tackle aggregation in a multi-agent scenario, next, we introduce the agents involved in the debate and their opinions. We denote as $Ag = \{1, \ldots, n\}$ the set of agents taking part in a $TODF$, and as $L_i$ the labelling encoding the opinion of agent $i \in Ag$. A labelling profile puts together the opinions of all the agents participating in a debate, as follows.

**Definition 3.3.2** (Labelling profile)**.** Let $L_1, \ldots, L_n$ be argument labellings of the agents in $Ag$, where $L_i$ is the argument labelling of agent $i$. A *labelling profile* is a tuple $\mathcal{L} = (L_1, \ldots, L_n) \in \mathbf{L}(TODF)^n = \mathbf{L}(TODF) \times \overset{n}{\ldots} \times \mathbf{L}(TODF)$, where $n = |Ag|$.

This way, the labelling profile holds the opinions of all the participants in the debate. Note that $\mathbf{L}(TODF)^n$ is the set of all possible labelling profiles over a $TODF$.

**Example 3.3.1** (The opinions of the neighbours)**.** We continue the example 3.2.1 from previous section.

Making the arguments and their relations explicit allows Alan, Bart, and Cathy to start sharing their opinions. Thus they can indicate whether they think each argument should be accepted or rejected or whether they have no opinion about it:

- On the one hand, Alan (shown as agent 1 in the first row in Table 3.1) loves getting up late, and so he rejects norm $N$ by assigning an `out` label to the target and accepts arguments $a_1$ and $a_2$ by labelling them as `in`. However, he concedes argument $a_3$ so that it also labels it as `in`.

- On the other hand, Bart (agent 2 in second row in Table 3.1) is used to getting up early and is clearly in favour of norm $N$. Consequently, he accepts both norm $N$ and argument $a_3$ and rejects arguments $a_1$ and $a_2$ which are against $N$.

- Finally, Cathy (agent 3 in third row in Table 3.1) is keen on routines, and thus she accepts norm $N$ and argument $a_3$ and rejects argument $a_1$. Nevertheless, she likes to get up at 7 a.m., so she accepts $a_2$.

| | | Arguments | | | |
|---|---|---|---|---|---|
| | | $N$ | $a_1$ | $a_2$ | $a_3$ |
| Agents | 1 | ✗ | ✓ | ✓ | ✓ |
| | 2 | ✓ | ✗ | ✗ | ✓ |
| | 3 | ✓ | ✗ | ✓ | ✓ |

Table 3.1: Opinions of neighbours in the discussion about street cleaning norm.

Figure 3.3 graphically depicts Alan's, Bart's, and Cathy's labellings (noted as $L_1, L_2, L_3$ respectively), representing their opinions about the $TODF$ from Figure 3.2[1].

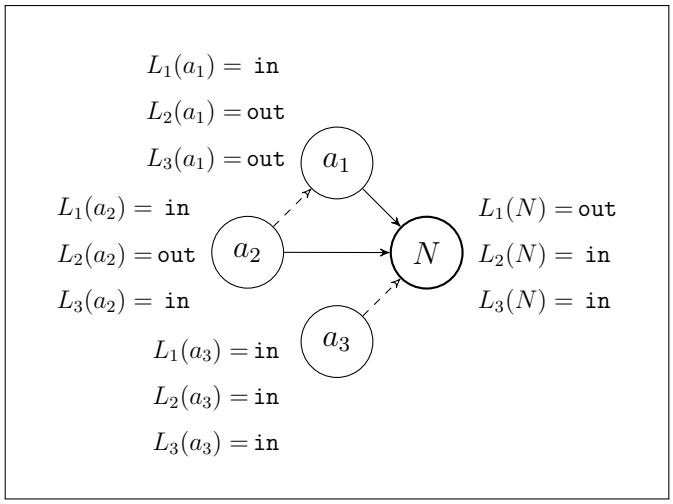



Figure 3.3: Associated $TODF$ graph with labellings of neighbours' street cleaning discussion.

Given this situation, the question that arises, and which this research answers, is: should the neighbours agree to accept this street cleaning norm? Alternatively, how should they aggregate their individual opinions into a single opinion representing their collective discussion? Should they aggregate the opinions on each argument by choosing the opinion that appeared more? Should they consider the existing dependencies

[1]Contrarily to the opinion diffusion models [Grandi et al. 2015], here we do not assume any influence between the agents that could change their opinions in a future stage. The opinions given are supposed to be final and without changes

between the different opinions through the relationships? Should they let the oldest one decide? To rephrase it, what is the best option to decide, a majority rule, a dependency approach or a dictatorship method?

These and further questions originate the topics of the next chapter, but before facing the computation of a decision, we perform a study on the "goodness" of participants' opinions in terms of consistency.

## 3.4 Coherent argument labellings

Mirroring the study of rationality in the argumentation systems performed in [Dung 1993] and used afterwards in [Awad et al. 2015] for a multi-agent scenario, a novel and relaxed notion of consistency is defined in this section.

In [Awad et al. 2015] the authors use the notion of *complete labelling* [Baroni et al. 2011], which starts in [Dung 1995], reflecting the idea that a rational agent will label arguments consistently. A complete labelling requires that an argument is labelled `in` if, and only if, all the arguments which attack it are labelled `out`; and an argument is labelled `out` if, and only if, at least one of the arguments that attack it is labelled `in` (formal definition in Chapter I). Thus an argument can only be accepted (`in`) if all of its attackers are not accepted (`out`) and so on.

The restrictions imposed by a complete labelling do not seem suitable for human participation systems. In a real-world scenario, where humans are the issuers of opinions, this notion strikes as an excessive and rigid way to characterise consistent opinions. Participants are not machines bound to classical logic thinking, nor do they limit their thought process to one type of logic. We, humans, are capable of contradicting ourselves when arguing, capable of supporting contradictory claims at the same time. Thus, it should be preferable to have notions of consistency that at least support some degree of inconsistency and not to identify an opinion as inconsistent when a single opinion on an argument is not supporting the claim it relates to. Instead, we propose *coherence* as a more flexible notion of consistency more appropriate for human debate. Coherence imposes less restrictive conditions for a labelling to be classified as consistent. Moreover, this concept will contribute to the analysis we provide subsequently with an important property to expect from an aggregation function, namely a coherent

collective decision.

Given an argument $a$ in the TODF, we contrast the opinions about the argument, named *direct opinion*, with the opinions about its immediate descendants in the TODF, called *indirect opinion*.The labelling over an argument will be *coherent* if its indirect opinion agrees with its direct opinion, or in other words, when the *majority* of labels in its indirect opinion is in line with its direct label. In the following, we introduce some notations that we will use to define coherence.

First, given an argument $a$, its set of attacking arguments is

$$A(a) = \{b \in \mathcal{A} | b \mapsto a\};$$

and its set of defending arguments is

$$D(a) = \{c \in \mathcal{A} | c \Vdash a\}.$$

Hence, the labels attached to the arguments in $A(a) \cup D(a)$ form the indirect opinion of $a$.

Let $L$ be an argument labelling[2] and $S$ a set of arguments, we denote the number of arguments accepted in $S$ as

$$\texttt{in}_L(S) = |\{b \in S \,|L(b) = \texttt{in}\}|;$$

and the number of rejected arguments as

$$\texttt{out}_L(S) = |\{b \in S \,|L(b) = \texttt{out}\}|.$$

Given this notation, the number of accepted defending arguments of $a$ is $\texttt{in}_L(D(a))$ and the number of rejected defending arguments is $\texttt{out}_L(D(a))$. Similarly, the number of accepted and rejected attacking arguments, respectively, is represented by $\texttt{in}_L(A(a))$ and $\texttt{out}_L(A(a))$. We define next the *positive* and *negative support* of the indirect opinion about an argument as follows. The positive support counts how many indirect opinions support an argument, and the negative support counts the reverse, the number of indirect opinions not supporting an argument.

[2]Notice that we are considering a single argument labelling ($L$), being from an agent or not, not a labelling profile ($\mathcal{L}$).

**Definition 3.4.1** (Positive support)**.** Let $a \in \mathcal{A}$ be an argument and $L$ an argument labelling on $\mathcal{A}$. We define the *positive (pro) support* of $a$ as: $Pro_L(a) = \mathtt{in}_L(D(a)) + \mathtt{out}_L(A(a))$. If $Pro_L(a) = |A(a) \cup D(a)|$ we say that $a$ receives *full positive support* from $L$.

**Definition 3.4.2** (Negative support)**.** Let $a \in \mathcal{A}$ be an argument and $L$ an argument labelling on $\mathcal{A}$. We define the *negative (con) support* of $a$ as: $Con_L(a) = \mathtt{in}_L(A(a)) + \mathtt{out}_L(D(a))$. If $Con_L(a) = |A(a) \cup D(a)|$, $a$ receives *full negative support* from $L$.

We observe that the positive support of an argument adds the numbers of accepted defending arguments with the numbers of rejected attacking ones, whereas the negative support counts the accepted attacking arguments with the rejected defending ones.

The remainder of this section introduces the notion of coherence by combining an argument's positive and negative support. A labelling is *coherent* if, for each argument, the following conditions hold: (1) if an argument is accepted, that is, it is labelled `in`, then its positive support has to be higher than its negative support and (2) if an argument is rejected, that is, it is labelled `out`, then its negative support has to be higher than its positive support. Formally,

**Definition 3.4.3** (Coherence)**.** Given a $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$, a *coherent labelling* is a total function $L : \mathcal{A} \rightarrow \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ such that for all $a \in \mathcal{A}$ with $A(a) \cup D(a) \neq \emptyset$:

(1) if $L(a) = \mathtt{in}$ then $Pro_L(a) \geq Con_L(a)$;

(2) if $L(a) = \mathtt{out}$ then $Pro_L(a) \leq Con_L(a)$.

Note that when $L(a) = \mathtt{undec}$, coherence does restrict the indirect opinion, which is one of the differences with the following definition. We introduce a stronger and more general notion of coherence that considers the possible difference between positive and negative support.

**Definition 3.4.4** (c-Coherence)**.** Let $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ be a target oriented discussion framework. A *c-coherent labelling* for some $c \in \mathbb{N}$ is a total function $L : \mathcal{A} \rightarrow \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ such that for all $a \in \mathcal{A}$ with $A(a) \cup D(a) \neq \emptyset$:

(i) if $L(a) = \mathtt{in}$ then $Pro_L(a) > Con_L(a) + c$;

(ii) if $L(a) = \texttt{out}$ then $Pro_L(a) + c < Con_L(a)$;

(iii) if $L(a) = \texttt{undec}$ then $|Pro_L(a) - Con_L(a)| \leq c$.

The $c$ value from $c$-coherence can be understood as a measure of the consistency of a labelling. The higher the $c$, the better correspondence between the direct and indirect opinions. Note that the weakest form of c-coherence, 0-coherence, is almost the coherence from the previous definition 3.4.3, but they are not equivalent. The next proposition shows where they differ.

**Example 3.4.1.** Carrying on the Neighbour's example, Table 3.2 illustrates the positive and negative support for the arguments involved in their discussion.

| | Labellings | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $L_1$ | | | $L_2$ | | | $L_3$ | | |
| | Positive support | Negative support | Direct label | Positive support | Negative support | Direct label | Positive support | Negative support | Direct label |
| $a_1$ | 1 | 0 | in | 0 | 1 | out | 1 | 0 | out |
| $a_2$ | 0 | 0 | in | 0 | 0 | out | 0 | 0 | in |
| $a_3$ | 0 | 0 | in | 0 | 0 | in | 0 | 0 | in |
| $N$ | 1 | 2 | out | 3 | 0 | in | 2 | 1 | in |
| **Coherence** | ✓ | | | ✓ | | | ✗ | | |

Table 3.2: The coherence of the labellings from the neighbourhood discussion.

**Proposition 3.4.1.** 0-coherence implies coherence, but coherence does not imply 0-coherence.

*Proof.* That 0-coherence implies coherence is straightforward. Given an argument $a$, assume we have a labelling $L$ that is 0-coherent. If $L(a) = \texttt{in}$, then $Pro_L(a) > Con_L(a)$, so it satisfies condition (1) of coherence. Analogously, if $L(a) = \texttt{out}$, then $Pro_L(a) < Con_L(a)$, which satisfies condition (2) of coherence. Therefore, $L$ is also coherent on $a$.

The reverse implication, coherence implies 0-coherence, is false due to the third condition added to the definition of $c$-coherence. For a labelling to be 0-coherent on an argument labelled as undec, it needs the pros and cons of its indirect opinion to be equal, while to be coherent, it does not need equality. We can use as a counterexample

for this case the TODF and labellings depicted in figures A.3(a) and A.3(b) from the Appendix A. In this examples the labellings $L$ and $L'$ are coherent but not $0$-coherent: for $L$, $Pro_L(a) < Con_L(a)$ when $L(a) = \texttt{undec}$ and for $L'$ the reverse. Hence, neither labelling satisfies condition (3) of $0$-coherence. □

$c$-coherence, as the previous proposition shows, is stronger than coherence. For this reason, we can use it to analyse some cases where coherence might not be enough. For example, in the next chapter, we see that one aggregation function needs $0$-coherence labellings to satisfy a social choice property when coherence is not an assumption strong enough.

We will note the class of all the argument labellings of a $TODF$ as $\mathbf{L}(TODF)$, the subclass of coherent argument labellings as $Coh(TODF)$, and the subclass of $c$-coherent argument labellings as $Coh_c(TODF)$ for some $c \in \mathbb{N}$.

As can be seen, the notion of coherence differs from the classical notion of rationality from [Dung 1995, Caminada 2006, Awad et al. 2015]. While in classical rationality an argument and one of its attackers can't be both accepted (`in`), in coherence, they can as long as there are other opinions on other descendants to counterbalance the negative opinion of the attacker.

**Example 3.4.2.** Now we apply the previous definitions to the running example of Figure 3.3. Table 3.2 shows that while labellings $L_1$ and $L_2$ are coherent, $L_3$ is not. $L_3$ is not coherent due to argument $a_1$: while the direct opinion on the argument indicates rejection ($L_3(a_1) = out$), its indirect opinion indicates acceptance (its positive support (1) is greater than its negative support (0)). Only $L_1, L_2$ belong to the subclass of its coherent argument labellings $Coh(TODF)$. Moreover, $L_1$ and $L_2$ are *0-coherent*.

Later on, besides using our definitions of coherence to assess the consistency of the agents' opinions, we will use them to study the quality of the aggregation functions. Coherence is used as a tool for characterising an aggregation function's domain and studying its behaviour. Furthermore, coherence will be used to evaluate the improvement of the aggregation functions when the labelling profiles are restricted to coherent ones. We aim to define aggregation functions giving as output a coherent labelling.

## 3.5 Summary

This chapter formally introduced the TODF and its related concepts.

- Section 3.1 introduced the TODF, the contributions of this part and simple protocol to produce a debate using the TODF.
- In Section 3.2, we established the TODF as a structure composed by arguments, attack and defence relationships and a target argument. The TODF is the structure of a debate aimed at resolving a single issue, the target, and where arguments are issued against or in favour of other arguments.
- In Section 3.3, we introduced the opinion over the structure using an argument labelling. Represented by a labelling, an agent can either accept, reject or even stay undecided about each argument of the debate.
- In Section 3.4, we presented our alternative notion of consistency, coherence, to replace the common notion of rationality adopted by other approaches. Using this concept, we can assess whether an opinion, represented by a labelling, is coherent with respect to the dependencies manifested via relationships. In short terms, a labelling is coherent when the direct opinion of an argument —the label attached to it— is in line with its indirect opinion —the labels relating the descendants of the argument. Furthermore, we introduced the $c$-coherence that can measure the degree of consistency between the direct and indirect opinions.

We have provided answers to research questions RQ-1 and RQ-2 introduced in Chapter 1. More precisely, regarding RQ1, we have proposed a new model to represent a multi-agent debate. Indeed, the TODF is a new model to structure a multi-agent debate to achieve a collective decision about a single issue under discussion (the target). As to research question RQ-2, we have introduced the novel notions of coherence and $c$-coherence to evaluate the consistency of an opinion.

# Chapter 4

# Aggregation and analysis

In this chapter, we build on the definitions of the previous chapter, which established the basic notions of the Target oriented discussion framework (TODF). Section 4.1 introduces the contributions of this chapter. Section 4.2 introduces the aggregation problem and the social choice properties that will be used to analyse an aggregation function; Section 4.3 defines several aggregation functions to reach a collective decision; Section 4.4 analyses and compares the aggregation functions in terms of their social choice properties; and, Section 4.5 offers a complexity evaluation of their implementation. Finally, Section 4.6 offers a summary of the contributions of this chapter.

## 4.1 Introduction

Altogether, the TODF structure plus the labelling systems form a sophisticated picture of the participants' beliefs on the debate. From it, several aggregation functions are defined to extract a decision on each argument in the debate, including the target argument.

Among many features to analyse, we focus on finding how beneficial it is for an aggregation function to consider the relations between the different arguments of a debate, i.e. the dependencies, when producing an outcome. Furthermore, the coherence notion defined for the participants' opinions can be used to analyse the aggregation functions in terms of the consistency of the participants in the debate.

The following are the contributions of our research presented in this chapter:

- *Aggregation functions exploiting dependencies*. As mentioned before, given the

interconnected nature of the arguments in a discussion, our research focuses on using such connections to create aggregation operators that exploit these dependencies differently. From none to full use of dependencies, a set of aggregation functions are defined, comprising different degrees of dependencies.

- *Formal analysis*. For each aggregation function, we analyse its social choice properties to provide guarantees on its behaviour. This analysis allows us to compare the different functions in terms of the social choice properties they fulfil and to point out the one that offers the best trade-off to use in a participation system.

- *Computational analysis*. In addition to the formal analysis, our research studies the computational complexity of the aggregation functions proposed.

## 4.2 The aggregation problem

Although the approach allowing the `undec` label may lead to an undecided outcome, the goal of this work is to reach a collective decision or label on the debate proposal. The following research develops step 4 of the protocol in Section 3.2, applying an aggregation function to decide each argument in the debate. In this section, we cast the goal as a *judgement aggregation problem* [List and Pettit 2002] solved by having a set of agents collectively decide how to label a target oriented argumentation framework. To solve such a problem, we advocate the use of an aggregation function which will provide a label for the target and the remaining arguments, i.e. a *collective labelling*, as an outcome.

First, following the work in [Awad et al. 2015], we formalise the discussion problem by putting together the TODF and the agents, i.e. the debate and the participants. Next, we define the main tool to compute the collective decision on each of the arguments of the debate, the *aggregation function*.

**Definition 4.2.1** (Labelling discussion problem)**.** A *labelling discussion problem* $\mathcal{LDP}$ is a pair $\langle Ag, TODF\rangle$, where $Ag$ is a finite, non-empty set of $n \in \mathbb{N}$ agents, and $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau\rangle$ is a target oriented discussion framework.

**Definition 4.2.2** (Aggregation function)**.** Given a labelling discussion problem $\langle Ag, TODF\rangle$, a function $F : \mathcal{D} \longrightarrow \mathbf{L}(TODF)$, where $\mathcal{D} \subseteq \mathbf{L}(TODF)^n$ $(n = |Ag|)$

is the domain of the function, is called an *aggregation function* for the discussion problem.

An aggregation function $F$ outputs a single argument labelling, called *collective labelling*, from the opinions of the agents contained in a labelling profile. The resulting collective labelling encodes the collective decision over the target and the arguments of the debate. Next, we formalise the goal of the aggregation process, the collective label on the target.

**Definition 4.2.3** (Decision over a target)**.** Given an aggregation function $F$ for a labelling discussion problem $\langle Ag, TODF \rangle$ and a labelling profile $\mathcal{L} \in \mathcal{D}$, the label $F(\mathcal{L})(\tau)$ stands for the *decision over the target* of the $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$.

### 4.2.1 Social choice properties

Social choice theory provides a collection of formal properties that make it possible to assess the behaviour and attributes of the aggregation functions [Dietrich 2007]. Based on [Awad et al. 2015], our research adapted some of these properties to characterise the desirable properties of an aggregation function. In addition to the adapted properties from [Awad et al. 2015], we define novel properties for aggregation functions to address the notion of coherence and consider dependencies among arguments. We recall that our research provides an original way to relax existing assumptions on argument independence in the context of collective decisions.

The first two properties characterise aggregation functions in terms of the domain, i.e. the labelling profiles they can take as input. In particular, the first one adapts from [Awad et al. 2015] the notion of universal domain to characterise aggregation functions defined for any labelling profile; then, we modify this property to consider the aggregation function which is at least defined for coherent labelling profiles.

**Exhaustive domain (ED) [Awad et al. 2015].** An aggregation function $F$ satisfies ED if its domain is $\mathcal{D} = \mathbf{L}(TODF)^n$, namely if the function is defined for all labelling profiles.

**Coherent domain (CD).** An aggregation function $F$ satisfies CD if its domain contains all coherent labelling profiles, namely $Coh(TODF)^n \subseteq \mathcal{D}$.

Moreover, we define *Collective coherence* as a property characterising the aggregation functions that compute coherent collective labellings.

**Collective coherence (CC).** An aggregation function $F$ satisfies CC if for all $\mathcal{L} \in \mathcal{D}$ $F(\mathcal{L}) \in Coh(TODF)$.

We regard the CC property as the most important for an aggregation function. An aggregation function fails at satisfying Collective coherence when it is not able to provide a coherent labelling. This is the case when the collective label (direct opinion) for some argument and its collective indirect opinion are not in line. Such inconsistency may prevent the acceptability of collective decisions [Thagard 2002]. We note that Collective coherence is the counterpart of the *collective rationality* property defined in [Awad et al. 2015]. There, Awad et al. require that the outcome of aggregating labellings to be a *complete* labelling. As argued in Chapter 3, our notion of coherence can be viewed as a relaxation of the notion of complete labelling used in [Awad et al. 2015]. Hence, Collective coherence can be regarded as the relaxation of collective rationality.

In a debate, the opinions of all the agents involved must be considered equally, and Anonymity is the social choice property that captures such requirements.

**Anonymity (A) [Awad et al. 2015].** Let $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile in $\mathcal{D}$, $\sigma$ any permutation over $Ag$, and $\mathcal{L}' = (L_{\sigma(1)}, \ldots, L_{\sigma(n)})$ the labelling profile resulting from applying $\sigma$ over $\mathcal{L}$. An aggregation function $F$ satisfies Anonymity if $F(\mathcal{L}) = F(\mathcal{L}')$.

The Non-dictatorship property requires that no agent overrules the opinions of the rest of the agents. Since Non-dictatorship follows directly from Anonymity, the former is a weaker version of the latter.

**Non-dictatorship (ND) [Awad et al. 2015].** An aggregation function $F$ satisfies ND if there is no agent $i \in Ag$ such that, for every labelling profile $\mathcal{L} = (L_1, \ldots, L_n)$ for which $\mathcal{L} \in \mathcal{D}$, we have $F(\mathcal{L}) = L_i$.

Another important property in the social choice literature is unanimity, which characterises the behaviour of aggregation functions when there is agreement among the agents' opinions. Here, two unanimity properties will consider the relationships between the arguments in the TODF. In particular, the notion of unanimity is adapted

to the TODF to express a desirable property: if all agents share the very same direct opinion on an argument, the collective opinion should be that same one. We name this property *Direct unanimity* to reflect that only the direct opinions are taken into account.

**Direct unanimity (DU).** Let $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile, where $\mathcal{L} \in \mathcal{D}$. An aggregation function $F$ satisfies DU if, for any $a \in \mathcal{A}$ such that $L_i(a) = l$ for all $i \in Ag$, where $l \in \{\texttt{in}, \texttt{out}, \texttt{undec}\}$, then $F(\mathcal{L})(a) = l$ holds.

Additionally, we provide an alternative property relating to unanimity, *Endorsed unanimity*, that considers the dependencies between the arguments and studies the cases when the indirect opinions are unanimous. In particular, this is the counterpart of Direct unanimity for indirect opinions: if there is unanimity in the indirect opinion of an argument, the collective opinion for that argument must be in line with it.

**Endorsed unanimity (EU).** Let $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile such that $\mathcal{L} \in \mathcal{D}$. An aggregation function $F$ satisfies EU if :

(i) For any $a \in A$ such that $a$ counts on full positive support for all $L_i$, then $F(\mathcal{L})(a) = \texttt{in}$;

(ii) For any $a \in A$ such that $a$ counts on full negative support for all $L_i$, then $F(\mathcal{L})(a) = \texttt{out}$.

We notice that neither Direct unanimity implies Endorsed unanimity nor the reverse[1]. Each property characterises unanimity on a different part of the opinion, direct and indirect, respectively.

In addition to unanimity, supportiveness is also considered a complementary property. This property requires that an aggregation function does not label an argument with a label that any agent has not employed.

**Supportiveness (S) [Awad et al. 2015]** An aggregation function $F$ satisfies S if for every argument $a \in A$ and for all labelling profiles $\mathcal{L} = (L_1, \ldots, L_n)$, $\mathcal{L} \in \mathcal{D}$, we can find some agent $i \in Ag$ for which $F(\mathcal{L})(a) = L_i(a)$.

[1] At least, not without adding any assumption on the opinion profiles, maybe some type of coherence

Monotonicity is a property aimed at capturing how the result of an aggregation function changes as opinions, expressed as labellings on arguments, change. In particular, if some of the direct opinions of an argument change to become the same as its collective label, then this collective label should remain the same. Here we adapt Monotonicity and `in-out`-Monotonicity properties from [Awad et al. 2015]. Unlike Monotonicity, `in-out`-Monotonicity (we prefer the name *Binary monotonicity*) only considers the `in` and `out` labels[2].

**(Binary) Monotonicity (M) [Awad et al. 2015]** Let $l \in \{\texttt{in}, \texttt{out}, \texttt{undec}\}$ (resp. for binary $l \in \{\texttt{in}, \texttt{out}\}$) be a label, $a \in \mathcal{A}$ an argument, and $\mathcal{L} = (L_1, \ldots, L_i, \ldots, L_{i+k}, \ldots, L_n)$, $\mathcal{L}' = (L_1, \ldots, L'_i, \ldots, L'_{i+k}, \ldots, L_n)$, $\mathcal{L}, \mathcal{L}' \in \mathcal{D}$, two labelling profiles that only differ on the labellings of agents $i, \ldots, i+k$. An aggregation function $F$ is (resp. binary) monotonic if $L_j(a) \neq l$ while $L'_j(a) = l$ for $j, i \leq j \leq i+k$, then $F(\mathcal{L})(a) = l$ implies that $F(\mathcal{L}')(a) = l$.

Next, the notion of Monotonicity is expanded with two novel properties that, unlike the notion of Monotonicity presented in [Awad et al. 2015], consider the opinions of an argument's descendants. The first of these novel properties, which we call *Familiar monotonicity*[3] (binary if we only consider the labels `in` and `out`), determines that when the direct support for the collective label of an argument increases, the collective label must not change, provided that the opinions on the descendants of the argument do not change either. The need for the latter condition stems from the fact that an argument's collective label might change after the opinions on its descendants are changed.

**(Binary) Familiar Monotonicity (FM).** Let $l \in \{\texttt{in}, \texttt{out}, \texttt{undec}\}$ (respectively for binary $l \in \{\texttt{in}, \texttt{out}\}$) be a label, $a \in \mathcal{A}$ an argument, and $\mathcal{L} = (L_1, \ldots, L_i, \ldots, L_{i+k}, \ldots, L_n)$, $\mathcal{L}' = (L_1, \ldots, L'_i, \ldots, L'_{i+k}, \ldots, L_n)$, $\mathcal{L}, \mathcal{L}' \in \mathcal{D}$, two profiles that only differ on the labellings of agents $i, \ldots, i+k$. An aggregation function $F$ satisfies FM (resp. BFM) if $L_j(a) \neq l$ while $L'_j(a) = l$ and $L_j(b) = L'_j(b)$ for any $j, i \leq j \leq i+k$ and any argument $b$ descendant of $a$, then $F(\mathcal{L})(a) = l$ implies that $F(\mathcal{L}')(a) = l$.

[2]Ensuring the monotonicity in the case of an `undec` as the outcome is particularly hard because replacing a `in` or `out` for the label `undec` can break the tie that was allowing the outcome to be `undec`.

[3]Is called "familiar" because it captures the monotonicity of the descendants — the family — of an argument.

The previous notions of monotonicity are related in the following proposition:

**Proposition 4.2.1.** If an aggregation function satisfies Monotonicity (respectively Binary monotonicity), it also satisfies Familiar monotonicity (respectively Binary familiar monotonicity).

*Proof.* The proof is straightforward because the conditions for Familiar monotonicity (resp. Binary familiar monotonicity) are precisely the same as monotonicity (resp. Binary monotonicity) plus one additional restriction relating to the descendants. The "familiar" versions are more restrictive, hence they hold if the non-familiar versions hold. □

Finally, the notion of Independence [Awad et al. 2015] states that the aggregated label for an argument must depend only on the labels that different agents have for that argument. That is, the aggregated label does not depend on the labels for other arguments. This property is included here for completeness though it is not considered a desirable property due to the general intention of exploiting the dependencies in the aggregation process.

**Independence (I) [Awad et al. 2015].** Let be two labelling profiles $\mathcal{L} = (L_1, \ldots, L_n)$ and $\mathcal{L}' = (L'_1, \ldots, ' L_n)$, such that $\mathcal{L}, \mathcal{L}' \in \mathcal{D}$. For all $a \in \mathcal{A}$, if $\forall i \in \{1, \ldots, n\}$ $L_i(a) = L'_i(a)$ then $F(\mathcal{L})(a) = F(\mathcal{L}')(a)$.

Although this section introduces a set of properties to characterise aggregation functions, it is important to note that not all of them are equally important. We can argue that Collective coherence is the most important property for an aggregation function. If an aggregation function is collectively coherent, it provides a coherent collective labelling, regardless of the coherency of the individual opinions being aggregated. Along with Collective coherence, the aggregation functions should also satisfy the two domain-related properties — Exhaustive domain and Coherent domain (preferably Exhaustive domain to allow broader applicability) — and the usual social choice properties of Anonymity and (if that is not possible) Non-dictatorship. Also, it is considered that an aggregation function should be monotonic, though, given the assumption we take in this research about using dependencies, Familiar Monotonicity (binary or otherwise) is

desirable. Unanimity is also important, but since there could be cases in which unanimity is not satisfied to achieve other important properties such as coherence, then it is less important[4].

Indeed, it may well be the case that aggregation functions do not satisfy the remaining properties, namely Monotonicity (binary or otherwise), Supportiveness and Independence [5]. They are just included here to provide a complete characterisation of aggregation functions.

## 4.3 Aggregation functions for collective decision-making

This section presents the aggregation functions to compute a collective labelling for a labelling discussion problem and, thus, the decision over a target. Therefore, this work tackles the same problem that the work in [Awad et al. 2015]. However, our approach takes an important step beyond since it aims to establish the collective opinion over the arguments without assuming independence between arguments. That is, we do not assume that the aggregated opinion on an argument depends only on the opinions issued on that single argument. Instead, given the relationships between the arguments (attack or defence), it is natural to consider these connections between arguments when computing a collective opinion from it. As a result, following our approach, we define three aggregation functions computing a collective decision, each exploiting a different degree of dependency between the opinions.

The question is how to exploit dependencies, which fundamentally amounts to deciding how much indirect opinion a function uses when computing the aggregated labelling for a given argument. This motivates a family of aggregation functions that exploit indirect opinions in different ways, namely: (i) by giving priority to direct opinions over indirect opinions; (ii) by giving priority to indirect opinions over direct opinions; and (iii) by combining both direct opinions and indirect opinions considering that

[4]For example, in a scenario with two arguments, one attacking the other, if everyone has voted the pair as `in`, we would prefer a function that gives up unanimity and identifies that one of these arguments must be `out` to ensure Collective coherence to a function that ensures unanimity and insists that they must both be `in`. Admittedly, real-life groups, such as the current Republican caucus in the US Congress, would prefer unanimity to Collective coherence in such cases.

[5]Note these properties are related to the argument independence assumption that we are relaxing here.

they are valuable to the same degree. Section 4.4 investigates and compares the social choice properties that each one satisfies to elucidate the aggregation function that best performs. This will allow us to analyse, as part of the discussion in Section 4.4, the benefits and drawbacks, in social choice terms, of the different degrees of exploiting indirect opinions.

Before introducing such functions, for the sake of completeness, the majority rule is also defined as an aggregation function that completely disregards indirect opinion.

Throughout the whole section, we will employ the following notation to represent an argument's direct positive and negative support. Let $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile[6] and $a$ an argument, $\texttt{in}_{\mathcal{L}}(a) = |\{i \in Ag \mid L_i(a) = \texttt{in}\}|$ denotes the *direct positive support* of $a$, whereas $\texttt{out}_{\mathcal{L}}(a) = |\{i \in Ag \mid L_i(a) = \texttt{out}\}|$ denotes its *direct negative support*.

## Disregarding dependencies: a majority rule

The majority function compares the acceptances and rejections received by an argument. The argument will be accepted or rejected depending on whether acceptances or rejections are the majority. Conversely, it will be labelled as undecided if there is a tie. Formally,

**Definition 4.3.1** (Majority function)**.** Given a labelling profile $\mathcal{L}$, the *majority function* for any argument $a$ is defined as:

$$M(\mathcal{L})(a) = \begin{cases} \texttt{in}, & \text{if } in_{\mathcal{L}}(a) > out_{\mathcal{L}}(a) \\ \texttt{out}, & \text{if } in_{\mathcal{L}}(a) < out_{\mathcal{L}}(a) \\ \texttt{undec}, & \text{otherwise} \end{cases}$$

**Example 4.3.1** (Majority rule in the neighbourhood discussion)**.** Following the neighbours' example 3.3.1, we use the majority function to compute the collective labels of each argument. Figure 4.1 represents the collective labelling obtained. For arguments $a_2, a_3$ and $N$ there are more `in`'s than `out`'s, therefore the collective labels using $M$ for such arguments is `in`. For argument $a_1$ is the reverse, there are more `out`'s than `in`'s. Thus, its collective label is `out`.

[6]Notice that, differently from the notation in Section 3.4, here we consider a labelling profile ($\mathcal{L}$), not an argument labelling (L).

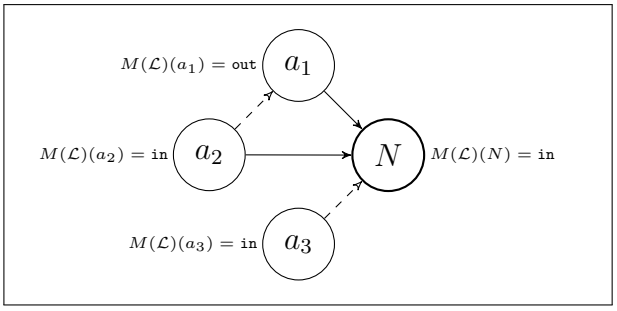



Figure 4.1: Associated TODF graph with the collective labelling (and decision over target $N$) computed by function $M$.

## Exploiting dependencies: prioritising direct opinions

The next function, the *Opinion First function (OF)*, is a variation of the majority function that exploits dependencies but prioritises direct opinions over indirect opinions. Thus, the function first considers direct opinions to obtain an aggregated opinion on an argument. If using direct opinions leads to a *tie* (equal number of acceptances and rejections), then $OF$ uses the collective indirect opinions. Formally,

**Definition 4.3.2** (Opinion first function)**.** Given a labelling profile $\mathcal{L}$, the *opinion first* function for any argument $a$ is calculated as:

$$OF(\mathcal{L})(a) = \begin{cases} \texttt{in}, & \text{if } in_{\mathcal{L}}(a) > out_{\mathcal{L}}(a) \\ \texttt{in}, & \text{if } in_{\mathcal{L}}(a) = out_{\mathcal{L}}(a) \text{ and } Pro_{OF(\mathcal{L})}(a) > Con_{OF(\mathcal{L})}(a) \\ \texttt{out}, & \text{if } in_{\mathcal{L}}(a) < out_{\mathcal{L}}(a) \\ \texttt{out}, & \text{if } in_{\mathcal{L}}(a) = out_{\mathcal{L}}(a) \text{ and } Pro_{OF(\mathcal{L})}(a) < Con_{OF(\mathcal{L})}(a) \\ \texttt{undec}, & \text{otherwise} \end{cases}$$

We observe that, when there is a tie in the direct opinion, the function uses the collective indirect opinions, i.e. the indirect opinions from the collective labelling. This makes the computation of this function a recursive process.

**Example 4.3.2.** Figure 4.2 shows the collective label produced by $OF$ for each argument in the neighbours' example. Since there are no ties for any argument, $OF$ behaves like $M$, and so its collective labelling accepts $a_2, a_3$ and $N$, and rejects $a_1$.

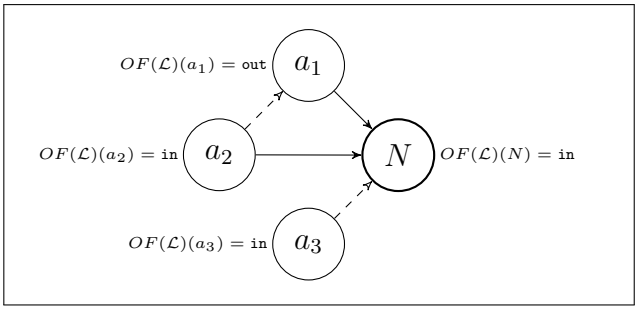



Figure 4.2: Associated TODF graph with the collective labelling (and decision over target $N$) computed by function $OF$.

## Exploiting dependencies: prioritising indirect opinions

As a counterpart of $OF$, the *Support First function (SF)* is introduced, which prioritises indirect opinions over direct opinions. $SF$ considers first the collective indirect opinions to obtain an aggregated opinion on an argument. If using indirect opinions leads to a tie, then $SF$ uses direct opinions. Formally,

**Definition 4.3.3** (Support first function)**.** Given a labelling profile $\mathcal{L}$, the *support first* function for any argument $a$ is calculated as:

$$SF(\mathcal{L})(a) = \begin{cases} \texttt{in}, & \text{if } Pro_{SF(\mathcal{L})}(a) > Con_{SF(\mathcal{L})}(a) \\ \texttt{in}, & \text{if } Pro_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L})}(a) \text{ and } in_{\mathcal{L}}(a) > out_{\mathcal{L}}(a) \\ \texttt{out}, & \text{if } Pro_{SF(\mathcal{L})}(a) < Con_{SF(\mathcal{L})}(a) \\ \texttt{out}, & \text{if } Pro_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L})}(a) \text{ and } in_{\mathcal{L}}(a) < out_{\mathcal{L}}(a) \\ \texttt{undec}, & \text{otherwise} \end{cases}$$

Similar to the $OF$, this function also computes the indirect opinions from the collective labelling. So the computation of this function is also a recursive process.

**Example 4.3.3.** Figure 4.3 shows the collective label produced by $SF$ for each argument in the neighbours' example. We recall that $SF$ considers first indirect opinions. Since arguments $a_2, a_3$ have no descendants, their collective labellings stem from the majority in the direct opinion, and hence, $SF(\mathcal{L})(a_2) = SF(\mathcal{L})(a_3) = \texttt{in}$. As to argument $a_1$, $SF$ first considers the collective labelling of $a_2$, that is $\texttt{in}$, and thus $SF(\mathcal{L})(a_1) = \texttt{in}$. Finally, target $N$ is attacked by arguments $a_1, a_2$, both with collective label $\texttt{in}$, and defended by argument $a_3$ with label $\texttt{in}$. Therefore, the indirect collective support of $N$ is against $N$, and hence $SF$ rejects it, namely $SF(\mathcal{L})(N) = \texttt{out}$.

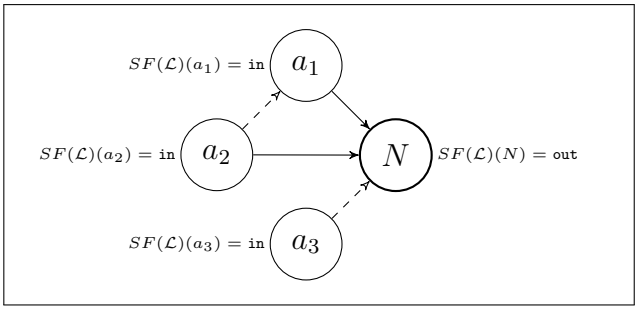



Figure 4.3: Associated TODF graph with the collective labelling (and decision over target $N$) computed by function $SF$.

## Exploiting dependencies: balancing direct and indirect opinions

Finally, after studying functions that prioritise direct opinions, $OF$, or indirect opinions, $SF$, in what follows, we introduce an intermediate function balancing both. Next, we define the *Balanced Function* $BF$, which equally combines direct and indirect support. The following definition might seem a bit complex, but the underlying rationale is simple: for each argument, the balanced function computes both its direct and indirect support to choose the label that best represents both. Formally,

**Definition 4.3.4** (Balanced function)**.** Given a labelling profile $\mathcal{L}$, the *balanced function* over $\mathcal{L}$ for any argument $a$ calculated as:

$$BF(\mathcal{L})(a) = \begin{cases} \texttt{in}, & \text{if } \; IO(\mathcal{L})(a) + DO(\mathcal{L})(a) > 0 \\ \texttt{out}, & \text{if } \; IO(\mathcal{L})(a) + DO(\mathcal{L})(a) < 0 \\ \texttt{undec}, & \text{if } \; IO(\mathcal{L})(a) + DO(\mathcal{L})(a) = 0 \end{cases}$$

where the functions $IO$ (indirect opinion) and $DO$ (direct opinion) are defined as:

$$IO(\mathcal{L})(a) = \begin{cases} 1, & \text{if } \; Pro_{BF(\mathcal{L})}(a) > Con_{BF(\mathcal{L})}(a) \\ 0, & \text{if } \; Pro_{BF(\mathcal{L})}(a) = Con_{BF(\mathcal{L})}(a) \\ -1, & \text{if } \; Pro_{BF(\mathcal{L})}(a) < Con_{BF(\mathcal{L})}(a) \end{cases}$$

$$DO(\mathcal{L})(a) = \begin{cases} 1, & \text{if } \; \texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a) \\ 0, & \text{if } \; \texttt{in}_{\mathcal{L}}(a) = \texttt{out}_{\mathcal{L}}(a) \\ -1, & \text{if } \; \texttt{in}_{\mathcal{L}}(a) < \texttt{out}_{\mathcal{L}}(a) \end{cases}$$

As can be seen, first, the balanced function compares the positive opinion against the negative opinion for both the direct and indirect opinion. Then, from the types of

impact extracted from direct and indirect opinion (positive, negative or neither), the function provides the outcome that best represents their combination.

**Example 4.3.4** (Neighbourhood discussion)**.** Figure 4.3 shows the aggregated opinion and the decision over the target for the neighbourhood example obtained by the balanced aggregation function. As shown in the picture, neighbours collectively accept arguments $a_2$ and $a_3$, whereas argument $a_1$ is undecided. Finally, the decision over the target is to accept it (i.e., $BF(\mathcal{L})(N) = \texttt{in}$), so the norm is accepted.

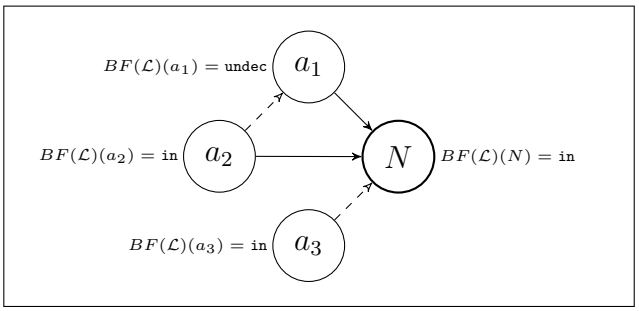



Figure 4.4: Associated TODF graph with the collective labelling (and decision over target $N$) computed by function $BF$.

We observe that, as the $OF$ and $SF$, $BF$ uses the collective labels already computed by the function itself on the indirect opinion, using a recursive process to compute an outcome. Later, Section 4.5 analyses this feature in more detail.

## 4.4 Formal analysis of the aggregation functions

This section compares the results from analysing the aggregation functions presented in the previous sections. It is considered in social choice theory that the problem of merging opinions about arguments is an instance of collective decision-making. As such, following [Awad et al. 2015], the quality of each aggregation function is analysed using the social choice properties defined in Section 7.2.2.

The purpose of these properties is to characterise the behaviour of the aggregation functions in different situations. Thus, this analysis results in a complete characterisation of each aggregation function and an in-depth comparison between them. To make a comprehensive comparison, the results of the three novel aggregation functions that use dependencies are contrasted against an adapted version of the "majority-rule" function,

which disregards these dependencies completely. As a result, we will be able to see the trade-offs of using more or fewer dependencies and determine the best aggregation function to solve the collective decision problem, i.e. the function that performs better in terms of its social choice properties.

The formal propositions and proofs of the results shown next are in Appendix A. Table 4.1, shows the social choice properties fulfilled by each aggregation function. Table 4.1 splits social choice properties into two groups: those identified as desirable in Section 4.2.1 when exploiting dependencies between arguments and those that are not so relevant to the purposes of this research but are typically referred to in the social choice literature. Afterwards, we use these results to compare the aggregation functions. As stated in the previous sections, we recall that the most important property is *Collective coherence* because it ensures a minimum of consistency at the outcome of an aggregation function. Table 4.1 shows the relationship between this property and the "degree" of indirect opinion involved in the decision-making represented by each aggregation function.

From left to right in Table 4.1: $M$ disregards indirect opinions; $OF$ prioritises direct opinions over indirect opinions; $BF$ equally considers direct and indirect opinions; and finally, $SF$ prioritises indirect opinions.

By analysing Table 4.1 one can make several interesting observations regarding (i) the positive and negative effects of exploiting dependencies; (ii) the aggregation function that offers the best compromise between exploiting direct and indirect opinions; and (iii) the positive and negative effects of introducing uncertainty using the `undec` label.

First, we observe that exploiting dependencies yields two main benefits:

- *Indirect opinions must be exploited at least as much as direct opinions to ensure Collective coherence.* Indeed, either disregarding indirect opinions or prioritising direct opinions over indirect opinions is not enough to achieve Collective coherence. Thus, neither the majority rule, which disregards indirect opinions, nor $OF$, which prioritises direct opinions, satisfy Collective coherence. However, both $SF$, which prioritises indirect opinions, and $BF$, which equally combines indirect and direct opinions, do satisfy Collective coherence.

| Desirable properties | M | OF | **BF** | SF |
|---|---|---|---|---|
| **Collective coherence** | ✗ | ✗ | ✓ | ✓ |
| Exhaustive domain | ✓ | ✓ | ✓ | ✓ |
| Coherent domain | ✓ | ✓ | ✓ | ✓ |
| Anonymity | ✓ | ✓ | ✓ | ✓ |
| Non-dictatorship | ✓ | ✓ | ✓ | ✓ |
| Endorsed unanimity | (✓) | (✓) | (✓) | ✗ |
| Unanimity | ✓ | ✓ | ✗ | ✗ |
| Binary familiar monotonicity | ✓ | ✓ | ✓ | ✓ |
| Familiar monotonicity | ✗ | ✗ | ✗ | ✗ |
| Other properties | | | | |
| Binary monotonicity | ✓ | ✓ | ✗ | ✗ |
| Monotonicity | ✗ | ✗ | ✗ | ✗ |
| Supportiveness | ✗ | ✗ | ✗ | ✗ |
| Independence | ✓ | ✗ | ✗ | ✗ |

Table 4.1: Comparison of social choice properties fulfilled by the aggregation functions —Majority function ($M$), Opinion first function ($OF$), Balanced function ($BF$) and Support first function ($SF$). Symbol code: ✓means fully satisfied; (✓) represents satisfied under some assumptions; and ✗ stands for unsatisfied.

- *Exploiting indirect opinions preserves Anonymity*. The aggregation functions exploiting indirect opinions ($OF$, $BF$, and $SF$) only consider different amounts of positive and negative opinions while disregarding the sources of opinions. Hence, because of such general treatment of agents' opinions, they all satisfy the Anonymity and Non-dictatorship properties.

Second, despite obtaining major benefits, particularly in terms of satisfaction of Collective coherence, we pay the price for exploiting dependencies, namely:

- *The exploitation of indirect opinions impacts the unanimity and monotonicity properties.* We notice that as we move from left to right in Table 4.1, the Direct unanimity and Monotonicity are satisfied by fewer functions, clearly relating the satisfaction of the properties with the level of indirect opinion involved: the

higher the importance of indirect opinions in an aggregation function, the fewer satisfied unanimity and monotonicity properties.

- *Exploiting dependencies between arguments impedes Independence.* As expected, even a little involvement of indirect opinions in the decision-making prevents the fulfilment of this property, and, therefore, the fulfilment of other social choice properties (not considered here) stronger than Independence. However, note that we do not regard this observation as a negative result. Recall from the discussion in Section 4.2.1 that Awad et al. [2015] consider the necessity of independence questionable (because of the existing dependencies of attack between arguments), while the literature, [Mongin 2008, Pigozzi et al. 2008, Dietrich and Mongin 2010, Lang et al. 2016], and the approach taken here consider Independence as too strong and not very plausible.

At this point, given the pros and cons above-mentioned regarding the exploitation of dependencies, we are ready to identify what is considered the best-in-class-aggregation operator:

- *BF provides the best trade-off between exploiting direct and indirect opinions.* On the one hand, $OF$ does not satisfy Collective coherence, but it satisfies both types of unanimity and the weaker versions of monotonicity. On the other hand, while $SF$ satisfies Collective coherence and Binary familiar monotonicity, it fails at satisfying any unanimity property or the more restrictive monotonicity properties. $BF$ sits between $OF$ and $SF$. Regarding the other properties such as Exhaustive domain, Anonymity or Familiar monotonicity, all the aggregation functions satisfy them to the same degree, making no difference when deciding the best one.

Last but not least, we turn our attention to the benefits and drawbacks of introducing the `undec` label to cope with uncertainty:

- *The introduction of uncertainty favours the general treatment of any labelling profile.* Implicitly, in the approach taken here, we use the `undec` label to obtain an outcome even in those cases where there is no clear decision over an argument. The introduction of the `undec` label helps to process ties (when the number of acceptances equals the number of rejections) that would occur without this label.

Not allowing the `undec` label would restrict the domain of the aggregation functions and hamper decision-making despite the existence of valid opinions. Note that this is not the case for all the aggregation functions that have been introduced since they all fulfil the Exhaustive and Coherent domain properties.

- *The introduction of uncertainty negatively affects monotonicity properties.* Using `undec` label may cause the lack of a "positive" or "negative" decision regarding the acceptance of an argument. This fact directly impacts the satisfaction of the monotonicity properties, hence the need for weaker versions such as Binary monotonicity and Binary familiar monotonicity. This behaviour is made explicit when proving the satisfaction of the monotonicity properties by the aggregation functions (see appendix A).

Besides the general observations above, Table 4.1 is also valuable to help individually analyse each of the aggregation functions introduced in this chapter:

- Previously is shown that $M$ does not satisfy the most important property, Collective coherence. Therefore, $M$ does not ensure the coherence of the labelling obtained as a collective decision; consequently, it might contain irrational sets of argument labellings. Despite this, the majority function satisfies many of the other desired social choice properties without any restrictions, except the Endorsed unanimity property, which is restricted to 0-coherent profiles. We also observe that while $M$ satisfies restricted versions of monotonicity properties, it does not satisfy their non-restricted versions due to the existence of the `undec` label. Finally, the non-exploitation of dependencies guarantees the satisfaction of the Independence property, but the possible `undec` label resulting from a tie prevents the satisfaction of supportiveness.

- At first sight, the $OF$ function satisfies several desirable social choice properties without restrictions, except for Endorsed unanimity, which requires coherent labelling profiles to hold. Nonetheless, $OF$ still fails, just like $M$, to satisfy Collective coherence; hence, it cannot ensure a minimum of rationality for the collective decision. Finally, $OF$ does not satisfy the non-binary monotonicity properties, and, as a result of exploiting indirect opinions, it loses the Independence property. To summarise, the way $OF$ exploits indirect opinions is not

enough, as observed above in the general analysis.

- $SF$ increases the relevance of indirect opinions when computing a collective labelling. On the one hand, this entails the satisfaction of Collective coherence. On the other hand, this negatively impacts the satisfaction of monotonicity since $SF$ loses Binary monotonicity compared to $OF$. Furthermore, $SF$ is further from satisfying Endorsed unanimity than $OF$, since $SF$ does not satisfy Endorsed unanimity even when we impose some coherence on agents' labellings. Finally, likewise $M$ and $OF$, $SF$ also satisfies exhaustive and coherence domain, Anonymity, Non-dictatorship, and Binary familiar monotonicity.

- $BF$ provides a trade-off between $OF$ and $SF$. First, $BF$ satisfies most of the desirable properties identified in Section 4.2.1, including Collective coherence. However, note that $BF$ only satisfies Endorsed unanimity in the case of 0-coherent labellings (though it is better than $SF$ that does not satisfy any of the unanimity properties). Second, $BF$ does not satisfy properties such as Direct unanimity and supportiveness but recall that the first one was considered the least desirable property and that the second one was not even considered desirable.

## 4.5 Computational complexity

In addition to the theoretical assessment of the aggregation functions, we analyse their computational complexity here. To do so, this section provides an algorithm for computing the collective decision on a target using an aggregation function.

Let us consider a $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$, with a target $\tau$ for which we aim at computing a collective label. Thus, we require a profile $\mathcal{L}$ reflecting the opinions of the agents involved in the discussion and a function to aggregate the opinions in the profile (be it either $SF$, $OF$, or $BF$).

We observe that, according to proposition 3.2.1, the graph associated to the TODF is a DAG and that three of the aggregation functions compute the collective labelling recursively. Therefore, the computation of the collective labels for the arguments in the discussion framework can be performed while traversing its associated graph, henceforth referred to as $G_{TODF}$. This is where we can resort to topological sorting [Kahn

1962] to perform graph traversal. Thus, we propose to compute the collective labels for the arguments and the target of a $TODF$ after a topological sorting algorithm. From this follows that the computation of the collective label for the target, which is not recursive thanks to the topological sorting, is linear in the number of nodes (arguments) plus edges (attack and defence relationships) in the associated graph of the discussion framework, asymptotically, namely $O(|\mathcal{A}| + |\mapsto| + |\Vdash|)$.

Function COMPUTECOLLECTIVEDECISION in Algorithm 1 calculates the collective decision for the target $\tau$ of a $TODF$ from a profile $\mathcal{L}$ and aggregation function $F$. The algorithm uses $PendingArguments$ to store the arguments in the correct order so as to compute the collective labellings recursively. An argument is stored in $PendingArguments$, at first if it has no descendants (line 2), and after, within the algorithm (line 9), when all its descendants already have a collective labelling computed. An argument is deleted (line 5) from $PendingArgument$ when its collective labelling is being computed using the aggregation function $F$ (lines 6)[7].

**Algorithm 1** Compute collective decision

```
1:  function COMPUTECOLLECTIVEDECISION(G_TODF, τ, F, L)
2:      PendingArguments ← arguments with no descendants
3:      while PendingArguments is not empty do
4:          b ← remove argument from PendingArguments
5:          F(L)(b) ← compute collective label for b
6:          for each argument c such that (b, c) is an edge of G_TODF do
7:              remove (b, c) from G_TODF
8:              if there are no incoming edges for argument c then
9:                  add argument c to PendingArguments
10:     return F(L)(τ)                    ▷ Collective label for target τ
```

Due to the low complexity of the algorithm when computing an aggregated labelling, we can consider that this approach has the potential to be used in practice, even in large-scale scenarios.

[7] In [Ganzer et al. 2017] we provide an implementation of Algorithm 1 available to the general public together with all the functions introduced in Section 4.3.

## 4.6 Summary

This chapter was dedicated to the aggregation of opinions on the TODF.

- Section 4.2 defined the aggregation problem and the purpose of an aggregation function, to compute a collective labelling. We presented the social choice properties we used to analyse the behaviour of an aggregation function. Among them, we regard Collective coherence as the most desirable property.

- In Section 4.3 we introduced specific aggregation functions: the Majority function, computing the label supported by the majority of agents; the Opinion first function, prioritising the direct opinion over the indirect opinion of the agents to compute a collective label; the Support first function, the reverse, prioritising indirect opinion over direct opinion; and, the Balanced function, which considers equivalent both direct or indirect opinion.

- Section 4.4 contains the formal analysis of the aggregation functions mentioned above regarding their social choice properties. Among many, a few conclusions of the analysis are: the Balanced function is the function satisfying more desirable properties; taking into account the dependencies of the debate helps to satisfy Collective coherence and impacts negatively unanimity and monotonicity properties.

- Finally, Section 4.5 offers a complexity assessment of the implementation of the aggregation functions, showing that an implementation can face large-scale scenarios.

Among the different research questions proposed in Chapter 1 we answered positively the following:

RQ-3 We presented three new aggregation functions that exploit dependencies between arguments to find a collective opinion. These functions to aggregate opinions respond to a novel approach with respect to the state of the art.

RQ-4 Considering dependencies in the aggregation made us revisit the social choice properties regarding coherence, unanimity and monotonicity. This resulted in the introduction of three new social choice properties that take into account the

dependencies between arguments in a debate: Collective coherence, Endorsed unanimity and Familiar monotonicity.

RQ-5 By analysing the aggregation functions, we have seen that exploiting dependencies can benefit the fulfilment of several social choice properties while being a disadvantage to others. In particular, the analysis shows that the Balanced function is the function offering the best trade-off.

# Chapter 5

# Conclusion and discussion

In the context of participatory systems and collective reasoning, this part I advances the state of the art on several lines of research. The developments of this part, from the previous chapters, are comprehensively reviewed in the Section 5.1. A discussion about the TODF model is provided in Section 5.2, exposing the advantages and disadvantages of this line of work, the latter being the motivation for the research in part II.

Throughout this part, some assumptions have been guiding the investigation. First, we assumed that the agents in a multi-agent debate might not participate rationally, and hence agents are allowed to express their opinions in an inconsistent or contradictory manner. Second, uncertainty is inherent in any debate, be it because participants signal that they do not have a clear opinion about certain topics or because they do not express any opinion at all. Third, a collective opinion aggregation function must not ignore the dependencies existing in a debate. Therefore, we have exploited the connections between the arguments and their opinions in a debate. Finally, aggregation functions must aim to be minimally consistent (coherent), thus computing coherent collective opinions.

## 5.1 Contributions

The work shown in this part for solving the collective decision problem tackles different areas providing the contributions summarised below.

***The Target oriented discussion framework* (TODF), a model to support multi-agent debates.** Answering positively the research question RQ-1, the TODF aims to

articulate a multi-party discussion around a given topic or proposal. The TODF enables to represent a multi-agent debate whose participants express arguments for and against the proposal and other arguments being discussed. The structure of the debate is captured by means of *abstract arguments*, the *attack* and *defence* relationships, and a *target* argument. Moreover, the opinions of the agents, issued on the arguments, are captured by means of *labelling systems* (likewise in [Awad et al. 2015]), which allow representing qualitative opinions. The labels `in`,`out` and `undec` represent the acceptance, rejection, or indecision, respectively, on an argument under evaluation.

***Coherence*, a new method to characterise the consistency of an opinion.** Going beyond the typical notion of rationality used in argumentation theory ([Dung 1995, Caminada 2006, Awad et al. 2015]), coherence relaxes the constraints that an opinion must satisfy to be considered sufficiently reasonable. Namely, an agent's opinion is coherent when at each argument, the opinion issued directly on it is in line with the opinions on its related arguments. In addition, we provide the means to specify the degree of coherence that an opinion may hold. Thus, we answer the research question RQ-2 positively.

**Novel social choice properties to assess better the aggregation functions exploiting dependencies.** These new properties enrich the collection of social choice properties employed to characterise the behaviour of an aggregation function. Thus, besides Anonymity, Unanimity, Monotonicity, and further properties borrowed from the social choice literature, we introduce three novel properties:

- *Collective coherence*, the most desired property, describes the behaviour of those aggregation functions that compute a coherent collective opinion from the agents' opinions;
- *Familiar monotonicity* characterises monotonicity on the aggregation functions taking into account the dependencies to compute an outcome;
- Lastly, *Endorsed unanimity*, distinguishing the aggregation functions that respect a unanimous indirect opinion relating to an argument.

The previous properties consider the dependencies of a debate to characterise the aggregation functions. Therefore, we can answer yes to research question RQ-4.

**Novel aggregation functions exploiting dependencies to compute collective opinions.** As argued in Section 4.3, since we are dealing with interrelated arguments, we did not consider independence as a reasonable characteristic for the aggregation process. As a result of this approach, three different aggregation functions have been introduced, each capable of exploiting dependencies between arguments in different manners. Thus offering a positive answer to research question RQ-3, we defined:

- the *Opinion first function* prioritising direct opinions over indirect opinions;
- the *Support first function* prioritising indirect opinions; and,
- the *Balanced function* that balances both types of opinions.

All the aggregation functions, including the Majority function for completeness purposes, have been analysed and compared by means of the social choice properties. Two of those functions, Support first and Balanced functions, guarantee the *Collective coherence* of the outcome, even when *all* sort of labelling profiles are allowed as input, namely even those in which participants' opinions are not individually coherent.

Furthermore, the analysis in Chapter 4 produced insight into the design of an aggregation function and the price paid to ensure coherence and handle uncertainty. Although introducing the `undec` label favours the general treatment of any labelling profile, it negatively affects monotonicity properties. We have seen that disregarding indirect opinions or prioritising direct opinions over indirect opinions is not enough to achieve Collective coherence. However, the necessary exploitation of indirect opinions to obtain Collective coherence comes at a price: the higher the importance of indirect opinions in an aggregation function, the fewer the number of unanimity and monotonicity-related properties that can be satisfied. In the end, the Balanced Function proves to be the best trade-off between exploiting direct and indirect opinions.

Overall, the contributions with the TODF make headway in bringing together the fields of argumentation and computational social choice.

## 5.2 Strengths and limitations

This section explores several aspects about the design of the TODF model that can become advantages or disadvantages for different contexts. Furthermore, this review

motivates the development of the Relational model, described in part II.

### 5.2.1 Strengths

The representation of the TODF in a graph-like form sometimes can be a handicap in terms of visibility of the information, especially when the participation rate is high, for it may be a challenge to show large amounts of information appropriately to a participant. Despite this limitation, having similar characteristics to existing participation systems makes the TODF a model with familiar and more accessible features to interact with. The separate representation of information as arguments related via relationships, attack or defence, and the participants' opinions expressed using three labels on it already has some similarities to with some existing implementations such as [Decidim Barcelona, Kialo] or [Klein 2012]. Even current social networks [Facebook, Twitter, Reddit], which are not designed to decide collectively, have similar features, such as allowing the participants to issue arguments (generally referred to as comments) on other arguments or allowing them to give an opinion —like and dislike— on arguments made by others. In more detail, we list the strengths the TODF shares with the platforms mentioned above.

- **Target of the discussion**. In the same way that arguments in the TODF can be related to a proposal or other arguments, in a platform, the users can issue comments on a post or on other comments issued by others, forming threads.

- **Positive and negative relationships.** In participation systems, usually, we can relate each comment with a positive or negative impact, comparable to having an attack or a defence, though they are not classified as such, and only the interpretation of the text can lead to this subjective classification. Thus, adapting the platform to the TODF format or translating the content into a TODF representation can be relatively easy.

- **Opinion in labels**. Moreover, the like or dislike options for the posts and comments, *typically available* available in platforms such as [Facebook, Twitter, Reddit], relate closely to the `in` and `out` labels from the labelling system to express the participants' opinions on the content.

- **Target-oriented.** Besides that, the target-oriented characterisation of the TODF, which does not allow cycles, bodes well for a participation system purpose. In social networks or participatory platforms, the comments that arguments would represent grow in a single direction, rooted in a single post or issue, i.e. a target-oriented and acyclic relationship.

### 5.2.2 Limitations

The distribution that the TODF captures from the information of a debate can also have some drawbacks, especially regarding expressiveness. We list in detail the possible limitations of the TODF.

- **Limited relationships** The two-fold relationship endowed in the TODF, though based on the natural form in which humans discuss either against or supporting other claims, somehow limits the participants' expressiveness. Neutral information, neither attacking nor defending, which sometimes could be useful to provide factual or other types of information, cannot be issued in the discussion. To address this issue, the participants could label an argument as undecidable, but that remains an individual option rather than a general classification in the debate.

- **Merging of structural and subjective elements**. The structure of the debate contains opinion, i.e. the debate is connected via attack and defence relationships, which are subjective connections (opinion) that all the participants must share. That could be a problem if some participants do not agree with the type of relationship chosen between two arguments. Once an argument has been issued against or supporting another, all the participants are bound to share that type of relationship, thus sharing the subjective classification endowed in that relationship and unable to classify it from their point of view. Although some participants might agree on the classification, others might not and would classify the relationship the other way around. To illustrate this problem consider the next case extending example 3.2.1.

  **Example 5.2.1.** Let's assume that together with Alan, Bart, and Cathy, another person, Diana, also participates in the discussion. Diana, who works eight hours a day and knows that Alan works half of what she does, thinks the task should be

distributed according to the amount of free time each person possesses. Therefore, she considers that the task is equally distributed among the neighbours to be an argument against the proposal, i.e. $a_3$ ="Fair task distribution" should be related by an attack relationship to the proposal. But it is not the case since, in this example, $a_3$ is defending the proposal. Even if it were the case, the other neighbours would disagree on the attack relationship. To conclude, in this scenario, the TODF cannot represent all the points of view that the participants need.

Such a problem is caused by combining two different forms of content, information and opinion, into one single object. The attack or defence relationships merge the structural purpose of a connection with the qualitative evaluation. This merging induces a type of opinion in the debate structure susceptible of not being shared among the individuals. The Relational Model (RM), in next part II, owes to this issue one of its significant characteristics.

- **Limited opinion on the arguments**. Continuing with the topic of individual opinions, the TODF allows the opinion on one argument to be one of three options: `in`, `out` and `undec`. This approach offers a basic classification that would be straightforward to use by the users of a platform. The resemblance that these labellings have with the current rating options in social platforms (like, dislike, thumbs up, etc.) makes it easy for users to grasp the function of such labels and use them properly and unambiguously. However, its simplicity also reduces too much the agents' expressiveness. A human discussion usually requires more complexity to evaluate an argument than a label with only three optional values. The options only allow full acceptance, full rejection or indecision halfway between the previous two. There is no room for middle terms or subtlety, which could easily be needed many times. The labelling system does not offer an accurate and precise way to express an opinion.

  One form to solve this issue could be by allowing more than three values or even a continuous range of values for an opinion. This latter case is explored in the RM in part II, which uses real-valued functions to represent the participants' opinions.

- **Limited opinion on the relationships** Additionally, another limitation is related to the opinion allowed. The TODF does not consider other types of opinions be-

sides acceptance and rejection of the arguments. The human thinking process is more complex than that. Besides identifying good/bad situations, we can think in other ways, for example, in terms of truth. In the TODF, we have no way to differentiate between the opinion of a participant who likes an argument but thinks it is not true and the opinion of another participant believing that the argument is true but does not like what it states. For both cases, the label would be `out`. Such a problem is addressed in the RM by allowing two evaluation methods to express two different types of evaluations. An evaluation in terms of truth and an evaluation in terms of goodness.

Overall, the TODF offers a basic but valuable approach for studying a multi-agent debate. Although its limitations regarding expressiveness, the TODF features pay off by providing a manageable representation to study a debate and an easy and understandable functioning that eases its applicability in real scenarios.

# Part II

# Relational model

# Chapter 6

# Modelling a debate

This part II introduces the Relational model (RM) and explains how to perform collective reasoning on it. Chapter 6 presents the RM, its formal characteristics, and defines the notion of coherence. Chapter 7 provides social choice properties, defines the aggregation functions and analyses them in terms of the properties they fulfil. Finally, Chapter 8 discusses the research of this part.

The content of this chapter is organised as follows. Next section, introduces the RM and its contributions. Section 6.2 introduces the ideas to understand the RM and the formalisation of the model, and Section 6.3 characterises coherence in the RM. Finally, Section 6.4 summarises the research of this chapter.

## 6.1 Introduction

The RM is a new model to represent a multi-agent debate that, compared to the TODF, aims for a more general characterisation of a debate. Similarly to part I, the RM is the basis from which to study aggregation functions on a debate having a set of proposals for which a decision has to be made. This research aims to be a more sophisticated approach to capture discussions and process collective decisions from them.

The main purpose of the RM is to allow more expressiveness than the existing approaches permit, both in terms of the structure of the debate and the opinion expressed by the participants. To do so, the abstract argumentation approach is abandoned to advance toward a new and neutral structure for a debate, allowing more comprehensive and rich opinions from the participants. Furthermore, the RM accommodates a new

notion of coherence and provides large families of aggregation functions that exploit the dependencies in the debate to form a collective opinion from the participants.

The contributions of our research presented in this chapter are listed below.

- *No subjective structure.* Collective decision-making frameworks based on traditional argumentation models usually take as starting point a fixed argumentation structure that only models attack relationships between arguments or a combination of attack and support/defence relationships between arguments. These frameworks may allow participants to express opinions about the different arguments included in the debate [Leite and Martins 2011, Awad et al. 2015, Ganzer-Ripoll et al. 2019]. The fixed nature of the argumentation structure, even if the participants define it, represents a significant drawback for e-participation systems. Adopting a fixed argumentation structure using only attacks, or attacks plus supports/defences, limits what participants are allowed to express. For instance, a fixed attack relationship between two arguments might be problematic for some participants who disagree with the classification of the relationship as an attack and would have defined the same relationship as a defence. To solve this problem, the RM uses relationships that only represent reasoning connections between elements of the debate are, not subjectively classified attacks or defences. In other words, the reason why there is a connection between different elements, a *reasoning*, is captured formally by a relationship. The subjective assessment is then applied individually by each participant, not in terms of attack or defence, but in terms of acceptability of the connections. Thus, the structure of the debate is focused on organising relevant information, not on expressing the subjective opinion of the participants.

- *Going beyond abstract argumentation.* Several approaches [Coste-Marquis et al. 2007, Leite and Martins 2011, Awad et al. 2015] make use of abstract argumentation frameworks [Dung 1995], or some variations of them, as in [Ganzer-Ripoll et al. 2019], to represent the elements of a debate. In such frameworks, whole arguments are atomic elements. In work on argumentation, this limitation has led to work on "structured" [Modgil and Prakken 2013] or "rule-based" [García and Simari 2004] argumentation which constructs arguments out of lower level components like facts and rules. Along these lines, this work takes a similar but

more general approach. The debate is constructed using two types of abstract elements: statements, which represent sentences without any reasoning within (so they are not abstract arguments) and the relationships between statements, each one representing an existing reasoning connecting the statements[1]. To clarify, the RM represents formally a debate that takes place in the world, where we can find the words said by the participants. In this "real" debate, for each connection between statements, there is an actual reasoning (i.e. an explanation of why those statements are connected) which we formally represent in the RM using an abstract relationship. As statements and relationships represent different types of information, we allow them to be evaluated differently by the participants.

- *Compound and real-valued opinions.* Previous work on argumentation-based approaches has only allowed participants in a debate to express opinions about either the arguments [Leite and Martins 2011, Ganzer-Ripoll et al. 2019], or about the relationships between arguments [Dunne et al. 2011][2]. Here in the RM, the participants can provide opinions on both statements and the relationships between them. Opinions about relationships capture participants' acceptance, or otherwise, of the reasoning represented by the relationship, and opinions about statements reflect participants' satisfaction with the statement itself.

  Furthermore, the opinions about relationships and statements are expressed using continuous values rather than the discrete values of [Awad et al. 2015, Ganzer-Ripoll et al. 2019] (`in`, `out`, `undec`). This feature allows the participants to express their opinions in a wider range of values, making the approach more flexible and applicable for those participation systems that can provide degrees of agreement or disagreement[3].

- *A more flexible notion of coherence.* Previous work on determining collective opinions makes use of a notion of "rationality" in which an opinion is either

[1] We could relate the statements with axioms, premises or conclusions, and the relationships by the rules or demonstration steps that lead from premises to conclusions, just as in structured argumentation.

[2] [Dunne et al. 2011] is not about combining collective opinions on relationships between arguments. Still, it provides the groundwork for such a system by studying argumentation where the relationships between arguments have different weights.

[3] It should be noted though that most existing e-participation system just allow users to express agreement or disagreement.

determined to be acceptable or not acceptable (where "acceptable" has different interpretations but reflects the constraints on distributions of opinions across statements to be consistent) [Awad et al. 2015, Rago and Toni 2017, Ganzer-Ripoll et al. 2019]. Though, the common notion of rationality from [Dung 1995] is somewhat limiting. Since the opinions originate from human participants, and humans are not always consistent in their views, insisting on this rigid form can lead to misrepresenting valuable information. Hence, our research proposes a less restrictive notion of rationality, called "coherence", which evolved from the one defined for the TODF to assess the degree to which an opinion is coherent, be it from an agent or from the collective aggregation.

A debate can be built in many ways following the RM. Here, we propose a simple procedurethat would produce an RM and serve as a support guide for future sections. There are four main steps:

**Step 1 –** *Start debate.* A set of statements for which we intend to obtain a collective opinion are chosen as targets of the RM.

**Step 2 –** *Extend debate.* Participants are then allowed to put forward relationships that will represent relevant reasoning. A relationship may either be put forward in conjunction with new statements –so it connects from existing statements to a new statement– or it may connect existing statements. This step continues until no participant wishes to add a further relationship.

**Step 3 –** *Input opinions.* Participants express their opinions on the relationships and statements in the RM by providing *subjective evaluations* on them. The evaluation of a statement expresses preferences over them, while the evaluation of a relationship expresses agreement, or otherwise, with the reasoning represented by the relationship.

**Step 4 –** *Obtain collective opinion.* The participants' opinions are merged to establish a consensus view of each statement and relationship in the framework. Hence, the collective opinion of the target statements is also obtained.

## 6.2 Formalising the Relational model

This section introduces the main features of the RM, a model designed to represent a collective debate where participants discuss a proposal by putting forward additional information relevant to the discussion and giving their opinions about it. The RM comprises two main parts: the structural part, which organises the *content* of a debate, and the interpretative part, which represents the participants' *opinions* in a debate. Figure 6.1 depicts the distinction between the two components of the RM, as well as their basic elements introduced next.

### 6.2.1 Structure

The RM will have two main elements to capture the structure of a debate: *statements* and the *relationships* among them. The *statements* represent plain sentences that describe facts such as, "Increase of house prices in the neighbourhood". Participants will afterwards be able to express their opinion about the desirability or undesirability of the sentence (Section 6.2.2). A *relationship* represents a reasoning connecting statements. Each relationship connects a set of source statements to some destination statement. In this model, a reasoning is understood as "the reason that connects" different statements, which can be logical, in the mathematical sense, or not. Each reasoning then may or may not be acceptable to some degree to the participants of the debate, which will provide an opinion about it by evaluating the relationship representing them.

We first introduce the formal notion of a *relational framework* to capture the relationships between statements. The notion of relationship will consider a non-empty set of source statements related to a destination statement. In general, connecting a set of (source) statements to a (destination) statement indicates that the source statements support inferring the destination statement. However, the framework here is agnostic about the form that the support and the inference mechanism takes[4]. For instance, in the example in Figure 6.3, statements $s_2$ and $s_3$ support inferring $s_4$. Formally:

**Definition 6.2.1** (Relational framework)**.** A *relational framework* $RF$ is a pair $\langle \mathcal{S}, \mathcal{R} \rangle$, where $\mathcal{S}$ is a set of statements and $\mathcal{R} \subset \mathcal{P}(\mathcal{S}) \times \mathcal{S} \times \mathbb{N}$ is a relation fulfilling:

- *Contingency*: for any $s \in \mathcal{S}$ and any $c \in \mathbb{N}$, $(\emptyset, s, c) \notin \mathcal{R}$.

[4]We do not restrict the reasonings to be from a particular logic.

- *Acyclicity*. There are no cycles in $\mathcal{R}$, that is, there is no subset of relationships $\{(\Sigma_0, s_1, c_1), \ldots, (\Sigma_{k-1}, s_k, c_k)\} \subset \mathcal{R}$ such that $s_i \in \Sigma_i$, $i \in \{1, \ldots, k-1\}$, and $s_k \in \Sigma_0$.

- *Indirect connection*: for any two statements $s_1, s_2 \in \mathcal{S}$:

  - $s_1$ is connected with $s_2$, i.e., there is a reasoning $(\Sigma, s, c) \in R$ such that $s_1 \in \Sigma$ and $s = s_2$ or $s_2 \in \Sigma$ and $s = s_1$; or,

  - there exists a statement $s \in \mathcal{S}$ that *connects indirectly* $s_1$ and $s_2$. That is, there is a path connecting $s_1$ with $s$ and another path connecting $s_2$ with $s$[5].

**Observation 6.2.1.** We note that since the relation $\mathcal{R}$ is acyclic, it follows that $\mathcal{R}$ is neither *reflexive* ($\forall s \in \mathcal{S}$, $(\Sigma \cup \{s\}, s, c) \notin \mathcal{R}$) nor *symmetric* ($\forall s_1, s_2 \in \mathcal{S}$, if $(\Sigma_1 \cup \{s_1\}, s_2, c_2) \in \mathcal{R}$ then $(\Sigma_2 \cup \{s_2\}, s_1, c_1) \notin \mathcal{R}$). Note that there are no conditions regarding the transitivity of the relationship $\mathcal{R}$.

We also notice a natural number in the relationship to differentiate several relationships that can exist between the same set of statements $\Sigma$ and $s$. From a practical perspective, this allows us to represent that alternative relationships can be assigned between to the very same statements (as shown in Figure 6.3, where target $\tau$ is related to statement $s_1$ through relationships $r_1$ and $r_6$).

Moreover, the acyclic property relates to two purposes of the RM: first, to represent a chronological addition of the statements (and reasoning) in a debate, i.e., the order in which the participants have put forward the statements during the discussion; second, as the following definition will show, to allow the structure to be directed towards the targets of the debate that represent the topic under discussion.

Commonly, debates discuss a particular subject or proposal and may even consider a set of proposals. In the RM, we consider the *target* (of the debate) to be a set of statements forming the proposal. The distinguishing feature of these statements, the target of the debate, is that none of them can be a destination statement in any relationship because they are the initiators of the debate. Thus, the target acts as the root of the structure that captures the debate. This structure composed of the statements, relationships and target is called a *Directed Relational Framework*, or DRF for short.

[5]A path connecting a statement $a_0$ with a statement $a_k$, or vice versa, is a sequence of relationships $r_1, r_2, \ldots, r_k$ such that $r_i = (\Sigma_{i-1} \cup \{a_{i-1}\}, a_i, c_i) \in \mathcal{R},\ 1 \leq i \leq k$.

**Definition 6.2.2** (Directed relational framework)**.** A *directed relational framework* (DRF) is a tuple $\langle \mathcal{S}, \mathcal{R}, T \rangle$ such that:

(i) $\langle \mathcal{S}, \mathcal{R} \rangle$ is a relational framework;

(ii) $T \subset \mathcal{S}$ is a set of target statements;

(iii) Target statements in $T$ can only be the source of relationships, namely for any relationship $(\Sigma, s, c) \in \mathcal{R}$, $s \notin T$; and

(iv) All non-target statements are connected to targets so that for any statement $s \in S$, $s \notin T$, there is a path $\{(\Sigma_0, s_1, c_1), \ldots, (\Sigma_{k-1}, s, c_k)\} \subset \mathcal{R}$ such that $T \cap \Sigma_0 \neq \emptyset$.

**Observation 6.2.2.** We note that a DRF is constrained to be a *connected acyclic* graph, albeit one that can have several targets. This reflects the idea that, since a DRF represents a single debate, every statement in that debate should have some connection to the rest of the debate.

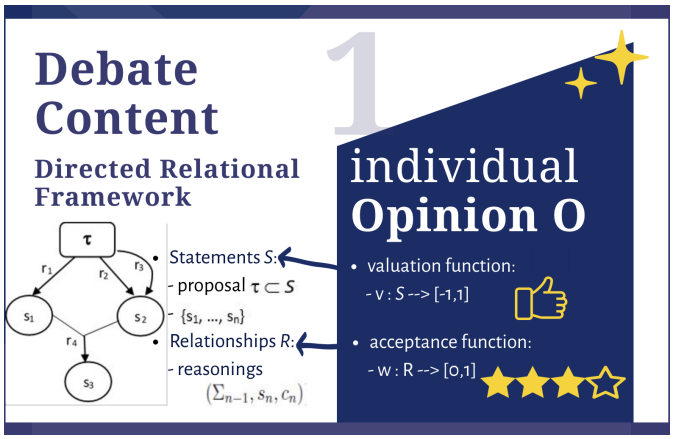



Figure 6.1: The basic elements of the RM.

In what follows, we propose an example to clarify the concepts presented so far. This example will be extended and developed throughout this part II to illustrate the concepts being introduced.

**Example 6.2.1.** Let us consider a debate about building a sports centre. Let's assume, first, that we have only two statements and one reasoning between them. The proposal $\tau$, "Construction of a sports centre in a particular location in the neighbourhood", is the target being discussed. In this setting, we consider the reasoning "The construction

of the sports centre will imply the demolition of existing buildings that give historical relevance to the neighbourhood" is represented by the relationship $r_1$ that connects the target $\tau$ to a new statement $s_1$ "Destruction of the character of the neighbourhood", as depicted in Figure 6.2.

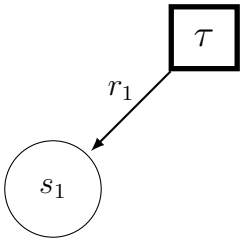



Figure 6.2: Graphical representation of the relationship between proposal $\tau$ (building a sports centre) and statement $s_1$ (destroying the neighbourhood's character).

Now that the basic elements have been illustrated, we enlarge the debate by presenting further statements and reasonings that could appear in the discussion. This addition will lead to a bigger $DRF$[6].

The statements and reasonings of the debate are listed in tables 6.1 and 6.2 respectively. Finally, Figure 6.3 depicts the connections between statements through relationships. We note that $r_4$ is a relationship connecting three statements[7].

[6]In fact, the example could grow larger with more statements and relationships than the ones we include here. However, including more information would make the example difficult to manage, obstructing its purpose.

[7]Notice that one participant could introduce an extra relationship from $\tau$ to $s_5$, representing the reasoning "A new community centre will give more relevance to the neighbourhood, which will increase the house price", which is not the sum of $r_2, r_3$ and $r_4$, but represents a whole new way to connect $\tau$ to $s_5$. This shows that transitivity can exist in the model, connecting statements from non-consecutive levels of the debate via a single relationship, but should not be understood as the combined reasoning formed by the reasoning steps in between.

| Statement | Description |
|---|---|
| $\tau$ | Construction of a sports centre in a particular location in the neighbourhood |
| $s_1$ | Destruction of the neighbourhood character |
| $s_2$ | Attraction of more affluent residents to the neighbourhood |
| $s_3$ | Attraction of new business to the neighbourhood |
| $s_4$ | Crime reduction in the neighbourhood |
| $s_5$ | Property values raise in the neighbourhood |

Table 6.1: Statements for the sports centre example.

| Relationship | Reasoning | Connection |
|---|---|---|
| $r_1$ | The construction of the sport centre will imply the demolition of existing buildings which now give historical relevance to the neighbourhood | $\tau$ to $s_1$ |
| $r_2$ | The new sport centre will make the neighbourhood more attractive for wealthy residents because they are more interested in leisure activities | $\tau$ to $s_2$ |
| $r_3$ | A new community centre will attract more businesses to the neighbourhood. | $\tau$ to $s_3$ |
| $r_4$ | Having richer residents and more businesses will increase security around the neighbourhood and therefore reduce criminal activities. | $\{s_2, s_3\}$ to $s_4$ |
| $r_5$ | The reduction of crime will increase the price of the houses in the neighbourhood | $s_4$ to $s_5$ |
| $r_6$ | The architecture of a new building may clash with other buildings, changing the character of the neighbourhood | $\tau$ to $s_1$ |

Table 6.2: Reasoning for the sports centre example.

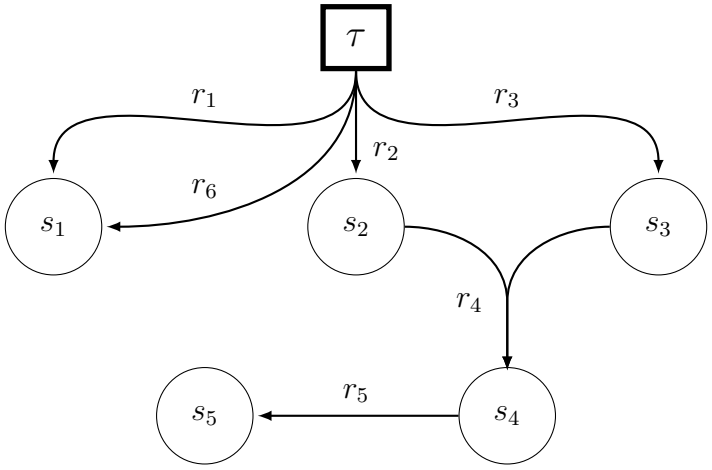



Figure 6.3: DRF for the sports centre example. The nodes represent the statements, and the arcs between nodes represent the relationships between the statements.

The next section defines how the opinions of the participants will be included in the RM.

### 6.2.2 Opinions

This section addresses the formalisation of the opinions put forward by participants in a debate. The participants' opinions are encoded in the form of functions that assign values to the elements of the debate's structure. The *valuation function* assigns values to statements and the *acceptance function* assigns values to the relationships.

A participant's opinion is twofold to obtain the two types of subjectivity involved in the debate. The first, relating to the valuation function, provides a participant with a way to express their opinion about the statements in the debate. The second, relating to the acceptance function, allows a participant to express the truth they see in each reasoning, captured by a relationship, that is included in the debate. Thus, the desirability or undesirability that each participant feels about each statement of the debate is represented by a positive or negative value assigned with the valuation function, and the agreement that each agent relates to reasoning between the statements is represented by an acceptance value. If a participant does not provide any opinion on the statement or relationship, the default value for each function is 0.

Formally, the functions forming the opinion are the following.

**Definition 6.2.3** (Valuation function)**.** Given a DRF $\langle \mathcal{S}, \mathcal{R}, T \rangle$, a *valuation function* $v : \mathcal{S} \longrightarrow I$ maps each statement to a value in $I = [-a, a]$, $a \in \mathbb{R}^+$.

Given a statement $s \in \mathcal{S}$: we say that $s$ has *full positive valuation* if $v(s) = a$; $s$ has *full negative valuation* if $v(s) = -a$ ; and if $v(s) = 0$ then $s$ has *neutral valuation*.

**Definition 6.2.4** (Acceptance function)**.** Given $a$ a positive number and a DRF $\langle \mathcal{S}, \mathcal{R}, T \rangle$, an *acceptance function* $w$ maps each relationship to a value in $I^+$, $w : \mathcal{R} \longrightarrow I^+ = [0, a]$.

Given a relationship $r \in \mathcal{R}$ and an acceptance function $w$, we will refer to the value $w(r)$ as the *acceptance degree* of $r$. If $w(r) = a$ then the acceptance function expresses *full agreement* with the relationship, whereas if $w(r) = 0$ it expresses *full disagreement*. Without loss of generality, henceforth the value of $a$ is set to 1, and hence $I = [-1, 1]$ and $I^+ = [0, 1]$.

Now, we introduce the participants' opinions on the running example 6.2.1 started in the previous section.

**Example 6.2.2.** Given the statements and relationships of the $DRF$, the participants can express their opinions by providing a value to every statement and relationship of the debate. The values assigned to the statements will represent how they feel about each respective statement, and each degree given to a relationship will indicate how much the participant believes in the truth of the reasoning behind it.

Next, the three participants taking part in the debate are introduced[8]. For clarity and simplicity, we show the rationale behind their opinions for only statements $\tau, s_1$ and the reasoning $r_1$. We recall that $\tau$ represents "Construction of a sports centre in a particular location in the neighbourhood", $r_1$ represents the reasoning "The construction of the sports centre will imply the demolition of existing buildings that give historical relevance to the neighbourhood", and $s_1$ represents "Destruction of the character of the neighbourhood". The complete set of opinions will be directly represented on Figures 6.4 and 6.5 for valuations and acceptances respectively.

We now consider three people taking part in the debate described above with the following opinions:

[8]For the sake of simplicity, the example is limited to three participants and a small number of statements and relationships.

- **Participant 1** is a middle-aged woman with a family who lives in the neighbourhood where the sports centre is supposed to be built. She values the proposal ($\tau$) very positively because her family practices sports. Although she has been living in the neighbourhood for a long time, she has no childhood memories of the neighbourhood and doesn't care too much for its historical character, so she values the statement $s_1$ as neutral. Finally, she assigns a small acceptance value to relationship $r_1$ because she acknowledges that building the sports centre will imply the demolition of some buildings, but she doesn't believe that the character of the neighbourhood will be too affected by such a loss.

- **Participant 2** is a retired older man. He has always lived in the neighbourhood and would like to preserve its unique features. Hence, he values the statement $s_1$ negatively and also the target $\tau$ because he is not interested in sports and would prefer another kind of public building instead. He entirely agrees with the relationship $r_1$, since he considers that the buildings in the proposed location for the new sports centre are important to the neighbourhood and would be demolished if the centre is built.

- **Participant 3** is a young postgraduate student who does not practice any regular sports that can be held in the planned sports centre, so values $\tau$ quite negatively. He agrees with the relationship $r_1$ because he acknowledges that the existing buildings could be catalogued as of special architectural interest. However, he is neutral concerning $s_1$ because he does not care about preserving the neighbourhood's character.

From the description of the agents' opinions, now we would translate their preferences into valuations and acceptances[9]. Thus, for instance, agent one is "highly positive" on the target $\tau$ ($v_1(\tau) = 0.9$), but neutral regarding statement $s_1$ ($v_1(s_1) = 0$). Furthermore, agent one considers that the plausibility of relationship $r_1$ is "little" ($w_1(r_1) = 0.2$). The DRF graphically represented in the previous Figure 6.3, is completed with the opinions in Figures 6.4 and 6.5 that show respectively the valuation functions and acceptance functions of agents 1, 2, and 3: $v_1$, $v_2$ and $v_3$, while $w_1$, $w_2$

[9]There are multiple ways to convert a description to actual values, in this example we chose one. In a real debate the participants would provide the actual values.

and $w_3$ encode agents' acceptances of the relationships.

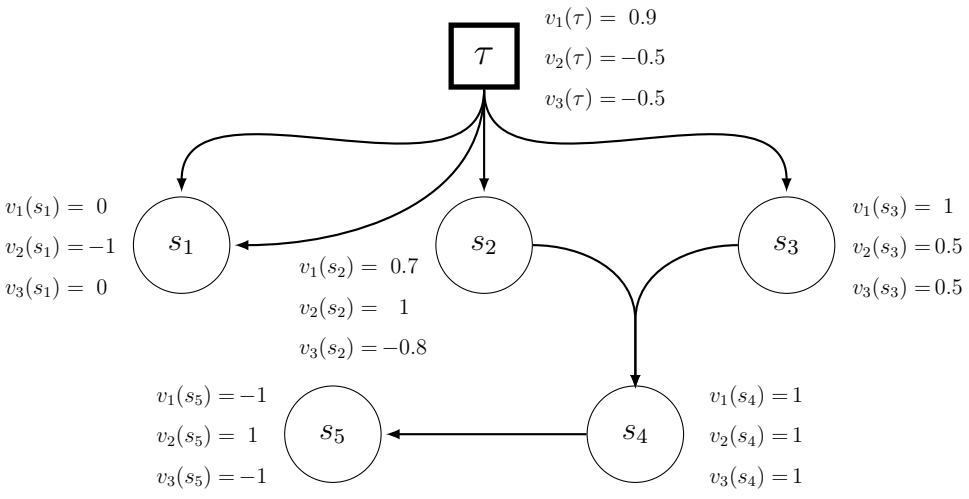



Figure 6.4: Agents' valuation functions. $v_1$, $v_2$ and $v_3$ encode agents' valuations on statements for agents 1, 2 and 3, respectively.

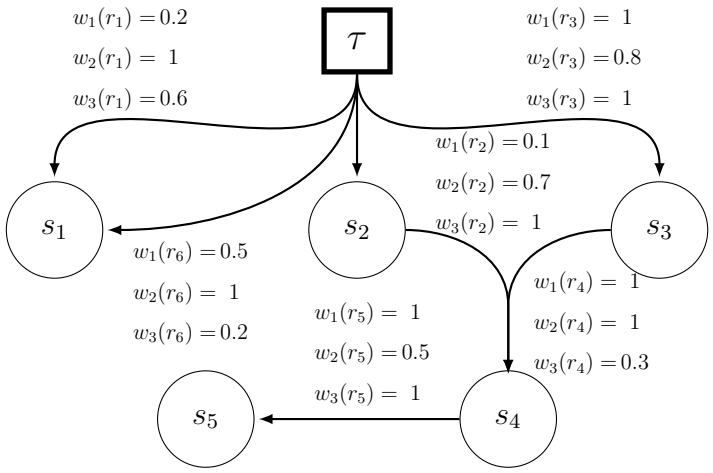



Figure 6.5: Agents' acceptance functions. $w_1$, $w_2$ and $w_3$ encode agents' acceptances of relationships for agents 1, 2 and 3, respectively.

**Observation 6.2.3.** Considering this example, it is worth elaborating on the direction of the relationship regarding content and opinion. We note that relationship $r_1$ is directed from $\tau$ towards statement $s_1$ so that $s_1$ can be interpreted as its consequence. Conversely, the opinion about $s_1$ —be it either positive or negative— will also affect the opinion on $\tau$ through relationship $r_1$. The opinion that a participant will have on a statement, in this case the target, will be conditioned by the consequences of this

statement. Generally, a direct relationship between statements has to establish which direction to take when evaluating the framework. This evaluation could be set to relate one statement $s$ (or a set of statements) to (i) its consequence statements or to (ii) the statement whose opinion is affected by the opinion about $s$. Our approach chooses the first option to reflect the direction of reasoning (from premises to conclusions) in the debate structure. However, the subsequent process aimed at opinion aggregation will follow these relationships in the opposite direction.

We formally introduce next the notion of individual opinion over a $DRF$.

**Definition 6.2.5** (Opinion)**.** Given a $DRF = \langle \mathcal{S}, \mathcal{R}, T \rangle$, an *opinion* over the $DRF$ is a pair $O = (v, w)$ such that $v$ is a valuation function and $w$ is an acceptance degree.

Henceforth, the class of all opinions over a $DRF$ is noted as $\mathbb{O}(DRF)$.

In a practical and realistic application, we can assume that for each relationship, at least one agent values it different from 0. Otherwise, in practical terms, it would be the same as not having that relationship in the debate.

As depicted in Figures 6.4 and 6.5, each agent $i$ involved in a debate will have its individual opinion $O_i = (v_i, w_i)$. The principal goal will be to compute a collective opinion from the opinions issued by the agents, but first, the next section tackles the consistency of an opinion.

## 6.3 Characterising coherent opinions

Previous works on the formal modelling of debates have placed restrictions on the opinions that individuals can put forward. For example, Awad et al. [2015] interpret the opinions expressed by individuals as labels, in the sense of [Baroni et al. 2011], for the arguments that they are expressing opinions about. Thus, an argument can be labelled `in`, meaning that the individual thinks that it holds, `out`, meaning that the individual thinks it does not hold, or `undec`, meaning that the individual is not sure whether it holds or not. These labellings are restricted to be *complete* labellings [Baroni et al. 2011], broadly meaning that they conform to a notion of rationality where arguments are `out` if they are attacked by arguments that have been established to be `in`, and are `in` if they are only attacked by arguments that are `out`.

Though different elements form the previous perspective, for example, arguments or labellings, while the RM has statements and real-valued opinions, we could pursue a similar notion of rationality for the RM. However, when dealing with human participation, the opinions cannot be expected to be rational — contradictions or inconsistencies between the direct and indirect opinions expressed by participants may arise, just as in the example shown above. Humans may express opinions that are far from rational, so imposing this kind of constraint would be too restrictive for modelling human debates, and therefore it is so for the notion of consistency that will be applied to the RM. Instead, weaker conditions for an individual opinion are imposed as to be classified as reasonable or *coherent*, along the lines of the work in part I.

The opinions in the RM are expressed in terms of the desirability or undesirability for the different statements (maybe not all of them), and the acceptability of the relationships linking these statements (maybe not all of them). As mentioned above, relationships are a crucial element of the RM: they indicate that the opinion about one statement affects the opinion about another statement. In particular, we will use the terms: (i) *direct opinion* to refer to the value directly given to a statement by a participant; and (ii) *indirect opinion* to refer to the values given to the related statements whose opinions may affect the direct opinion. By comparing the direct and indirect opinions of each statement, we can determine "reasonable" opinions and provide a notion of *coherence* [Thagard 2002]. Hence, given a statement, the characterisation of coherence consists of contrasting the opinion expressed about that statement, the *direct opinion*, with the opinions expressed about the immediate descendants of the statement, what is called the *indirect opinion*.

Informally, coherence works in the following way. First, an estimated opinion for a statement based on its indirect opinion is computed. If the estimated opinion for the statement aligns with its direct opinion, then the opinion is said to be coherent with respect to the statement. That will be the case when the opinions (valuations) on the descendants are similar to the opinion (valuation) on the statement.

**Example 6.3.1.** Considering the example in Figure 6.4 again, we consider statement $\tau$, its descendants ($s_1$, $s_2$, and $s_3$), and the opinion of agent 2 ($v_2$). We observe that although the direct opinion about $\tau$ is negative ($v_2(\tau) = -0.5$), the valuations for its

descendants are diverse: while the valuation for $s_1$ is also negative ($v_2(s_1) = -1$), and hence in line with $\tau$, the valuations on the other descendants are positive ($v_2(s_2) = 1$ and $v_2(s_3) = 0.5$), and hence not in line with $\tau$. Thus, at first sight[10] it would seem that an overall assessment of the indirect opinion would not be in line with the direct opinion.

In what follows, the estimated opinion is first formalised as an aggregated measure formed from the indirect opinion about a statement —i.e. the collection of values on the descendants and their relationships. The estimation function will consider descendants' valuations of a statement and their acceptance degrees from the relationships connecting them to the statement so that the higher the value of a relationship between a statement and a descendant is, the more important the opinion about that descendant should be. After that, we define the notion of coherence by measuring how close the direct opinion about a statement is to the estimated opinion.

First, we introduce some concepts and notations that will aid us in later steps. Given a $DRF = \langle \mathcal{S}, \mathcal{R}, T \rangle$, we define the *set of relationships from* $s \in \mathcal{S}$ as the set of relationships having $s$ in the set of initial statements. Formally,

$$R^+(s) = \{r = (\Sigma, s', c) \in \mathcal{R} \mid s \in \Sigma\}. \tag{6.1}$$

The term *descendants of a statement* $s$, denoted by $D(s)$, will refer to any statement $s_r$ connected to $s$ by a relationship $r$ that has $s$ as one of the initial statements and $s_r$ as final statement. Formally,

$$D(s) = \{s_r \in \mathcal{S} \mid r \in R^+(s), r = (\Sigma, s_r, c)\}.$$

**Definition 6.3.1** (Direct and indirect opinion)**.** Following from this notation then, given an opinion $O = (v, w)$ the *direct opinion* of $s$ is $v(s)$ and the *indirect opinion* of $s$ is

$$IO(s) = \{(v(s_r), w(r)) |\ r \in R^+(s), r = (\Sigma, s_r, c) \text{ for some } \Sigma \subset \mathcal{S}\}$$

Thus, the indirect opinion of $s$ is the collection of opinions attached to the statements and relationships descending from $s$, grouping each acceptance degree with its respective statement evaluation. The estimation function is defined then as a function that provides a representative value for the indirect opinion of a statement, $IO(s)$.

[10]Note that we are not considering acceptances at this point.

**Definition 6.3.2** (Estimation function)**.** Given a $DRF = \langle \mathcal{S}, \mathcal{R}, T \rangle$ and $O = (v, w)$ an opinion over the $DRF$, the *estimation function* is a valuation function mapping each statement to a value in the set $I$:

$$\begin{array}{rccc} e: & \mathcal{S} & \longrightarrow & I \\ & s & \longmapsto & e(s) \end{array}$$

such that:

- if $I(s) = \emptyset$ then $e(s) = 0$; otherwise,
- $e(s) = f(IO(s))$, i.e. the result of computing the values from $IO(s)$.

For a statement $s$ the estimation function computes a value from the set $IO(s)$ by means of an aggregation function $f$. The estimation function computes an estimated value for a statement using the valuations and acceptance degrees of the indirect opinion about that statement. This definition is generic, allowing for many estimation functions to be specified to compute different approximations for the direct opinion. It, therefore, specifies a broad family of estimation functions rather than any specific function.

The rest of part II though, will use a particular estimation function to compute specific results later on: the weighted average of the valuations of the descendants, where the weights are the acceptance degrees on the relations leading to each descendant. In this manner, the higher the acceptance of a relationship, the more valued the opinion on the descendant related. Formally,

**Definition 6.3.3** (Particular estimation function)**.** Given a $DRF = \langle \mathcal{S}, \mathcal{R}, T \rangle$ and $O = (v, w)$ an opinion over the $DRF$, the *estimation function* used for specific computations is, for all $s \in \mathcal{S}$:

$$e(s) = \begin{cases} 0, & \text{if } R^+(s) = \emptyset \text{ or } \sum_{r \in R^+(s)} w(r) = 0, \\ \dfrac{\sum_{r \in R^+(s)} w(r) v(s_r)}{\sum_{r \in R^+(s)} w(r)}, & \text{otherwise.} \end{cases}$$

Notice that when a statement has no descendants, there is no value to gather from an empty indirect opinion. Notice also that this specification of estimated opinion can be regarded as a general and straightforward approach to approximate a direct opinion using the indirect opinion. Such simplicity and generality bode well with the intention

of allowing the relationships between statements to represent any kind of reasoning, so a specific behaviour for the estimated value cannot be specified.

Informally, an opinion is characterised as coherent for a given statement when the value assigned by the participant (issuing the opinion) to the statement (i.e., its direct opinion) is aligned with the values assigned to its descendants (i.e., its estimate opinion). Furthermore, given the continuous values allowed in the opinion, the degree of coherence can be chosen by using a parameter $\epsilon$. Formally:

**Definition 6.3.4** (Coherence)**.** Consider a $DRF = \langle \mathcal{S}, \mathcal{R}, T \rangle$ and an $\epsilon \in (0, 1)$[11] difference. An opinion $O = (v, w)$ is $\epsilon$-*coherent* on $s \in \mathcal{S}$ when if $D(s) \neq \emptyset$ then:

$$|v(s) - e(s)| < \epsilon$$

where $e$ is the estimation function.

In general, an opinion $O$ will be $\epsilon$-coherent if it is $\epsilon$-coherent for every statement in $\mathcal{S}$. We will denote as $\mathbb{C}_\epsilon(DRF)$ the class of all the $\epsilon$-coherent opinions. Thus, if $O$ is an $\epsilon$-coherent opinion, then $O \in \mathbb{C}_\epsilon(DRF)$.

**Observation 6.3.1.** Although the TODF and RM have very different characteristics and perspectives of a debate, the overall idea of coherence for both models is the same: a labelling or opinion on an argument or statement respectively is coherent if the respective indirect labellings or opinions are supporting it. Since each model has its own characteristics, the method to appreciate the supportiveness of the indirect labellings or opinions is different. The TODF uses the $Pro$ and $Con$ functions and the RM an estimation function.

**Example 6.3.2.** Continuing with the running example introduced in example 6.2.2, we now have the means to compare the values from the estimation function and the actual values given by participant 1 to each statement, the direct opinions, see $v_1$ in Figure 6.6. We can see that if $\epsilon \in (0.3, 1)$ then the opinion of participant 1 for the statements $s_1, s_2, s_3$ and $s_5$ is $\epsilon$-coherent but not for statement $s_4$ due to the difference between direct opinion and estimated value, which is the maximum possible. Because of this

[11] We choose the interval (0,1) for the value of $\epsilon$ as the minimum interval that guarantees that if the direct opinion is 1 (or -1) then an opinion cannot be classified as coherent when the estimation function value is of the opposite sign, i.e., $e(s) \not\leq 0$ (or $e(s) \not\geq 0$ respectively).

statement $s_4$, the opinion of participant 1 cannot be classified as $\epsilon$-coherent for any $\epsilon \in (0, 1)$.

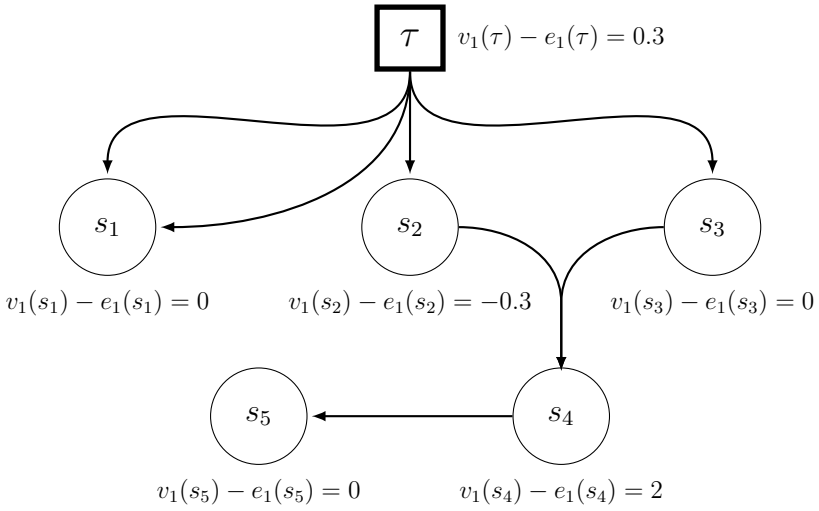



Figure 6.6: Coherence of Agent 1.

## 6.4 Summary

This chapter introduced the RM and its coherence notion.

- Section 6.1 introduced the RM, the contributions in this chapter and a simple protocol to build a debate with the RM.

- Section 6.2 defined the RM formally. The RM is formed by a structure to organise a debate and an opinion to represent the participants' points of view. The structure of the RM is the directed relational framework, composed of statements, elemental sentences without reasoning, and relationships between statements, representing the reasoning behind the connections. Similar to the TODF, a set of targets is defined as the root of the debate. The opinion of the RM is represented by two real-valued functions, a valuation function for the statements and an acceptance function for the relationships.

- Section 6.3 defines the notion of $\epsilon$-coherence in the RM. Using an estimation function, a function to approximate a value for the indirect opinion of a statement, $\epsilon$-coherence can assess the consistency of an opinion considering the dependen-

cies of the debate. In short, an opinion is coherent if, for each statement, the direct opinion and the estimated value of the indirect opinion are similar, closer than $\epsilon$.

We have provided answers to research questions RQ-1 and RQ-2 from Chapter 1. Regarding RQ1, we have proposed the RM, a new model to represent a multi-agent debate. The representation of reasonings as relationships, the two real-valued functions to express the opinions and the clear distinction between structure and opinion in a debate, are novel features that characterise the RM. Furthermore, as in the TODF, a set of statements are the targets of the debate. This new multi-agent debate model called for a new notion of consistency. We introduced the new notion of $\epsilon$-coherence, which allows to evaluate the consistency of an opinion considering some ”tolerance” coherence degree ($\epsilon$). Thus, answering the research question RQ-2.

# Chapter 7

# Aggregation and analysis

This chapter tackles the aggregation functions and the analysis of such functions. The Relational model (RM) is the basis on which to perform the aggregation. Therefore the aggregation functions must be defined accordingly to work with the model.

The following sections of this chapter will be organised as follows. Section 7.1 introduces the research of the following sections. Section 7.2 defines the opinion aggregation problem and introduces the social choice properties used to analyse the aggregation functions defined in Section 7.3. The formal analysis of the aggregation functions is shown in Section 7.4 together with a computational analysis of implementing them. Finally, Section 7.5 summarises the research of this chapter and list the specific contributions.

## 7.1 Introduction

The RM, as defined in the previous chapter represents a debate by organising the information using statements and relationships between them. The agents' opinions are issued on these objects using the valuation and acceptance functions to, respectively, value statements and relationships. Thus, the RM establishes a clear separation between what is the structure of the debate and what are the opinions issued on it. Additionally, the notion of coherence will serve to analyse the collective opinion in terms of the consistency of the opinions, either the opinions of the participants and the aggregated opinion from them.

Once all users have expressed their opinions on statements and relationships, maybe

not on every statement and relationship, opinions must be aggregated to calculate a collective opinion. This aggregation can take into account direct opinions, indirect opinions, or a combination of both direct and indirect. So, the use of dependencies for the aggregation is a factor to consider. In establishing suitable aggregation functions, we must be aware that individual opinions may be incoherent. Nonetheless, the goal is to define aggregation functions that combine these "imperfect" individual opinions into a "reasonable" collective opinion.

The contributions of this chapter are listed in detail below:

- *Aggregation functions exploiting dependencies*. This work proposes several opinion aggregation functions that use the participant's opinion on a debate to compute a collective opinion. We define two families of functions that compute the collective opinion using the dependencies between opinions in different forms.

- *Formal analysis*. There is an assessment of these families of functions against a wide-ranging set of properties designed to provide a detailed characterisation of their behaviours. Several properties are adapted to the RM from the social choice literature [List and Pettit 2002] to assess the aggregation functions in such a model. The same study is carried out in four scenarios considering different worthwhile assumptions regarding the participants' opinions.

- *Computational analysis.* The formal analysis is complemented with a computational analysis. This analysis computes the computational complexity of the aggregation functions, showing their applicability for real-life scenarios.

## 7.2 Formalising the collective decision-making problem

As stated previously, the main goal is to help agents reach a collective decision on target statements. This corresponds to step 4 of the protocol in Section 6.1.In Section 7.2.1 this goal is cast to be an opinion aggregation problem, which we can solve by using an aggregation function that computes a single opinion from all agents' opinions. Although opinions can be aggregated in different ways, here, like in part I, it will be required that the outcome of aggregation should be fair. In particular, Section 7.2.2 introduces desirable social choice properties to help analyse and compare different opinion aggregation

functions.

### 7.2.1 The opinion aggregation problem

First, we need the notion of an opinion profile, which brings together the opinions of the individuals involved in a debate.

**Definition 7.2.1** (Opinion profile)**.** Let $Ag = \{1, \ldots, n\}$ be a set of $n$ agents and a $DRF = \langle \mathcal{S}, \mathcal{R}, T \rangle$. An *opinion profile* is a collection of opinions $(O_1 = (v_1, w_1), \ldots, O_n = (v_n, w_n)) \in \mathbb{O}(DRF)^n$ over the $DRF$ such that $O_i = (v_i, w_i)$ stands for the opinion of agent $i$.

The problem at hand, the opinion aggregation problem, is how to aggregate the agents' opinions from an opinion profile to produce a single opinion. If the opinion profile represents the views expressed by individuals in a debate, the combination should represent the collective opinion of all the individuals. The opinion aggregation function, which we formalise below, is the mechanism for establishing this collective opinion.

**Definition 7.2.2** (Opinion aggregation function)**.** Given a $DRF$ and a set of $n$ agents $Ag$, a function $F : \mathcal{D} \subseteq \mathbb{O}(DRF)^n \longrightarrow \mathbb{O}(DRF)$ mapping an opinion profile from a domain $\mathcal{D}$ to a single opinion is called an *opinion aggregation function*. Given an opinion profile $P$ in the domain $\mathcal{D}$, $F(P)$ is called the *collective opinion* by $F$ and it will be noted as $F(P) = (v_{F(P)}, w_{F(P)})$.

In terms of the components of the RM, the collective opinion, output by an opinion aggregation function, combines the collective valuations over statements and the collective acceptances of the relationships. In Section 7.3 we define specific opinion aggregation functions that compute a collective opinion in different manners. Before that, we introduce the properties that will be used to analyse an opinion aggregation function.

### 7.2.2 Social choice properties

Social choice theory provides formal properties to characterise aggregation functions in terms of outcome fairness [Dietrich 2007]. In what follows, we adapt to RM some of the desirable social properties of an aggregation function introduced in [Awad et al. 2015]

and part I. Besides adapting properties, some novel properties characterise aggregation functions motivated by the fact that here the opinions are real-valued in contrast to the discrete-valued opinions considered in [Awad et al. 2015] and part I.

First, we characterise aggregation functions in terms of the opinion profiles that they can take as input. Thus, from [Awad et al. 2015] we adapt to RM the notion of *Exhaustive domain* to characterise opinion aggregation functions that are defined for any opinion profile. After that, this property is modified to limit an opinion aggregation function to operate with $\epsilon$-coherent opinion profiles.

**Exhaustive domain (ED) .** An opinion aggregation function $F$ has an *Exhaustive domain* if its domain is $\mathcal{D} = \mathbb{O}(DRF)^n$, namely if the function can operate over all profiles.

**$\epsilon$-Coherent domain ($\epsilon$-CD).** An opinion aggregation function $F$ satisfies *$\epsilon$-Coherent domain* property if its domain $\mathcal{D}$ contains all $\epsilon$-coherent opinion profiles, namely $\mathbb{C}_\epsilon(DRF)^n \subseteq \mathcal{D}$.

From now on, sometimes we will refer to $\epsilon$-Coherent domain as "coherent domain".

**Lemma 7.2.1.** An opinion aggregation function satisfying Exhaustive domain property also satisfies $\epsilon$-Coherent domain property.

*Proof.* Straightforward, since an aggregation function taking any opinion profile will also take $\epsilon$-coherent opinion profiles. □

Moreover, next comes *collective $\epsilon$-coherence* as a property characterising opinion aggregation functions that produce $\epsilon$-coherent collective opinions. Therefore, thanks to the parameter $\epsilon$, the notion of collective $\epsilon$-coherence here is more flexible than the notion of coherence used in part I.

**Collective $\epsilon$-coherence ($\epsilon$-CC).** An opinion aggregation function $F$ has *$\epsilon$-Collective coherence* if for all $P \in \mathcal{D}$, then $F(P) \in \mathbb{C}_\epsilon(DRF)$.

Hereafter we will refer to collective $\epsilon$-coherence as "Collective coherence" or $\epsilon$-CC.

In accordance with part I, here we also consider $\epsilon$-CC the most desirable property that can be satisfied by an aggregation function since collective coherence is the foundation of the acceptability of collective decisions [Thagard 2002]. Notice also that,

as in part I, Collective coherence can be regarded as the counterpart of the notion of *Collective rationality* in [Awad et al. 2015].

Next, Anonymity and Non-dictatorship characterise the importance of the agents involved in a debate that yields a collective opinion. On the one hand, Anonymity is a social choice property requiring that the opinions of all the agents involved in a debate are considered to be equally significant. On the other hand, Non-dictatorship requires that no agent overrules the opinions of the rest.

**Anonymity (A)** Let $P = (O_1, \ldots, O_n)$ be an opinion profile in $\mathcal{D}$, $\sigma$ a permutation over $Ag$, and $P' = (O_{\sigma(1)}, \ldots, O_{\sigma(n)})$ the opinion profile resulting from applying $\sigma$ over $P$. An opinion aggregation function $F$ satisfies *Anonymity* if $F(P) = F(P')$.

**Non-Dictatorship (ND).** An opinion aggregation function $F$ satisfies *Non-dictatorship* if there is no agent $i \in Ag$, such that for every opinion profile $P \in \mathcal{D}$, $F(P) = O_i$.

Notice that Non-dictatorship is a weaker version of Anonymity since it follows directly from it — any aggregation function that satisfies Anonymity will satisfy Non-dictatorship.

Now we turn our attention to how an opinion aggregation function behaves when agents agree on their opinions about statements. Unanimity is the social choice property that characterises the behaviour of aggregation functions when there is agreement among agents' opinions. Next, the classic notion of unanimity in [Awad et al. 2015] is adapted as *Narrow unanimity*. *Endorsed unanimity*, from part I, which helps exploit dependencies between statements, is adapted as well to RM. Since the notion of Narrow unanimity is rather rigid, other more flexible unanimity properties are defined. Therefore, a family of unanimity properties are offered, and the study of the relationships between them.

**Narrow unanimity (NU).** Let $P = (O_1, \ldots, O_n)$ be an opinion profile, where $P \in \mathcal{D}$. An opinion aggregation function $F$ satisfies *Narrow unanimity* property if, for any $s \in S$ such that $v_i(s) = \lambda$ for all $i \in \{1, \ldots, n\}$, then $v_{F(P)}(s) = \lambda$ holds.

Narrow unanimity defines unanimity as to when all agents share the same opinion. While this is possible in settings where agents only have a few discrete possibilities for

expressing their opinion, as in [Awad et al. 2015], it is not likely to occur in the setting studied here, where opinions can take a wide range of values. As a result, relaxed variations are proposed, which are more useful for the RM setting. First, we say that *Sided unanimity* will hold when, for each statement, either all opinions on it are positive or negative. Formally,

**Sided unanimity (SU).** Let $P = (O_1, \ldots, O_n)$ be an opinion profile, where $P \in \mathcal{D}$. An opinion aggregation function $F$ satisfies *Sided unanimity* property if for every $s \in S$:

- if $v_i(s) > 0$ for all $i \in Ag$ then $v_{F(P)}(s) > 0$;
- if $v_i(s) < 0$ for all $i \in Ag$ then $v_{F(P)}(s) < 0$.

A weaker version of Sided unanimity is worth distinguishing:

**Weak unanimity (WU).** Let $P = (O_1, \ldots, O_n)$ be an opinion profile, where $P \in \mathcal{D}$. An opinion aggregation function $F$ satisfies *Weak unanimity* property if, for every $s \in S$:

- if $v_i(s) = 1$ for all $i \in Ag$ then $v_{F(P)}(s) > 0$;
- if $v_i(s) = -1$ for all $i \in Ag$ then $v_{F(P)}(s) < 0$.

Although WU requires that all agents agree on fully positive (1) or fully negative (-1) valuations on statements, it does not require that the output of the opinion aggregation function takes on those same values, as required by Narrow unanimity. This property has value when translating valuations expressed in a discrete model such as those in [Awad et al. 2015] into the RM, and so has value in allowing us to relate the model to those which came before.

From the definitions above, it follows that the three notions of unanimity are related.

**Proposition 7.2.1** (Unanimity relationships)**.** If an opinion aggregation function satisfies Sided unanimity, then it satisfies Weak unanimity. If an opinion aggregation function satisfies Narrow unanimity, then it satisfies Weak unanimity.

*Proof.* If an aggregation function cannot hold the sign when the assumptions of Weak unanimity are satisfied, then it is straightforward to see that it will not have Sided unanimity.

If an aggregation function satisfying Narrow unanimity returned a value $\lambda$ when all the agents valued the statement as $\lambda$, it would return the value when $\lambda$ is 1 or $-1$, hence satisfying Weak unanimity as well. □

Below, we see that further assumptions are needed to prove that NU implies SU.

As a final unanimity property, we define Endorsed unanimity, adapted from part I, to consider unanimity based on indirect opinions. In short, an opinion aggregation function will satisfy *Endorsed unanimity* if, for each statement, the collective opinion on the statement is in line with the unanimous indirect opinion on it. Formally,

**Endorsed unanimity (EU).** Let $P = (O_1, \ldots, O_n)$ be an opinion profile such that $P \in \mathcal{D}$. An opinion aggregation function $F$ satisfies *Endorsed unanimity* property if for every $s \in \mathcal{S}$:

(i) if $v_i(s_d) = 1$ for any $i \in Ag$ and $s_d \in D(s)$ (called *full positive support*), then $v_{F(P)}(s) > 0$; and

(ii) if $v_i(s_d) = -1$ for any $i \in Ag$ and $s_d \in D(s)$ (called *full negative support*), then $v_{F(P)}(s) < 0$.

We note that this property is closely related to the notion of coherence. In fact, we will show in the analysis that restricting the domain to coherent opinion profiles will help fulfil this property.

Next is the turn of the monotonicity properties to study how the result of an opinion aggregation function changes as opinions change. First, the notion of Monotonicity is adapted from [Awad et al. 2015]: if some of the direct opinions about a statement increase (or decrease), the collective opinion should increase (or decrease) accordingly.

**Monotonicity (M)** Let $s \in S$ be a statement, and $P = (O_1, \ldots, O_n)$ and $P' = (O'_1, \ldots, O'_n)$ such that for every $i$ $v_i(s) \leq v'_i(s)$. We say that an opinion aggregation function $F$ satisfies *Monotonicity* if $v_{F(P)}(s) \leq v_{F(P')}(s)$.

We notice that M only considers each statement's direct opinion. Since this work aims to handle opinion aggregation functions that combine direct and indirect opinions, we next provide a variation of Monotonicity that considers the dependencies. From

part I, the notion of *Familiar monotonicity* is adapted to the current model[1]. Familiar monotonicity requires that when the direct opinion on a statement increases, the collective opinion does not decrease, provided that the opinions on the descendants of the statement do not change either. Formally:

**Familiar monotonicity (FM).** Let $s \in S$ be a statement, and $P = (O_1, \ldots, O_n)$ and $P' = (O'_1, \ldots, O'_n)$ such that every opinion $i$ satisfies $v_i(s) \leq v'_i(s)$, and, $w_i(r) = w'_i(r)$ and $v_i(s_r) = v'_i(s_r)$ for every relationship $r \in R(s)$ and its associated descendant $s_r \in D(s)$. We say that an opinion aggregation function $F$ satisfies FM if $v_{F(P)}(s) \leq v_{F(P')}(s)$.

The following lemma establishes the relationship between monotonicity properties.

**Lemma 7.2.2.** An opinion aggregation function that satisfies Monotonicity also satisfies Familiar monotonicity.

*Proof.* Since FM assumes only an additional condition on the descendants' opinions compared to Monotonicity, fulfilling Monotonicity implies the fulfilment of Familiar monotonicity. □

The following proposition proves the relationship between Narrow and Sided unanimity via Monotonicity.

**Proposition 7.2.2.** An opinion aggregation function that satisfies Narrow unanimity and Monotonicity also satisfies Sided unanimity.

*Proof.* Let $F$ be an opinion aggregation function that fulfils Narrow unanimity and Monotonicity. Let $P = (O_1, \ldots, O_n)$ be an opinion profile over a $DRF$ and $s \in \mathcal{S}$. Assume that for any $i \in Ag$, $v_i(s) \geq \lambda$ for a certain $\lambda > 0$ and consider the opinion profile $P'$ such that for any $i$ $v'_i(s) = \lambda$. Then, by Monotonicity of $F$, $v_{F(P)}(s) \geq v_{F(P')}(s)$ holds, and by Narrow unanimity of $F$, $v_{F(P')}(s) = \lambda$. Then $v_{F(P)}(s) \geq \lambda > 0$, proving that $F$ also satisfies Sided unanimity. The proof for the negative case ($v_i(s) \leq \lambda$ for $\lambda < 0$) is analogous. □

[1]The name derives from the fact that this form of monotonicity takes into account opinion about the descendants of a statement which make up its family.

Finally, we introduce the Independence property, which will serve to emphasise the difference between opinion aggregation functions that exploit indirect and those that do not. Essentially, the property states that the collective opinion on a statement will only depend on the direct opinions of the statement at hand. Therefore, an opinion aggregation function satisfying Independence disregards the indirect opinions completely.

**Independence (I)** Let there be two profiles $P = (O_1, \ldots, O_n)$ and $P' = (O'_1, \ldots, O'_n)$, such that $P, P' \in \mathcal{D}$ ; and $s \in \mathcal{S}$ a statement, such that for all agents $i \in Ag$ $v_i(s) = v'_i(s)$. An opinion aggregation function $F$ satisfies *Independence* if $v_{F(P)}(s) = v_{F(P')}(s)$.

The next result shows the relationship between Monotonicity and Independence.

**Proposition 7.2.3.** An opinion aggregation function that satisfies Monotonicity also satisfies Independence.

*Proof.* Let $s \in \mathcal{S}$ be a statement and $P = (O_1, \ldots, O_n)$, $P' = (O'_1, \ldots, O'_n)$ two opinion profiles satisfying the assumptions of Independence on $s$, i.e., for every $i \in Ag$ $v_i(s) = v'_i(s)$, and $F$ an aggregation function satisfying Monotonicity.

For each $i$, the equality $v_i(s) = v'_i(s)$ is equivalent to (a): $v_i(s) \geq v'_i(s)$, and, (b): $v_i(s) \leq v'_i(s)$. Thus, assuming Monotonicity from (a) we can deduce $v_{F(P)}(s) \geq v_{F(P')}(s)$, and from (b) we can deduce that $v_{F(P)}(s) \leq v_{F(P')}(s)$. Hence, we conclude that $v_{F(P)}(s) = v_{F(P')}(s)$ proving that $F$ satisfies Independence. □

Having listed these properties, it is important to note that they are not all equally important. For a multi-party discussion, we consider Collective coherence the most important property in this work. If an aggregation function is collectively coherent, the resulting combined opinion will be coherent regardless of the coherence of the initial opinions that are being merged. In other words, an aggregation function that satisfies Collective coherence will always output a coherent collective opinion no matter how incoherent the opinions on which it is based. Along with Collective coherence, the properties that would also be important for an aggregation function are the two domain-related properties — Exhaustive domain and Coherent domain — because they allow for broad applicability of the function, and we naturally prefer the Exhaustive domain property because of its wider reach. Finally, the usual social choice properties of Anonymity, and therefore Non-dictatorship, are regarded as essential.

Among the unanimity properties, since Sided, Weak and Endorsed unanimity are less restrictive than Narrow unanimity, they are preferable for aggregation functions using dependencies. Narrow unanimity is disregarded as desirable because of its close relationship to Independence[2]. Thus, an aggregation function satisfying Narrow unanimity would forbid the use of the indirect opinion the way it is considered to be necessary.

Though it is natural to require some form of monotonicity, M is not desirable because of its relationship to Independence and discarding indirect opinion. Thus, Familiar monotonicity, which considers indirect opinion, is preferable in its place.

Finally, although the design of aggregation functions is focused on the use of both direct and indirect opinion, Independence, Monotonicity and Narrow unanimity are included in the set of properties to emphasise whether the aggregation functions take account of indirect opinion or not.

## 7.3 Aggregation functions for collective decision-making

This section defines a family of opinion aggregation functions for the RM. All aggregation functions presented use some combination of direct and indirect opinions to generate a collective opinion. The goal is to explore the spectrum of aggregation functions, from functions that only employ direct opinions to functions that only employ indirect opinions, so that they can be compared regarding the benefits that they yield in terms of social choice properties (in Section 7.4). This will allow us to learn how to exploit indirect opinions best to obtain collective opinions.

We first define an aggregation function that only aggregates direct opinions and thus disregards indirect opinions. This function will obtain a collective opinion by computing the average of the individual opinions in an opinion profile, valuations per statement and acceptance degrees. Formally:

**Definition 7.3.1** (Direct function)**.** Let $\langle \mathcal{S}, \mathcal{R}, T \rangle$ be a $DRF$ and $P = (O_1, \ldots, O_n)$ an opinion profile over the $DRF$. The *direct aggregation* of $P$ over the $DRF$ is defined as a function $D(P) = (v_{D(P)}, w_{D(P)})$, where $v_{D(P)}(s) = \frac{1}{n} \sum_{i=1}^{n} v_i(s)$ and $w_{D(P)}(r) =$

[2]Though it is not directly related —further minor assumptions must be added to Narrow unanimity to imply Independence.

$\frac{1}{n}\sum_{i=1}^{n} w_i(r)$ for any statement $s \in \mathcal{S}$ and relationship $r \in \mathcal{R}$.

**Example 7.3.1.** Figure 7.1 shows the result of applying the direct function to the opinion profile of the running example, using the profile values of Figures 6.4 and 6.5.

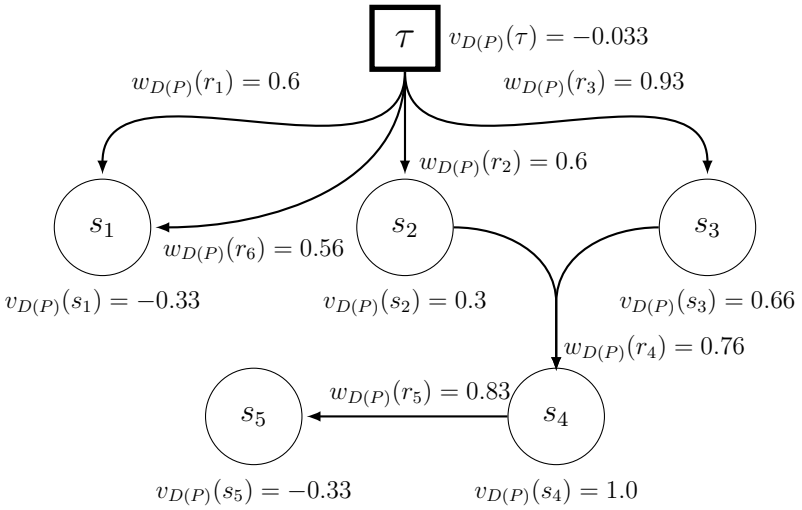



Figure 7.1: Direct function: aggregated valuations.

We now introduce an aggregation function that only aggregates indirect opinions and disregards direct opinions. It is thus the converse of the direct function defined above. The aggregation of indirect opinions aggregates the values extracted using an estimation function (defined in Section 6.3). Formally:

**Definition 7.3.2** (Indirect function)**.** $\langle \mathcal{S}, \mathcal{R}, T \rangle$ be a $DRF$ and $P = (O_1, \ldots, O_n)$ an opinion profile over the $DRF$. The *indirect aggregation* of $P$ over the $DRF$ is defined as a function $I(P) = (v_{I(P)}, w_{I(P)})$, where $v_{I(P)}(s) = \frac{1}{n}\sum_{i=1}^{n} e_i(s)$, where $e_i$ is an estimation function, and $w_{I(P)}(r) = \frac{1}{n}\sum_{i=1}^{n} w_i(r)$ for any statement $s \in \mathcal{S}$ and relationship $r \in \mathcal{R}$.

We notice that, while the direct function computes the average of individuals' direct opinions, the indirect function computes the average of individuals' indirect opinions. This is achieved by aggregating individuals' estimated opinions using an estimation function[3]. We also observe that both functions calculate the aggregation of acceptance degrees in the same way. That will be the case for all the aggregation functions defined

[3]Though for the indirect function the estimation function could be any, in the analyses of the following sections, we will use the particular function from definition 6.3.3.

in this section; hence, the difference between them will be only how they aggregate valuations.

**Example 7.3.2.** Figure 7.2 shows the aggregated (collective) valuations obtained by the indirect function for the opinion profile of the running example shown in Figure 6.4. We notice that $v_I(\tau) = (e_1(\tau) + e_2(\tau) + e_3(\tau))/3 = 0.079$ where $e_1(\tau) = (0.2 \cdot 0 + 0.5 \cdot 0 + 0.1 \cdot 0.7 + 1 \cdot 1)/1.8 = 0.59$, $e_2(\tau) = -0.257$, and $e_3(\tau) = -0.107$. The acceptance degrees of aggregated using the indirect function are the same as in Figure 7.1.

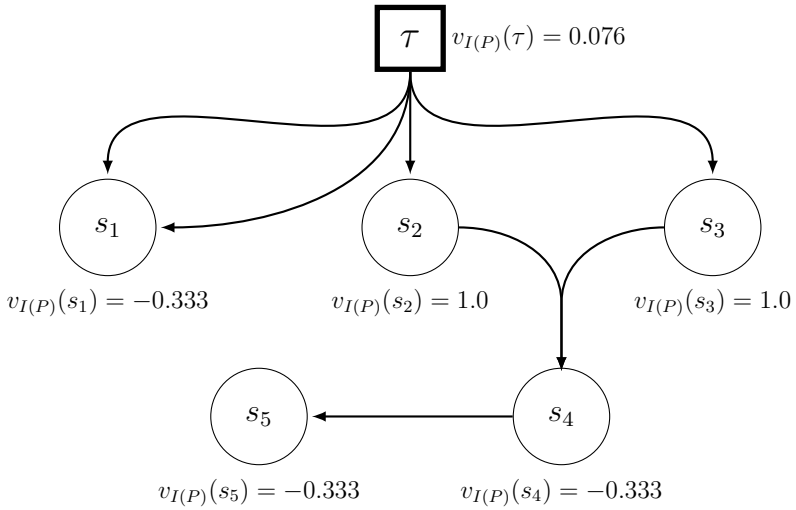



Figure 7.2: Aggregated valuations via indirect function.

Having defined two extremes of the spectrum of functions, next, we present the family of aggregation functions based on a linear combination of the direct and indirect functions.

**Definition 7.3.3** ($\alpha$-balanced function)**.** Let $\langle \mathcal{S}, \mathcal{R}, T \rangle$ be a $DRF$ and $P = (O_1, \ldots, O_n)$ an opinion profile over the $DRF$. Given the direct function $D(P) = (v_{D(P)}, w_{D(P)})$, the indirect function $I(P) = (v_{I(P)}, w_{I(P)})$, and $\alpha \in [0, 1]$, we define the aggregation function $B_\alpha(P) = (v_{B_\alpha(P)}, w_{B_\alpha(P)})$, where:

$$v_{B_\alpha(P)} = \alpha \cdot v_{D(P)} + (1 - \alpha) \cdot v_{I(P)}$$
$$w_{B_\alpha(P)}(r) = \frac{1}{n} \sum_{i=1}^{n} w_i(r)$$

for any statement $s \in \mathcal{S}$ and relationship $r \in \mathcal{R}$. We say that $B_\alpha$ is an $\alpha$-balanced aggregation function.

By changing the value of $\alpha$, we set the importance of the direct opinion with respect to the indirect opinion. The functions from definition 7.3.3 form a family of balanced aggregation functions: $\{B_\alpha\}_{\alpha\in[0,1]}$. In particular, by setting $\alpha$ to 0, we obtain the indirect function, and by setting it to 1, we obtain the direct function.

Next, we define an aggregation function that exploits indirect opinions differently from the indirect function. For a given statement, the *recursive function* calculates its aggregated valuation by using the collective opinion on its descendants, which, in turn, is recursively computed from their descendants, and so on. This recursive computing ends up reaching statements without descendants whose indirect opinion is empty. Therefore, unlike balanced aggregations, the recursive function disregards individual valuations in the indirect opinion and employs their collective opinions instead. This function was inspired by the aggregation functions used in the TODF, which compute the collective opinion recursively.

**Definition 7.3.4** (Recursive function)**.** Let $\langle \mathcal{S}, \mathcal{R}, T\rangle$ be a $DRF$ and $P = (O_1, \ldots, O_n)$ an opinion profile over the $DRF$. The *recursive aggregation* of $P$ over the $DRF$ is defined as a function $R(P) = (v_{R(P)}, w_{R(P)})$, where

$$v_{R(P)}(s) = \begin{cases} \frac{1}{\sum_{r\in R^+(s)} w_{R(P)}(r)} \sum_{r\in R^+(s)} v_{R(P)}(s_r)\cdot w_{R(P)}(r), & \text{if } R^+(s)\neq\emptyset \\ v_{D(P)}(s), & \text{otherwise} \end{cases}$$

and $w_{R(P)}(r) = \frac{1}{n}\sum_{i=1}^{n} w_i(r)$ for any statement $s \in \mathcal{S}$ and relationship $r \in \mathcal{R}$.

We recall that $R^+(s)$ stands for the relationships connecting $s$ to a descendant $s_r$ of $s$ through the relationship $r$.

The recursive function computes the average of the indirect collective opinion computed so far. In fact, we could say that, due to its recursive character, the function computes the estimated opinion for each statement in a bottom-up manner. Thus, the aggregation of opinions starts considering the direct opinions at the "leaves" of the debate, namely at the statements with no descendants, and moves up until reaching the targets.

**Example 7.3.3.** Figure 7.3 shows the aggregated (collective) valuations obtained by the recursive function for the opinion profile of the example shown in Figure 6.4. The aggregated acceptance degrees are the same as in Figure 7.1. Again, for the sake of illustrating the computation, please notice that we start by computing $v_R(s_1) = v_D(s_1) =$

$-0.33$ and $v_R(s_5) = v_D(s_5) = -0.33$ and, from these, we can compute $v_R(s_4) = v_R(s_5) \cdot w(r_5)/w(r_5) = -0.33$, $v_R(s_2) = v_R(s_4) \cdot w(r_4)/w(r_4) = -0.33 = v_R(s_3)$ so to finally compute $v_R(\tau) = (-0.33 \cdot 0.6 - 0.33 \cdot 0.56 - 0.33 \cdot 0.6 - 0.33 \cdot 0.93)/2.69 = -0.33$.

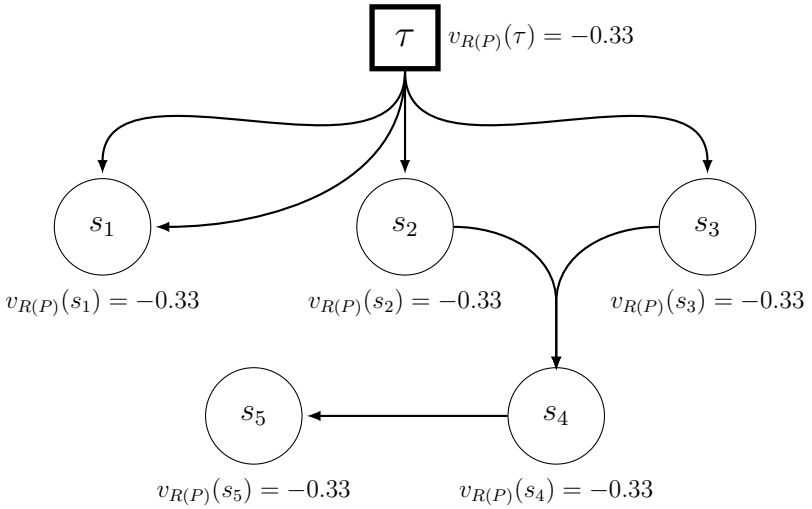



Figure 7.3: Aggregated valuations via recursive function.

Similarly to the balanced family above, we introduce a family of aggregation functions based on combining the direct and recursive functions.

**Definition 7.3.5** ($\alpha$-recursive function)**.** Let $\langle \mathcal{S}, \mathcal{R}, T \rangle$ be a $DRF$ and $P = (O_1, \ldots, O_n)$ an opinion profile over the $DRF$. Given the direct function $D(P) = (v_{D(P)}, w_{D(P)})$, the recursive function $R(P) = (v_{R(P)}, w_{R(P)})$, and $\alpha \in [0, 1]$, we define the aggregation function $R_\alpha(P) = (v_{R_\alpha(P)}, w_{R_\alpha(P)})$, where:

$$v_{R_\alpha(P)} = \alpha \cdot v_{D(P)} + (1 - \alpha) \cdot v_{R(P)}$$
$$w_{R_\alpha(P)}(r) = \frac{1}{n} \sum_{i=1}^{n} w_i(r)$$

for any statement $s \in \mathcal{S}$ and relationship $r \in \mathcal{R}$. We say that $R_\alpha$ is an $\alpha$-recursive function.

## 7.4 Analysing opinion aggregation functions

This section analyses the aggregation functions introduced in Section 7.3 in terms of the social choice properties introduced in Section 7.2. The analysis performed will run

along two dimensions: (1) the *coherence* of an opinion profile; and (2) the *consensus* on the acceptance degrees of an opinion profile. Formally:

**Definition 7.4.1** (Consensus and coherent profiles)**.**

(1) An opinion profile $P$ has *consensus on the acceptances degrees* if any two opinions $O_i = (v_i, w_i), O_j = (v_j, w_j)$ from the opinion profile satisfy $w_i = w_j$.

(2) A *coherent profile* is an opinion profile $P = (O_1, \ldots, O_n)$ such that for any $i \in \{1, \ldots, n\}$, $O_i \in \mathbb{C}_\epsilon(DRF)$ for some $\epsilon \in (0, 1)$.

Thus, the assumptions for the aggregation will consider whether agents' opinions are constrained to be coherent (the opinion profile is coherent) or whether agents agree on acceptance degrees (there is consensus on acceptance degrees). This results in four debate scenarios to analyse:

1. Unconstrained opinion profiles;
2. Constrained opinion profiles: assuming consensus on acceptance degrees;
3. Constrained opinion profiles: assuming coherent profiles; and
4. Constrained opinion profiles: assuming consensus on acceptance degrees and coherent profiles.

The analysis of these scenarios will help us assess the price that must be paid if the opinions stated by participating agents are not necessarily coherent. The scenarios will also help us assess the price that must be paid when the relationships between statements are open for discussion by means of acceptance degrees.

We present our analysis informally in the remainder of this chapter and include the mathematical formulation of the statements and their proofs in Appendix B.

### 7.4.1 Unconstrained opinion profiles

This is the most general scenario that we can consider for the analysis. We assume unconstrained opinion profiles, meaning that *any* opinion profile is deemed a possible input for the aggregation functions introduced in Section 7.3. In other words, the domain of the aggregation functions is the class $\mathbb{O}(DRF)^n$ itself, and hence opinions need not be coherent nor have consensus on the acceptance degrees.

Table 7.1 shows the social choice properties of the opinions aggregation functions defined in Section 7.3 in this general case. There is one column for each aggregation function and one row for each social choice property. We notice the distinction between desirable social choice properties and other properties, as discussed in Section 7.2.2. In the table, a green square (with a tick) indicates that a property is fulfilled, while a red square (with a cross) indicates that a property is not fulfilled. As to the more general aggregation functions, $\alpha$-B(alanced), and $\alpha$-R(ecursive), in some cases, the values $\alpha$ for which a given property holds are specified. For both families, the results are considered for $\alpha \in (0, 1)$, not considering 0 or 1. Thus the cases for the extreme values represent aggregations functions already shown in other columns.

| Desirable properties | D | I | R | $\alpha$-B | $\alpha$-**R** |
|---|---|---|---|---|---|
| **Collective coherence** | ✗ | ✗ | ✓ | ✗ | $\alpha < \epsilon/2$ |
| Exhaustive domain | ✓ | ✓ | ✓ | ✓ | ✓ |
| Coherent domain | ✓ | ✓ | ✓ | ✓ | ✓ |
| Anonymity | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-dictatorship | ✓ | ✓ | ✓ | ✓ | ✓ |
| Sided unanimity | ✓ | ✗ | ✗ | ✗ | ✗ |
| Weak unanimity | ✓ | ✗ | ✗ | $\alpha > 1/2$ | $\alpha > 1/2$ |
| Endorsed unanimity | ✗ | ✓ | ✗ | $\alpha < 1/2$ | ✗ |
| Familiar monotonicity | ✓ | ✓ | ✗ | ✓ | ✗ |
| Other properties | | | | | |
| Monotonicity | ✓ | ✗ | ✗ | ✗ | ✗ |
| Narrow unanimity | ✓ | ✗ | ✗ | ✗ | ✗ |
| Independence | ✓ | ✗ | ✗ | ✗ | ✗ |

Table 7.1: Social choice properties satisfied by aggregation functions D(irect), I(ndirect), R(ecursive), $\alpha$-B(alanced), and $\alpha$-R(ecursive) for: (i) a general scenario considering unconstrained opinion profiles; (ii) a scenario considering constrained opinion profiles: consensus on acceptance degrees.

**Domain and anonymity.** Table 7.1 shows that All the proposed opinion aggregation functions fulfil Exhaustive domain, Coherent domain, Anonymity and Non-dictatorship. This is because of the agnostic treatment of opinion profiles adopted by the aggregation functions. Since no constraints are imposed on opinion profiles received as input, ED is satisfied, and since no agent in an opinion profile receives special treatment, Anonymity

holds. Satisfying this family of properties is important. On the one hand, fulfilling ED ensures that any opinion profile can be used as input, that is, the aggregation functions do not filter out participants' opinions before computing a collective opinion. On the other hand, satisfying Anonymity guarantees that all participants are equally important when calculating a collective opinion.

**Collective coherence.** The Direct and Indirect functions do not satisfy Collective coherence. As a result, neither does any $\alpha$-Balanced aggregation function because they result from the linear combination of D and I. The result of such aggregation functions largely depends on the coherence of the opinion profile at hand, which can be as incoherent as possible in this scenario. More positively, the Recursive function does satisfy Collective Coherence (CC). Out of the family of recursive functions ($\alpha$-R), which relies on D and R, those for which $\alpha < \epsilon/2$, where $\epsilon$ is set to assess the coherence of the output, also satisfies CC. This tells us that the closer $\alpha$ is to 0 (the less the use of the direct opinion), the more coherent the collective opinion obtained by an $\alpha$-R function will be. The closer $\alpha$ is to $\epsilon/2$, the less coherent the collective opinion obtained by an $\alpha$-R function will be. When $\alpha$ goes above $\epsilon/2$, the $\alpha$-R function depends more on the direct opinion (which does not satisfy CC), so CC does not hold.

**Unanimity.** Narrow, Sided and Weak unanimity are not satisfied by the Indirect and Recursive functions. This is because the indirect opinion, employed by all these aggregation functions, ignores unanimity on the direct opinion and in some cases, these functions can produce a result in the opposite direction. On the other hand, the Direct function, which only depends on the direct opinions of a statement, does satisfy all the unanimity properties. This benefits the Balanced and Recursive families, which satisfy Weak unanimity for some values of $\alpha$. Notice that only Balanced and Recursive aggregation functions for which $\alpha$ is greater than $1/2$ satisfy Weak unanimity. This is to lessen the influence of the indirect opinion and sway the result towards the Direct function, which does satisfy the property. Regarding the Narrow and Sided unanimity properties, not even the influence of the Direct function is enough to guarantee that unanimity is preserved, and therefore no aggregation function in the Balanced or Recursive families fulfils them for any value of $\alpha$.

Regarding Endorsed unanimity, the situation changes for the Direct and Indirect functions. They flip sides so that the Direct function does not fulfil Endorsed unanimity,

but the Indirect function does. This is because the unanimity, in this case, resides in indirect opinions, and hence it is in line with the Indirect function, which only depends on indirect opinions. However, this works against the Direct function, which disregards indirect opinions and hence unanimity on its values. Conversely to the Weak unanimity case for the Balanced family, now the values of $\alpha$ have to be less than $1/2$ to sway the balanced aggregation towards the Indirect function and hence, to satisfy Endorsed unanimity. Next, although it might seem reasonable that aggregation functions in the Recursive family also fulfil Endorsed unanimity, they do not. This is caused by the recursive behaviour of these aggregation functions, which can overlook unanimity on indirect opinions to use instead opinions deep in the debate on which there might be no unanimity. And last, due to the failure of the Direct and Recursive functions to fulfil Endorsed unanimity, so do all the aggregation functions in the Recursive family, no matter the value of $\alpha$.

**Monotonicity.** The Familiar monotonicity property is fulfilled by the Direct function (as a consequence of fulfilling Monotonicity), the Indirect function, and therefore by the whole family of Balanced functions that are combinations of the Direct and Indirect functions. The Recursive function, and therefore the Recursive family, fails to satisfy Familiar monotonicity because, given a statement, the aggregated opinion about its descendants does not solely depend on the valuations of these descendants alone. Instead, the aggregated opinion about its descendants recursively depends on descendants down the relational framework. Thus, changes of opinion on "grandchildren" statements can cause a change of opinion independently of any change of the direct opinion.

**Other properties.** For completeness, we also analysed the fulfilment of further non-desirable properties. As expected, Independence is not fulfilled by any of the functions making use, at any degree, of indirect opinion, namely the Indirect function, the Recursive function, and the Balanced and Recursive families (for any $\alpha < 1$). Also related to the use of indirect opinions, we observe that any of these four functions do not fulfil Narrow unanimity and Monotonicity. These properties can only be satisfied when employing the direct opinion alone to obtain the aggregated collective opinion. This reinforces the discussion in Section 7.2.2 about disregarding these properties to consider alternative properties more orientated to aggregation functions that take account of dependencies.

### 7.4.2 Constrained opinions: assuming consensus on acceptance degrees

In this scenario, we assume the opinion profiles have a consensus on the acceptance functions, i.e., that all the participants value each reasoning of the debate equally. Adding this assumption does not change the results shown in the previous Section 7.4.1. Each aggregation function satisfies exactly the same properties as when there were no assumptions on the opinion profiles. Therefore Table 7.1 shows the results for this scenario as well. Nonetheless, it was deemed worthy to analyse this debate scenario because of the multiple already-in-use participation systems that do not allow participants to value the relationships between statements differently. In this case, if it is assumed that participants agree on acceptance degrees, the collective opinion will only depend on valuations over statements. This is equivalent to considering a debate where participants are allowed to value statements but do not express their opinions on the relationships between them.

Appendix B contains the proofs regarding the fulfilment of social choice properties in Section 7.4.1. By analysing such proofs, we observe that assuming consensus on acceptances does not yield any further benefit that did not appear in the analysis of Section 7.4.1. This justifies the lack of additional proofs for this scenario in Section B.2.

### 7.4.3 Constrained opinions: assuming coherent profiles

In this section, the opinion profile is assumed to be constrained as coherent to some degree (according to some value $\epsilon \in (0, 1)$). We recall that coherence occurs when the direct and indirect opinions are in line. Therefore, assuming coherence is expected to positively impact aggregation functions that exploit indirect opinions to compute a collective opinion.

Table 7.2 shows the desirable properties satisfied by the aggregation functions when assuming coherence. The light green squares with check marks identify properties that are now satisfied but were not (in Table 7.1) when not imposing coherence. Therefore, assuming coherence yields new positive results. More precisely, Table 7.2 shows that assuming coherence leads to the satisfaction of desirable unanimity properties for

| Desirable properties | D | I | R | $\alpha$-B | $\alpha$-**R** |
|---|---|---|---|---|---|
| **Collective coherence** | ✗ | ✗ | ✓ | ✗ | $\alpha < \epsilon/2$ |
| Exhaustive domain | ✓ | ✓ | ✓ | ✓ | ✓ |
| Coherent domain | ✓ | ✓ | ✓ | ✓ | ✓ |
| Anonymity | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-dictatorship | ✓ | ✓ | ✓ | ✓ | ✓ |
| Sided unanimity | ✓ | ✗ | ✗ | ✗ | ✗ |
| **Weak unanimity** | ✓ | ✓ | ✗ | ✓ | $\alpha > 1/2$ |
| **Endorsed unanimity** | ✓ | ✓ | ✗ | ✓ | $\alpha > \frac{1}{2-\epsilon}$ |
| Familiar monotonicity | ✓ | ✓ | ✗ | ✓ | ✗ |

Table 7.2: Highlighted, in light colour, the fulfilment of additional desirable properties, in addition to those shown in Table 7.1, when assuming coherent opinions.

several functions. First, given the coherence assumption, the unanimity on the direct opinion drags the indirect opinion to become more similar to it, and therefore the Indirect function gains Weak unanimity. Now, since the Direct function also satisfies it, it follows that all $\alpha$-Balanced functions now fulfil it too. Furthermore, thanks to the alignment that the coherence assumption brings between the direct and indirect opinions, the Direct function fulfils the Endorsed unanimity property. Therefore, having Endorsed unanimity fulfilled now by the Indirect and Direct functions, the aggregation functions in the Balanced family also fulfil it for any $\alpha$.

We observe that unanimity and the coherence assumption work well together. Unanimity on one statement brings together its direct and indirect opinions, making it impossible for both to be far apart and therefore allowing the Direct and Indirect functions to fulfil more unanimity properties.

Finally, the family of Recursive function now fulfils Endorsed unanimity, though not for any $\alpha$. Depending on the degree of coherence allowed in the opinion profile, i.e. the value of $\epsilon$, the interval of $\alpha$ values allowing $R_\alpha$ to fulfil Endorsed unanimity will change. In this case, $\alpha$ has to be greater than $1/(2-\epsilon)$, representing the need to overcome the bad result obtained by the Recursive function with respect to the Endorsed unanimity property.

### 7.4.4 Constrained opinions: assuming consensus on acceptance degrees and coherent profiles

In this section, we assume both previous constraints on the opinion profiles: coherence on the opinions and consensus on acceptance degrees. First, consensus on acceptance degrees on relationships represents a more simplified debate where participants only provide their opinions on statements. Second, the coherence assumed on opinions aligns direct and indirect opinions. Overall, both assumptions yield significant benefits regarding the satisfaction of desired social choice properties, as discussed next.

Table 7.3 shows the gain in fulfilment of desirable properties with respect to Table 7.2. The light green squares with check marks identify properties that are now satisfied but were not (in Table 7.2) when not imposing consensus on acceptance degrees.

| Desirable properties | D | I | R | $\alpha$-B | $\alpha$-**R** |
|---|---|---|---|---|---|
| **Collective coherence** | ✓ | ✓ | ✓ | ✓ | ✓ |
| Exhaustive domain | ✓ | ✓ | ✓ | ✓ | ✓ |
| Coherent domain | ✓ | ✓ | ✓ | ✓ | ✓ |
| Anonymity | ✓ | ✓ | ✓ | ✓ | ✓ |
| Non-dictatorship | ✓ | ✓ | ✓ | ✓ | ✓ |
| Sided unanimity | ✓ | ✗ | ✗ | ✗ | ✗ |
| Weak unanimity | ✓ | ✓ | ✗ | ✓ | $\alpha > 1/2$ |
| Endorsed unanimity | ✓ | ✓ | ✗ | ✓ | $\alpha > \frac{1}{2-\epsilon}$ |
| Familiar monotonicity | ✓ | ✓ | ✗ | ✓ | ✗ |

Table 7.3: Highlighted, in light colour, the fulfilment of additional desirable properties, in addition to those shown in Table 7.2, when assuming coherent opinions and consensus on acceptance degrees.

Now, besides the aggregation functions in the Recursive family, which now satisfies Collective coherence for any $\alpha$, the rest of the aggregation functions under study also satisfy $\epsilon$-Collective coherence when the opinion profiles are $\delta$-coherent, $0 < \delta \leq \epsilon$.

This significant improvement is because the consensus on acceptance degrees forbids the participants to value a relationship as 0, which is key to ensuring Collective coherence for the Direct and Indirect functions when the opinion profiles are coherent. As discussed in Section 6.2.2, we assume that for each relationship, at least one agent has valued it other than 0 because, otherwise, it would be as if the relationship did not

exist. This forces all the participants to have a positive value too.

In this manner, if the opinion profile is assumed to have both $\epsilon$-coherent opinions and consensus, then all the aggregation functions can guarantee $\epsilon$-coherent aggregated opinions, which may increase the acceptability of the results from the participants.

We notice that assuming consensus on acceptance degrees is quite reasonable. Indeed, such consensus is very likely to be found in many debates where the relationships are classified first and, after, the participants are allowed to give their opinions. Though the RM allows different acceptance degrees, it can fit perfectly with these scenarios by setting all the acceptance degrees as a constant value for every participant. Furthermore, the procedure to create the debate, and therefore the $DRF$, could be adapted so there is a first stage in which a collective value is established for each of the relationships, and then there is a second stage in which values are assigned to the statements.

### 7.4.5 Conclusions of the analysis

From the analysis for each debate scenario above, we can draw the following general observations:

- The aggregation functions of the recursive family achieve Collective coherence provided that they place little weight on direct opinions (or opinions are coherent, and there is consensus on acceptance degrees).

- Coherence in opinion profiles favours unanimity (specifically, WU and EU), though in different ways. $I$ and $\alpha$-Balanced are fully satisfied, while the family of recursive functions leans on the direct function to fulfil some unanimity properties with restrictions. As a result, the $\alpha$-Recursive family only satisfy WU and EU under strong conditions on $\alpha$ because the $R$ function never satisfies them.

- Coherent opinion profiles are insufficient for $D$, $I$, and $\alpha$-Balanced functions to achieve Collective coherence. They also require consensus on acceptance degrees. Recursive functions do not require such consensus (in fact, not even the non-coherent opinion profiles); hence, they are *robust* to the divergence of opinions on the relations between statements in a debate.

- While the $D$, $I$, $B_\alpha$ functions manage to achieve Familiar monotonicity in all

scenarios, the aggregation functions in the recursive family cannot even when counting on coherent opinion profiles and consensus on acceptance degrees. This is because the aggregated opinion on descendants recursively depends on descendants down the $DRF$. Thus, changes of opinion on "grandchildren" or further down statements can cause a change of opinion independently of any change of the direct opinion.

Based on these general observations above, it is the task of the person or group in charge of the debate to decide the aggregation operator to choose, considering: (1) the features of the debate scenario at hand; and (2) the desirable properties to guarantee. As a rule of thumb, since we cannot assume individual rationality (coherence) in real-world debates, the recursive aggregation functions would be the best choice to achieve Collective coherence. However, there is a price to pay because of the loss of some other valuable properties, in particular, unanimity for values of $\alpha$ that promote a significant use of the direct opinion. Otherwise, if the coherence of the collective output is not deemed so important, or we can guarantee that somehow the opinions of participants are coherent and the participatory system at hand does not allow for divergence on acceptances degrees, the Direct function becomes the aggregation function of choice. Within such constrained settings, the Direct function fulfils almost every property considered. Actually, all of them in the most restricting debate scenario explored in Section 7.4.4. In general, we can conclude that it would be advisable to consider the Recursive family, which can behave as similar to the Direct or to the Recursive function as wanted, and set the value of $\alpha$ depending on the features and goals in hand.

### 7.4.6 Computational complexity

Given the opinion aggregation problem in Section 7.2.1, this section explains the complexity of the different algorithms for computing a collective decision on its target. In particular, an algorithm for computing the recursive aggregation function is provided.

All aggregation functions proposed in Section 7.3 can be calculated by tractable algorithms. For example, the direct function calculates the average for all statements and relationships in a $DRF$ $\langle \mathcal{S}, \mathcal{R}, T \rangle$ considering the direct opinions in an opinion profile $P = (O_1, \ldots, O_n)$. Hence, its complexity is given by $\mathbf{O}((|\mathcal{R}| + |\mathcal{S}|) \times |P|)$, where $|\mathcal{R}|, |\mathcal{S}|$ are the number of relationships and statements, respectively; and $|P|$

**Algorithm 2** Compute recursive aggregation

**function** COMPUTERECURSIVEAGGREGATION($\langle \mathcal{S}, \mathcal{R}, T \rangle, (O_1, \ldots, O_n)$)
  **for** $r \in \mathcal{R}$ **do** ▷ Compute averaged acceptances
    aggregated_acceptance[$r$]← average_acceptances($w_1(r), \ldots, w_n(r)$)
  $\mathcal{H}(\langle \mathcal{S}, \mathcal{R}, T \rangle)$ ← DRF_to_B-hypergraph($\langle \mathcal{S}, \mathcal{R}, T \rangle$) ▷ Generate B-hypergraph of DRF
  sorted_sentences ← reverse(topological_sorting($\mathcal{H}(\langle \mathcal{S}, \mathcal{R}, T \rangle)$)) ▷ Topological sorting B-hypergraph
  **for** $s$ in sorted_sentences **do** ▷ Compute aggregated valuations
    valuation[$s$] ← 0 ▷ To accumulate aggregated valuations over descendants
    normaliser[$s$] ← 0 ▷ To normalise aggregated valuations over descendants
    compute relationships $R(s)$ to descendants
    **if** $R(s) \neq \emptyset$ **then** ▷ if $s$ has descendants
      **for** $r \in R(s)$ **do**
        $s_r$ ← descendant from relationship $r$
        valuation[$s$] ← valuation[$s$] + aggregated_valuation[$s_r$] · aggregated_acceptance[$r$]
        normaliser[$s$] ← normaliser[$s$] + aggregated_acceptance[$r$]
      valuation[$s$] ← valuation[$s$] / normaliser[$s$]
    **else** ▷ $s$ has no descendants
      valuation[$s$] ← average_valuations($v_1(s), \ldots, v_n(s)$)
    aggregated_valuation[$s$] ← valuation[$s$]
  **return** aggregated_valuation,aggregated_acceptance

is the number of opinions in an opinion profile. Computing the indirect and balanced functions can be done by calculating the aggregated acceptance of each relationship as an average and by calculating the aggregated valuation of each statement as the average of the estimation function, which in turn is an average of the indirect opinions for that statement. Hence, their complexity is given by $\mathbf{O}(|\mathcal{R}| \times |\mathcal{S}| \times |P|)$. The calculation of the recursive function can be done by calculating the aggregated acceptance of each relationship as an average and calculating the aggregated valuation of each statement by starting with statements with no descendants and using these results to calculate the aggregated valuation of the statements directly connected to them. In algorithm 2 we offer the pseudocode for the recursive function.

The algorithm starts by computing aggregated acceptances ($w_{R(P)}$) as a weighted average (lines 2-3), which has a complexity of $\mathbf{O}(|\mathcal{R}| \times |P|)$. Then, the algorithm computes aggregated valuations ($v_{R(P)}$) starting from the statements with no descendants. To do that, first, a topological sorting of the DRF is performed. This can be achieved by: (1) transforming the graph associated to the DRF into an acyclic B-hypergraph [4], denoted by $\mathcal{H}(\langle \mathcal{S}, \mathcal{R}, T \rangle)$ (line 4); and (2) then performing the topological sorting over the B-hypergraph (line 5). Starting from the statements without descendants, the algorithm computes aggregated valuations until reaching the statements in $T$ (lines 5-18). Note that algorithm 2 is not recursive, its name "Compute recursive aggregation" relates to the fact that is an implementation of the *recursive* function. The calculation of the topological sorting for B-hypergraphs has been studied in [Gallo et al. 1993][5]

[4]B-hypergraphs are a particular type of hypergraph with efficient algorithms for path finding [Gallo et al. 1993]. We can benefit from such results to compute the topological sorting of a DRF. This is because the hypergraph associated with a DRF can be readily turned into a B-hypergraph. The statements in a DRF become the nodes in a B-hypergraph, whereas the relationships in the DRF become the hyperedges in the hypergraph. However, we must consider one particular case: when several relationships connect the very same statements, we will consider that they will be all represented by a single hyperedge in the B-hypergraph. For instance, consider relationships $r_1$ and $r_6$ in Figure 6.3 linking $\tau$ to $s_1$. Since, in general, there cannot be two or more hyperedges in a hypergraph over the same nodes, we will consider only one single hyperedge to represent that $\tau$ and $s_1$ are related. In the example, it suffices to consider either $r_1$ or $r_6$. We do not lose anything by doing this simplification because we want to obtain the topological sorting of a DRF, and hence considering one of the relationships connecting the very same statements is enough.

[5]In particular, [Gallo et al. 1993] provides an algorithm to calculate the inverse topological sorting in a F-hypergraph. Any given B-hypergraph can be transformed into a symmetric F-hypergraph by changing

and has a complexity of $\mathbf{O}(|\mathcal{R}| \times |\mathcal{S}|)$. Finally, the complexity from lines 6 to 18 is $\mathbf{O}(|\mathcal{S}| \times |\mathcal{R}|)$ —a **for** instruction traversing relationships (line 11) nested inside a **for** instruction traversing the statements (line 6). Since the different parts have complexity $\mathbf{O}(|\mathcal{S}| \times |\mathcal{P}|)$ —computing the weighted average (lines 2-3)—, $\mathbf{O}(|\mathcal{S}| \times |\mathcal{R}|)$ —sorting topologically the B-hypergraph (line 5)— and $\mathbf{O}(|\mathcal{S}| \times |\mathcal{R}|)$ —computing the valuations (lines 6-18)—, the overall complexity is $\mathbf{O}(|\mathcal{S}| \times max(|\mathcal{R}|, |P|))$, the maximum of the previous three.

In `https://bitbucket.org/jariiia/workspace/projects/DRF` we can find a publicly-available implementation of algorithm 2 together with all the aggregation functions defined in this part.

## 7.5 Summary

Here we summarise the contents presented so far in the chapter.

- Section 7.2 presented the aggregation problem and an the aggregation function, a function to compute a collective opinion in the RM. Additionally, presented the social choice properties that are used to analyse the aggregation functions. Of them, $\epsilon$-Collective coherence is the most desirable function for an aggregation function.

- In Section 7.3 we defined specific aggregation functions for the RM: the Direct function, computing the collective opinion from the direct opinions of the agents; the Indirect function, computing the collective opinion from the indirect opinions using an estimation functions; and, the Recursive function, using recursively the indirect opinion of the collective opinion already computed.

  From these three functions we defined two families of functions: the Balanced family, a set of functions linearly combining the direct and indirect functions; and the Recursive family, a set of functions linearly combining the direct and recursive functions. Within these two families we find functions using any degree of direct or indirect opinion, from a function only using the direct opinion via

the direction of the hyperedges. Note the inverse topological sorting of the symmetric F-hypergraph coincides with the topological sorting in the original B-hypergraph.

direct function to a function only using indirect opinion using either indirect or recursive function, respectively.

- Finally, Section 7.4 presented the assessment of the aggregation functions in terms of their social choice properties. We considered four scenarios regarding the restrictions on the opinion profiles: unconstrained opinion profiles, consensus on acceptance degrees, coherent profiles, or both assuming consensus on the acceptance degrees and coherent profiles.

  From these analyses, among many results, we have seen that restricting the opinion profiles increase the social choice properties satisfied by the aggregation functions functions. Especially, Collective coherence is fulfilled by all the functions in the most restrictive scenario.

  In addition to the formal analysis, we presented an algorithm for implementing the recursive function and assessed the computational complexity of the different aggregation functions.

With regards to our research questions from Chapter 1, we can report the following positive answers:

RQ-3 We defined two families of aggregation functions that exploit differently the dependencies in a debate.

RQ-4 To assess aggregation functions that exploit dependencies, as in the TODF, we created new social choice properties regarding coherence, unanimity and monotonicity. This resulted in the introduction of Collective coherence, Endorsed unanimity and Familiar monotonicity, three new properties that take into account the dependencies between arguments in a debate.

RQ-5 By analysing the aggregation functions, we have seen that exploiting dependencies can benefit the fulfilment of several desired social choice properties. Interestingly, in the most restrictive scenario, the direct function, which disregards dependencies, proved to be the best choice.

# Chapter 8

# Conclusion and discussion

Given the emerging interest in the use of information and communication systems to allow citizens to participate in the governance process, the research on this part II about the Relational model (RM) aims to contribute to such context and provide tools to strengthen the collaboration between governments and citizens. The model allows to formally represent debates where citizens participate in a policy decision-making discussion and offers several options to obtain the collective opinion to take into account.

Several existing participatory systems either fail at providing a comprehensive environment where participants are allowed to show their complex opinion satisfactorily or fail to aggregate the different views in a meaningful way, or they aggregate in a way that limits users to just voting in favour of or against an issue. The RM characterisation and its study of the aggregation functions intend to overcome such limitations.

To address these limitations, the RM aims to provide a better representation of human debates. Particularly, the model allows a better expressiveness in the debate from the participants.

The RM offers a new answer to research question RQ-1 by being the first multi-agent debate model that does not merge the information of the debate with the subjective view of the participants' opinions. It distinguishes clearly what should be the structure: statements and relationships directing the discussion toward a set of targets (the DRF); and what should be opinion: valuation and acceptance functions, which capture each of the participants' opinion issued upon the structural elements of the debate. This distinction makes it possible for the participants to completely provide their unique views and opinion of the debate without forcing themselves to accept a fixed structure where

all the participants must share a unique and subjectively-build structure of the debate– such as argumentation frameworks where the participants share a unique structure with the subjectively classified attack or defence relationships between arguments. Furthermore, participants' opinions can be more precisely represented thanks to the duality of the opinion representation system and its real-valued intervals. The acceptance and opinion functions provide the user with a helpful division of the opinion that allows them to express different types of subjectivity important in the debate. The range of possible values to choose at each evaluation, having continuous intervals, amplifies the precision with which the participants can express their point of view. Furthermore, this continuous nature of the opinion allows us to consider such a wide range of opinion aggregation functions featuring as a parameter the degree of dependencies or indirect opinion we want to use for the aggregation.

In addition, the model does not assume that the users' opinions are rational though a weaker notion of rationality is defined as to characterise *coherent* user opinions and take advantage of them. Thus, answering research question RQ-2 positively.

As a response to research question RQ-3, two wide families of opinion aggregation functions are provided to face the aggregation of opinions: the *balanced and the recursive families*, which combines the *direct function* with the *indirect functions* and the direct function with the *recursive function*, respectively. These families explore all the possible degrees of dependencies to exploit in the aggregation. However, each family considers a different type of computation from the indirect opinion. The balanced family uses the indirect opinions directly from the profile of opinions, a consequence of using the indirect function. The recursive family uses the indirect opinions already computed in the collective opinion, a consequence of using the recursive function.

Although the model aims to work on a general scenario, the study on the opinion aggregation functions differentiates between four scenarios assuming: unconstrained profiles, constraint coherent profiles, and profiles constraint to have consensus on the acceptance degrees or profiles constraint by the two previous assumptions. That allows us to make a complete analysis of their behaviour regarding the social choice properties they fulfil. For each of the four scenarios, a complete assessment of the social choice properties fulfilled by the aggregation functions is made, providing us with a deep understanding of their behaviour (in appendix B we find the respective proofs). Provided

that the families of aggregation functions can vary the degree of indirect opinion to use, the analysis also reflects the importance of applying more or less indirect opinion to perform better or worse regarding different properties.

As it is beneficial for an opinion aggregation function to produce a coherent collective opinion, regarding the social acceptance in a participation system, the recursive family of aggregation functions stand out for their better performance when the scenario is unrestricted. The analysis demonstrates that the recursive aggregation function can compute a coherent collective opinion in its more general scenario when individual opinions can be incoherent, and there is a lack of consensus on the debate structure. On the other side, diminishing the importance of a coherent collective opinion and searching for the best performance in terms of the number of properties fulfilled, the direct function stands as the preferable function to use. In the most general scenario, the direct function satisfies all the desired social choice properties except for Collective coherence and Endorsed unanimity. As the scenario assumes more restrictions on the coherence of individual opinions and consensus among users on the debate structure, more aggregation functions achieve to compute coherent collective opinions. In the end, we see that in the more restrictive scenario, all the functions satisfy Collective coherence. Thus, it seems clear that in such a restrictive scenario, the best function to choose is the one satisfying all the desired social choice properties, which in this case is the direct function (see Table 7.3).

The analysis of the opinion aggregation functions is concluded with a computational assessment of the computational cost of aggregating collective opinions that indicates that implementing the aggregation functions in real-sized debates would be feasible.

# Part III

# Abstract multi-agent debate

# Chapter 9

# Modelling a generalised debate

In the previous parts, this research studied collective reasoning on two different models for a debate, the Target oriented discussion framework (TODF) in part I and the Relational Model (RM) in part II.

Part I and II defined and studied two specific multi-agent debate models, the TODF and RM, each one giving a particular semantic interpretation of a debate and its components. This last part goes beyond a specific interpretation for a debate by exploring an even more general model that can embrace many different kinds of discussions, including the TODF and the RM. Within this general model, this part leaves aside the research on collective decision methods to introduce an approach to analyse the quality of a multi-agent debate.

This chapter introduces the *Abstract multi-agent debate*, in short AMAD, a model aiming to capture the essential features of a multi-agent debate to study, under the same umbrella, many interpretations for a debate and its corresponding notion of coherence. Next, Chapter 10 shows the applicability of AMAD to represent other models and introduces a method to analyse the quality of a multi-agent debate. Finally, in Chapter 11 we summarise and discuss the findings of this part.

In the following, next section introduces the main features of AMAD. Section 9.2 presents formally AMAD. Section 9.3 defines the corresponding notion of coherence for AMAD. Finally, Section 9.4 reviews the contributions of this chapter.

## 9.1 Introduction

The purpose of the AMAD is to extend the RM into an even more abstract model that captures the essential features of many types of debate models. This abstraction allows a more general study of multi-agent debates, in particular, to represent many different kinds of a debate using solely the AMAD model.

As might be expected, each type of debate model can associate different interpretations and behaviours with each of its components. However, we can extract the essential properties they share in common and create a model that embodies the essence of a debate.

Looking at different types of multi-agent debate models —for instance, these reviewed or developed in our research. Namely, the Argumentation Framework (AF) [Dung 1995] with labelling system from [Awad et al. 2015], the TODF developed in part I or the RM in part II—, they share common core characteristics. They are composed of several pieces of information —which can either represent arguments, statements or other concepts— that connect by means of relationships –in some cases, attack, defence or even reasoning relationships like in the AF, TODF or the RM, respectively. Thus, the information that makes up a debate can be organised using a graph or hypergraph-like structure using nodes and relationships.

Furthermore, these elements are susceptible to being evaluated by the participants in the discussion. Whether this evaluation process uses labellings, voting or valuation functions on the arguments or statements, or even attack, defence or acceptance attributes on the relationships, the agents' opinions are issued over the elements forming the discussion.

In conclusion, a multi-agent debate is essentially formed by *structure*: interconnected information; and *opinion*: the agents' evaluations over the structure. These form the elemental configuration of a multi-agent debate, and they will constitute AMAD.

## 9.2 Abstract multi-agent debate

This section formally introduces the *Abstract multi-agent debate* (AMAD). We emphasise its abstract and general treatment of a debate, i.e., we do not intend to describe a

new model with its implicit semantics and interpretations restricting every object to a defined and precise concept. This general approach to a debate allows us to uniformly represent different types of debate models using the same basic elements and thus, create generic tools that provide insight into how a debate performs.

Since the work on the RM has inspired this abstraction, it is not unexpected that they have many characteristics in common. AMAD follows the path started with the RM by maintaining a clear distinction between what is the structure that organises the discussion and what are the opinions issued on the structure. The structure does not contain or represents the opinion from the participants. Differently to many particular models that have elements combining together opinions and structure —such as the AF or the TODF, where an attack relationship merges altogether a connection (structural object) with an attack interpretation (a subjective evaluation)— this sharp distinction between these two components is a centrepiece of AMAD.

### 9.2.1 The structure of AMAD

The debate structure of AMAD aims to represent the information (knowledge, facts, beliefs,...) abstractly, in many forms, shared by the participants. To do so, it restricts the structure to characterise only the principal features relating to the information contained in a debate. Thus, the structure of AMAD is composed by *nodes* and *relationships*. The nodes are the abstract representation of arguments, statements or any other type of information that can be connected to other information by some connection, may it be attack, defence, reasoning or another kind, which the relationship set will represent.

To maintain order in the structure, the relationships are *directed*, i.e. each connection has a start and an end. Therefore, the relationships connects nodes *towards* other nodes, instead of nodes *with* other nodes. As can be seen in our research, depending on the model, the direction of the relationship may vary according to the interpretation given to it. In the TODF, each relationship connects one argument to a target argument being attacked or defended by the first. Hence, the direction of the attack and defence goes against the chronological order of the debate, i.e. a new argument enters the discussion pointing to an old one. Contrarily, the RM goes the other way around. The relationship establishes the chronological order of the statements in the debate and the logical path to navigate through them. Each relationship represents some kind of rea-

soning, from a set of initial statements towards a single final statement. AMAD, like the RM, sets the relationships to point from a set of nodes toward one single node and also allows the structure to have multiple relationships to exist between the same initial and final nodes. As will be seen in the next section, the direction of the relationship will indicate from which nodes to gather the opinion for coherence purposes.

**Definition 9.2.1** (Debate structure)**.** A *debate structure* $\mathcal{S} = \langle N, R \rangle$ is a structure formed by a set of nodes $N$, representing pieces of information in a debate, and $R \subset P(N) \times N \times \mathbb{N}$, its relationships between nodes that represent their connections, satisfying:

- *Contingency*: for any $n \in N$ and any $c \in \mathbb{N}$, $(\emptyset, n, c) \notin R$.
- *Indirect connection*: for any two nodes $n_1, n_2 \in N$:
  - $n_1$ is connected with $n_2$, i.e., there is a relationship $(\Sigma, n, c) \in R$ such that $n_1 \in \Sigma$ and $n = n_2$ or $n_2 \in \Sigma$ and $n = n_1$; or,
  - there is a node $n \in N$ that *connects indirectly* $n_1$ and $n_2$. That is, there is a path connecting $n_1$ with $n$ and another path connecting $n_2$ with $n$[1].

We introduce some notation to improve presentation. A relationship $(\Sigma, n, c) \in R$, where $\Sigma \in P(N)$, $n \in N$ and $c \in \mathbb{N}$, will be written as $\Sigma \overset{c}{\succ} n$. We say that $n$ is *initial* in $r$ if there are $\Sigma \subset N$ and $n' \in N$ such that $r = \Sigma \overset{c}{\succ} n' \in R$ where $n \in \Sigma$ and it will be noted as $n \succ_r$. We say that $n$ is *final* in $r$ if $r = \Sigma \overset{c}{\succ} n \in R$ for some $\Sigma \subset N$ and this being noted as $\succ_r n$.

We note that the previous definition is very similar to the definition 6.2.1 of a relational framework. Similarly to the RM version, we define a relationship as an element of $P(N) \times N \times \mathbb{N}$. This definition allows the structure to represent many types of relationships, from the AF relationships [Dung 1995] (subset of $N \times N$) to the relationships in RM ($P(N) \times N \times \mathbb{N}$). The third component allows the structure to have different relationships that share the same initial and final nodes. However, unlike the RM, which defined the $DRF$ in definition 6.2.2 with the purpose of directing the discussion toward a set of targets, the debate structure of AMAD is less restrictive. In particular, AMAD

[1] A path connecting a node $a_0$ with a node $a_k$, or vice versa, is a sequence of relationships $r_1, r_2, \ldots, r_k$ such that $r_i = (\Sigma_{i-1} \cup \{a_{i-1}\}, a_i, c_i) \in R$.

allows cycles, while RM does not. Chapter 10 will present some examples that show that a debate model based on "vertices" and "edges", i.e. with a graph or hypergraph-like structure, is extended by the debate's structure and, therefore, can be expressed using the AMAD model, that is, with the opinion functions defined next.

Although the features chosen for the relationship (neutral relationship, directed from a set to one node, multiple connections between the same ends, etc.) may seem to prevent us from properly generalising models such as the Argumentation framework or the TODF, this is not the case. As will be seen in Chapter 10.1, each model can be generalised into the AMAD model preserving the properties that characterise them. Moreover, such translation into the AMAD model may enable us to distinguish better some of the intrinsic behaviour of each model that originally are more hidden, such as the nature of the attack relationship that, when represented in AMAD, clarifies the distinction between negative impact and connection between arguments.

### 9.2.2 The opinions

As has been established, the participants' opinions represent their particular views about the information shared in the discussion. Their opinions, printed on each piece of the structure, supply the discussion with the individuals' subjective points of view.

As in RM, the opinion of each agent is represented as two functions mapping each element of the structure, node or relationship, to a value. The function mapping nodes to values is named *node function* and the function mapping relationships to values is named *relationship function*. Given the difference between nodes and relationships, the opinions on them are allowed to be different, thus allowing the two functions to map into different sets of values so they can represent distinct types of opinions.

**Definition 9.2.2** (Opinion)**.** An agent opinion $O = (v, w)$ over a debate structure $\mathcal{S} = \langle N, R \rangle$ is a pair of functions where $v : N \longrightarrow I$ is a called the *node function* and $w : R \longrightarrow J$ is the *relationship function*, being $I, J \subset \mathbb{R}$ two sets of values.

We notice that this definition generalises the definitions of valuation (Def. 6.2.3) and acceptance (Def. 6.2.4) functions, for the RM. In this model, though, the model only restricts the opinion to map each element to a value from an undetermined set of real values, not a specific set of real values (discrete or continuous) nor any other

type of values (e.g. labels). We do not specify the sets $I$ and $J$, so they can represent very different types of opinions depending on the model we want to represent with AMAD. Either we could use a discrete set of real values to represent finite states of the opinion —which may be more appropriate to represent a model using labellings—, or we could use a continuous-value set for a more precise opinion —like in the RM. The only characteristic that we clearly distinguish is an agent's ability to evaluate a node and a relationship differently, which may differ on their evaluation method and thus on the set used for their respective functions. The role of the opinion is to provide the individual's evaluations of the information of the debate, i.e. each agent's point of view, but it also can serve to represent shared opinions on some elements of a debate. As an example, the notion of attack between nodes, as it is understood in the AF, can be represented by imposing a unique evaluation of the relationships for all the participants' opinions.

Finally, we introduce the general multi-agent debate formed by the two defined components, the structure and the opinion.

**Definition 9.2.3** (Abstract multi-agent debate)**.** Given $m$ agents and a debate structure $\mathcal{S} = \langle N, R \rangle$ we call an *Abstract Multi Agent Debate* the pair $\langle \mathcal{S}, \mathcal{O} \rangle$ formed by a debate structure $\mathcal{S}$ and a collection of opinions $\mathcal{O} = (O_1, \ldots, O_m)$ where each $O_i = (v_i, w_i)$ is the opinion of agent $i$ over the structure $\mathcal{S}$.

In other words, AMAD is formed by a debate structure, nodes that relate via relationships, and the opinions of the agents participating in the debate that provide the individuals' subjective views about the information provided in the structure. Choosing carefully how to translate the features of other models, many types of debate may be represented using this AMAD formalisation so then they can be the object of study using tools developed for AMAD. The following example illustrates the applicability of AMAD by translating a specific and well-known framework.

**Example 9.2.1.** In this example we illustrate how we can express the framework from [Awad et al. 2015] using the AMAD model.

Awad et al. use Dung's abstract argumentation framework (AF) to capture debate information using arguments, attacks between arguments, and a labelling system ([Caminada 2006]) with three labels (`in`, `out` and `undec`) to encode the opinions of the

participants. The next concepts relating the AF from [Awad et al. 2015] are formally introduced in Section 2.1.1 from Chapter 2.

The abstract argumentation framework is a structure $AF = \langle AR, Att \rangle$ formed by the arguments, $AR$, and the attack relationship between them, $Att \subset AR \times AR$. Taking as nodes the arguments, $N_{AF} = AR$ and as relationships the attack relationship, $R_{AF} = \overline{Att}$ where $\overline{Att}$ is the set of reversed attack relationships —i.e. $\overline{Att} = \{(b, a) \mid (a, b) \in Att\}$[2]. Thus, the resulting structure in AMAD is $\mathcal{S}_{AF} = \langle N_{AF}, R_{AF} \rangle$. This process was straightforward because the $AF$ was almost a structure fitting for AMAD.

Furthermore, each agent $i$ has a labelling function $L_i : AR \longrightarrow \{\texttt{in}, \texttt{undec}, \texttt{out}\}$ to express their opinion on the arguments —`in` to accept an argument, `out` to reject it, and `undec` to express indecision about it. As each $L_i$ provides opinions for the argument set, we can relate it directly to the node evaluation functions $v_i : N_{AF} \longrightarrow I$, which give values to the nodes previously seen as arguments. To preserve the symmetry of the labels, `in` is the opposite of `out` and `undec` is the middle term between `in` and `out`, we choose $I = \{1, 0, -1\}$ to respectively translate $\{\texttt{in}, \texttt{undec}, \texttt{out}\}$[3].

Finally, as we do not assume the negative interpretation of the attack relationship in our relationship set, $R_{AF}$, we will represent it via the relationship evaluation functions. Each $w_i$ will capture the attacking nature of the relationships by assigning the value $-1$, $w_i = w : R_{AF} \longrightarrow \{-1\}$, which could be understood as "all the relationships act negatively"[4]. Furthermore, we see that all participants share the attack nature of any relationship, i.e. $w_i = w$ for any agent $i$. Thus, the opinion for each participant $O_i = (v_i, w)$ is formed and the translation into the AMAD format is $AMAD_{AF} = (\mathcal{S}_{AF}, \mathcal{O}_{AF})$, where $\mathcal{O}_{AF} = (O_1, \ldots, O_m)$.

From this example we can extract different conclusions. First, the AMAD abstraction indeed can be used to express a model with a specific interpretation for its elements, though we might have to change the form in which those interpretations restrict the elements. In this case, the characteristics of the notion of attack are captured by both the

[2]Recall that the direction of the relationship in AMAD is the reverse of the attack direction.

[3]Many other sets could be used to represent $\{\texttt{in}, \texttt{undec}, \texttt{out}\}$. The importance of using specific values to represent the opinions comes into play when we define particular methods that have to use these values for computations that must relate to particular interpretations of the debate.

[4]As long as the value given to the relationship fits the interpretation intended for the debate, other values could be given to them.

relationships (structural attribute of an attack) and the relationship function (subjective attribute of an attack). Second, the translation into AMAD spells out the difference between opinion and structure in a relationship, in this case the attack relationship, clearly distinguishing what a relationship is and what an opinion that gives a negative impact to the relationship is.

## 9.3 Coherence in AMAD

Besides representing the participants' view on the debate, the opinion also serves as means to operate with its values and perform analyses of the debate. For instance, in [Awad et al. 2015] or the TODF, the opinion expressed by labelling functions is used to determine the consistency of the participants (may it be using the notion of complete labelling or coherence) or to create collective decision operators. Since one of the primary features of a debate is the relationships between the different pieces of information along the discussion, it is natural to attempt to extract meaning from them.

This type of analysis uses the intrinsic connections in the discussion, namely, the relationships, to derive influence among the opinions. This way, one node's opinion can impact another node's opinion for some method or analysis. This is called *dependency*, an opinion which depends on another. In other words, dependency represents the concept that an opinion issued in one place matters for the opinion in another place because it is linked from one to the other. In AMAD, considering the direction of the relationship, the dependency between nodes will act in the reverse direction, like in the RM. This is, given a relationship, the opinion of the final node of the relationship may affect the opinion of the initial set of nodes. The use of dependencies is central to characterise a general notion of coherence in AMAD. These dependencies indicate which opinions can interfere with other opinions and, therefore, are the basis for studying the consistency between them. Given a single node, the opinion issued to that node is called *direct opinion* and the set of opinions regarding its dependencies is called *indirect opinion*.

Given an opinion $O$, we say that the opinion value issued to a node $n \in N$ is the *direct opinion* of $n$. The values issued to the relationships coming from $n$ together with the values issued on their attached nodes form the *indirect opinion* of $n$. Formally, we use the following notation to define the previous concepts. We represent the relation-

ships coming from $n$ as:

$$R^+(n) = \{r \in R \mid n \succ_r\},$$

i.e., the set of relationships $r$ such that $n$ is initial in $r$. The set of descendants of $n$, $D(n)$, is the set of nodes that relate directly to $n$ by a relationship, i.e.,

$$D(n) = \{n_r \in N \mid \exists r \in R^+(n) \text{ s.t. } r = \Sigma \overset{c}{\succ} n_r\}.$$

Following the previous notation, we define the direct and indirect opinions.

**Definition 9.3.1** (Direct and indirect opinion)**.** Given an opinion $O = (v, w)$ the *direct opinion* of $n$ is $v(n)$ and the *indirect opinion* of $n$ is

$$IO(n) = \{(v(n_r), w(r))|\ r \in R^+(n) \text{ and } n_r \in D(s) \text{ s.t. } \succ_r n_r\} \subset (I \times J)^{|R^+(n)|}.$$

Thus, the indirect opinion is the collection of opinions attached to the nodes and relationships descending from a node, grouping each relationship evaluation with its respective node evaluation.

The interrelated structure of AMAD, i.e. nodes connected by relationships, is used to derive dependency between the opinions. The purpose of coherence is to determine whether the direct and indirect opinions are in line or not, i.e. if the indirect opinions support the direct opinion. To do so, an auxiliary function whose purpose is to approximate a value representing the indirect opinion has to be defined, named *estimation function*. As in the RM, we base the definition of coherence on a general function aimed to capture the general opinion on the dependencies. The estimation function provides a representative value for the indirect opinion that can be compared to the direct opinion and, therefore, determine whether or not there is coherence in that node.

**Definition 9.3.2** (Estimation function)**.** Given an $AMAD = \langle \mathcal{S}, \mathcal{O} \rangle$, and $O$ an opinion over the structure $\mathcal{S} = \langle N, R \rangle$. The *estimation function* is a node function mapping each node to a value in the set $I$ such that:

$$\begin{array}{rccc} e: & N & \longrightarrow & I \\ & n & \longmapsto & e(n) \end{array}$$

such that:

- if $I(n) = \emptyset$ then $e(n) = 0$; otherwise,

- $e(n) = f(IO(n))$, i.e. is the result of computing the values from $IO(n)$.

Namely, for a node $n$ the estimation function computes a value from the set $IO(n)$ by means of an aggregation function $f$.

We notice that the previous definition is analogous to the definition 6.3.2 from the RM part. Following the general character of AMAD, the estimation function can be adapted depending on the debate at hand. The estimation function can be designed to approximate a value for the indirect opinion accordingly to the interpretation intended for each debate. Therefore, the previous definition characterises a broad family of estimation functions rather than a specific function.

Given an estimation function designed to estimate a value for the indirect opinion of a node, coherence uses it to evaluate the consistency of the opinion. The purpose of coherence is to assess whether the direct and indirect opinions via the estimation function are in line, i.e., whether their representative values are close to each other. Namely, an opinion will be considered coherent with respect to one node when the value of its direct opinion is similar to the estimated value of its indirect opinion. Formally,

**Definition 9.3.3** ($\epsilon$-coherence)**.** Let be $AMAD = \langle \mathcal{S}, \mathcal{O} \rangle$, $O$ one of its opinions, $e$ an estimation function and a value $\epsilon \in \mathbb{R}$. We say that the opinion $O$ is $\epsilon$-coherent at $n \in N$ when if $D(n) \neq \emptyset$ then:

$$|v(n) - e(n)| < \epsilon$$

In general, we say that an opinion $O$ is $\epsilon$-coherent if it is $\epsilon$-coherent for every node in $N$. We will denote as $\mathbb{C}_\epsilon(\mathcal{S})$ the class of all $\epsilon$-coherent opinions over $\mathcal{S}$.

The value $\epsilon$ is the tightness parameter controlling how close the direct and indirect opinions must be to be considered coherent. A small value for $\epsilon$ will produce a more strict notion of coherence, while a bigger value will be easier to fulfil.

As in other work such as [Dung 1995, Caminada 2006, Awad et al. 2015] or in the TODF and the RM, in AMAD, the dependencies between nodes are used to assess the "correctness" or "rationality" of an opinion, i.e. its consistency. In each model, each one having its own semantic, consistency is characterised (and named) differently. Still, in all cases, the notion of consistency aims at the same purpose: to determine if an opinion is more or less acceptable depending on the connections or relationships among them.

AMAD aims to generalise the main features of a debate, so the notion of coherence works on any type of debate that can be represented using AMAD. Although, the estimation function has to be defined accordingly to the semantic approach to each model so the coherence notion fits properly the interpretation given to the specific discussion. As we will present through different examples in Chapter 10, when choosing the correct estimation function, we can use coherence to characterise the notion of consistency originally defined for other models. For instance, when translating the Argumentation Framework into AMAD [Dung 1995, Caminada 2006], there is an equivalence between an opinion being coherent and a labelling being a reinstatement labelling, thus proving that coherence on AMAD is a basic characterisation for consistency for many types of a multi-agent debate.

## 9.4 Summary

This chapter introduced the AMAD model, generalising the RM from part II. Following, we summarise the contents of each section.

- Section 9.2 formally introduced the AMAD. As in the RM, the AMAD is formed by two elements, structure and opinion. The debate structure organises the debate in the form of nodes and relationships between them, without any given interpretation aside from their structuring purpose. Unlike the RM, AMAD does consider a target of the debate nor restricts the relationship to be acyclic. The opinion in the AMAD, to represent the agents' opinions, is represented by a node function mapping each node to a value from a set and a relationship function mapping each relationship to a different set of values. AMAD is aimed to be a model that generalises several other more specific multi-agent debate models so we can study in general terms a multi-agent debate on it.

- Section 9.3 defines the notion of $\epsilon$ coherence for AMAD. In the same way that in the RM, we use an estimation function to represent the indirect opinion with a value to compare it to the direct opinion. An opinion is considered $\epsilon$-coherent when both values are similar with respect to $\epsilon$.

This chapter answers positively the following research questions, introduced in Chapter 1.

RQ-1 The AMAD is a new representation of a multi-agent debate using nodes, relationships to structure the debate and two functions to represent the agents' opinions. Especially, AMAD is a general model capable of representing other more specific multi-agent debate models, such as the Abstract argumentation framework [Dung 1995, Awad et al. 2015] or the TODF (shown in the next chapter).

RQ-2 Following the same path of the RM, we also introduced the $\epsilon$-coherence for AMAD, which also characterises the consistency of an opinion in a flexible way and depending on the estimation function we choose.

# Chapter 10

# Applicability

Next, Section 10.1 introduces the research of this chapter. Section 10.2 provides further examples of the ability of AMAD and its notion of coherence to represent other interpretations for a multi-agent debate. First, in Section 10.2.1 we show that using the adequate estimation function coherence serves to characterise the notion of complete labelling from [Awad et al. 2015] (equivalent to the complete extension in [Dung 1995] or the reinstatement labelling in [Caminada 2006], reviewed in Section 2.1.1). Sections 10.2.2 and 10.2.3, develop the AMAD adaptations from the TODF and RM, respectively, to then characterise their original notion of coherence by means of the abstract coherence defined in AMAD. Section 10.3 presents a new approach to the assessment of the quality of a debate, the Systematic incoherence analysis. Coherence will play a central throughout this chapter. Finally, Section 10.4 summarises the development of this chapter.

## 10.1 Introduction

AMAD captures many types of debates by including the essential features that a multi-agent debate should have. Thus, AMAD divides a discussion between structure and opinion. The structure organises the information using nodes and relationships. The opinion, from the agents, by means of the node evaluations and relationship evaluations. Respectively, they represent the information shared by the participants that organise the discussion and the subjective opinions issued about the information. Furthermore, a general notion of coherence for AMAD characterises consistency of the opinion.

As an abstract model, AMAD extends other particular models by being defined using less restrictive features and not restricting its components by specific interpretation of the debate. On the other way around, we can represent particular multi-agent debate models by restricting the features of AMAD so as to relate to a specific interpretation.

Although several characteristics of AMAD are different from other formal representations of a multi-agent debate (e.g. in [Awad et al. 2015, Coste-Marquis et al. 2007] or the TODF and RM), and therefore may seem incompatible, they are suitable to represent alternative approaches by capturing essential aspects of a multi-agent debate. In particular, consistency on these specific models can be characterised by using specific instances of coherence in AMAD, showing that the definition of coherence captures a general characterisation for consistency. Thanks to our notion of coherence, AMAD can be used to analyse the quality of a debate in multiple models. An analysis of a debate's quality studies the role of some features relating to some standards intended for the debate. In this research, we focus the analysis on one feature: the coherence of the opinions.

Coherence helps us determine the consistency of an agent's opinion, so it can be used to systematically analyse a debate in the search for specific nodes that are accumulating an excessive number of incoherent opinions. A participant being incoherent on one node can be considered an individual problem of the participant. However, several participants (above a given threshold) being incoherent in the very same place may indicate that the problem is shared among the participants. Systematic incoherence analysis determines these situations by uncovering those nodes of the debate that have *systematic incoherence* when an unusually large number of participants are deemed incoherent. Once these critical nodes of the debate are determined, the next step would be to examine each one to discover the cause of the problem. We perceive the possible causes as one of two kinds: redundancy of information or missing information.

## 10.2 Translating alternative debate models to AMAD

This section shows how AMAD relates to several types of multi-agent debates. In particular, we see how an AF [Dung 1995] with labellings [Awad et al. 2015], a TODF (part I) and RM (part II) are translated into AMAD, and, how AMAD's notion of coherence

is characterised for each one of them.

As can be seen in the previous Example 9.2.1 in Chapter 9, assuming consensus among the participants with respect to the relationship evaluations —by evaluating the relationships as $-1$— we can represent the AF structure. Relating to AF ([Dung 1995]) in some way, the TODF can be expressed as well in AMAD by changing the restrictions we apply: constraining the structure to be target-oriented, assuming the consensus on the relationships to be two-valued (attack and defence), and restricting the node evaluations to be discrete using 3 values. Section 10.2.2 shows this translation.

Besides the AF-related models, restricting AMAD to be "proposal-related" —i.e. such that any relationship comes from a unique node (the proposal)—, assuming consensus on the relationships using two values and restricting the node evaluations to issue three possible values, then we could represent the PAM model [Serramià et al. 2019]. Finally, the RM can be characterised in AMAD by constraining the structure to be target-oriented, where the target of the discussion is a set of nodes, and providing the opinion functions with the correspondent sets.

We note that taking an instance of a specific model and translating it into AMAD can be done without restricting some properties that characterise another debate model, which correspond to a specific interpretation of the debate —for example, the properties related to the target in a TODF (in definition 3.2.2). The properties of each model that correspond to a specific view of the debate may not be necessary in a translation to AMAD because they might not be relevant at the moment —for instance, the attacking nature of the relationship in the AF with labellings [Awad et al. 2015] is not relevant if the only purpose is to obtain the collective labelling using the majority function or to study the density of connections in the framework— or because the properties are already transferred into AMAD by making the translation —like the target-oriented properties of the TODF that, thanks to maintaining the same relationships between the nodes, would be automatically included in AMAD.

Next, we will see how the interpretation of the debate is important to apply a notion of coherence according to the model at hand. The estimation function, key to computing coherence, is the only element that needs to reflect the interpretation intended in the original debate model.

### 10.2.1 Translating AF with a labelling system to AMAD

As we have seen in Example 9.2.1, the translation from an AF to AMAD, as it is used in [Awad et al. 2015], is carried out in the following way.

- The structure of the debate $AF = \langle AR, Att \rangle$, composed by arguments and the attack relationship, is represented by the structure $\mathcal{S}_{AF} = \langle N_{AF}, R_{AF} \rangle$ where $N_{AF}$ represents the set of arguments and $R_{AF}$ the relationship between them as the reverse attack relationships from the $AF$. In this case though, $R$ only refers to the connections between nodes, not attack relationships. For simplicity, for an AF we can consider the relationships $r = \{n\} \overset{c}{\succ} n_r \in R_{AF}$ to be one to one $n \succ n_r \in R_{AF}$.

- Each labelling system $L_i : AR \longrightarrow \{\texttt{in}, \texttt{undec}, \texttt{out}\}$ is translated to a valuation function $v_i : N_{AF} \longrightarrow \{1, 0, -1\}$, by creating an alternative representation for discrete values. In this case we have `in` as $1$, `out` as $-1$, as long as they are opposite, and `undec` as $0$.

- Finally, the negative impact implicitly represented by the attack relationship is represented by the consensus on the relationship evaluation functions $w : R \longrightarrow \{-1\}$, which are the same for every agent.

Thus, $AMAD_{AF} = \langle \mathcal{S}_{AF}, \mathcal{O}_{AF} \rangle$ where $\mathcal{S}_{AF} = \langle N_{AF}, R_{AF} \rangle$ and $\mathcal{O}_{AF} = (O_1, \ldots, O_n)$, $O_i = (v_i, w)$, $1 \leq i \leq n$, is the translation into AMAD's format.

Next, we apply the notion of coherence to $AMAD_{AF}$ in such a way that it captures the concept of a *reinstatement labelling* from [Caminada 2006] (which is equivalent to the complete labelling from [Awad et al. 2015] or to the complete extension from [Dung 1995]). We recall from Chapter 2 definition 2.1.5.

Let $L$ be an $AF$-labelling. $L$ is a reinstatement labelling if it satisfies the following:

- For all $a \in AR$, $L(a) = \texttt{out}$ iff exists $b \in AR$ such that $b$ $Att$ $a$ and $L(b) = \texttt{in}$.

- For all $a \in AR$, $L(a) = \texttt{in}$ iff for all $b \in AR$ such that if $b$ $Att$ $a$ then $L(b) = \texttt{out}$.

Then we define an estimation function that is able to capture the interpretation given to the reinstatement labelling.

$$e(n) = \min_{r \in \mathbb{R}^+(n)} \{w(r)v(n_r)\}$$

where $\succ_r n_r$ (i.e. $n_r$ is final in $r$).

From this definition we can deduce that:

- a node $n$ has all its descendants $n_r$ valued as $-1$ (`out`) if, and only if

$$e(n) = \min\{(-1) \cdot (-1), \ldots, (-1) \cdot (-1)\} = 1$$

- a node $n$ has at least one descendant $n_r$ valued as $1$ (`in`) if, and only if

$$e(n) = \min\{(-1) \cdot 1, \ldots\} = -1$$

- otherwise $e(n) = 0$.

Therefore, if we set $\epsilon = 1$, then coherence determines exactly the condition on every node for the opinion to be equivalent to a reinstatement labelling[1], i.e., an opinion $O$ will be 1-coherent when for any $n \in N$

$$|v(n) - e(n)| < 1,$$

is equivalent to having a reinstatement labelling in the AF. Next, we prove this equivalence.

- An argument $n$ is labelled `in` if, and only if, for all $n_r$ such that $n \succ n_r \in R_{AF}$, $n_r$ is labelled `out`. That is to say that $v(n) = 1$ if and only if for all $n_r$, $v(n_r) = -1$, which is equivalent to $e(n) = 1$, and therefore $|v(n) - e(n)| = 0 < 1$.

- An argument $n$ is labelled `out` if, and only if, there exists $n_r$ such that $n \succ n_r \in R_{AF}$ and $n_r$ is labelled `in`. That is the same to say, that $v(n) = -1$ if and only if there is a $n_r$ that $v(n_r) = 1$, which is equivalent to $e(n) = -1$, and therefore $|v(n) - e(n)| = 0 < 1$.

With this presentation, we have established that both notions, coherence with that estimation function and the reinstatement labelling, can faithfully capture the same notion of consistency for AF.

[1] Actually, any $\epsilon \in (0, 1]$ could be used for such equivalence.

### 10.2.2 Translating TODF to AMAD

To translate the TODF into the AMAD, we follow the same steps as in the previous model.

First, we translate the TODF structure. As was the case with the AF translation, the argument set turns into the nodes set $N_{TODF} = \mathcal{A}$ and the relationship set is formed by union of the reversed attack and defence relationships $\overline{\mapsto \cup \Vdash}$, analogous to the AF translation, $R_{TODF} = \overline{\mapsto \cup \Vdash}$, thus $\mathcal{S}_{TODF} = \langle N_{TODF}, R_{TODF} \rangle$[2]. The target argument $\tau$, though it is not considered here, is not deleted as an argument and the properties relating it to the rest of the debate are incorporated automatically into the $AMAD_{TODF}$ structure by means of the intrinsic features of the relationships.

Then we translate the subjective attack and defence attributes of the relationships using the relationship evaluation functions of AMAD. Since all agents share the classifications of attack and defence relationships, each agent will have the same relationship function $w : R_{TODF} \longrightarrow \{1, -1\}$ mapping each relationship into one of two types: positive or negative impact. For every agent, if the relationship was a defence relationship then $w(r) = 1$, if the relationship was an attack relationship then $w(r) = -1$. And finally, the labelling system translates in the same way that in the first example with the AF. For each agent its valuation function is $v_i : N_{TODF} \longrightarrow \{1, 0, -1\}$ respectively matching the $1$, $0$ and $-1$ to the labels `in`, `undec` and `out`.

Next, we proceed to define an estimation function for $AMAD_{TODF}$ in order to propose a notion of coherence that will be formally equivalent to TODF's coherence from definition 3.4.3. We propose the function below to be the estimation function for $AMAD_{TODF}$:

$$e(n) = \frac{1}{|R^+(n)|} \sum_{r \in R^+(n)} w(r) v(n_r)$$

Where $\succ_r n_r$. Having the estimation function, for each opinion $O_i$ we can assess the $\epsilon$-coherence in $AMAD_{TODF}$ for some $\epsilon$ by checking the condition:

$$|v_i(n) - e_i(n)| < \epsilon$$

[2]We recall that the attack and defence relationships are disjoint sets, there is no need to differentiate between them.

Now we check the equivalence between the coherence in $AMAD_{TODF}$ and coherence in the TODF, using $\epsilon = 1$ and the previous estimation function. First, we recall definition 3.4.3.

A labelling is coherent if for all $n \in N_{TODF}$ with $A(n) \cup D(n) \neq \emptyset$:

(1) if $L(n) = \texttt{in}$ then $Pro_L(n) \geq Con_L(n)$;

(2) if $L(n) = \texttt{out}$ then $Pro_L(n) \leq Con_L(n)$.

Where $Pro_L(n) = \texttt{in}_L(D(n)) + \texttt{out}_L(A(n))$ the positive support of $n$ and $Con_L(n) = \texttt{in}_L(A(n)) + \texttt{out}_L(D(n))$ the negative support of $n$.

To check the equivalence between coherence in the TODF and its translation in AMAD, it suffices to see the case when $L(n) = \texttt{in}$, i.e. $v(n) = 1$ in $AMAD_{TODF}$. The other case is analogous.

If $L(n) = \texttt{in}$ then $Pro_L(n) \geq Con_L(n)$, which means that

$$\texttt{in}_L(D(n)) + \texttt{out}_L(A(n)) - \texttt{in}_L(A(n)) - \texttt{out}_L(D(n)) \geq 0.$$

Translating into $AMAD_{TODF}$ we can see that:

$$\texttt{in}_L(D(n)) - \texttt{out}_L(D(n)) = \sum_{n_r \in D(n)} w(r)v(n_r)$$

and

$$\texttt{out}_L(A(n)) - \texttt{in}_L(A(n)) = \sum_{n_r \in A(n)} w(r)v(n_r)$$

If $r$ is a defence then $w(r) = 1$ and if $r$ is an attack then $w(r) = -1$; and, if $n_r$ is $\texttt{in}$ then $v(n_r) = 1$, if $n_r$ is $\texttt{out}$ then $v(n_r) = -1$. If $n_r$ is $\texttt{undec}$ it does not count because $v(n_r) = 0$.

Therefore, as $D(n) \cup A(n) = R^+(n)$, we can express the same condition by using the estimation function defined for the $AMAD_{TODF}$:

$$e(n) = \frac{1}{|R^+(n)|} \sum_{n_r \in R^+(n)} w(r)v(n_r) > 0.$$

Now, notice that the estimation function $e(n)$ is bound to have values within the interval $[-1, 1]$ thanks to the weighting by the total of relationships of $n$. For this reason, choosing a value $\epsilon = 1$ is enough to ensure that the $\epsilon$-coherence condition in $AMAD_{TODF}$ is equivalent to the coherence in the $TODF$:

$L(n) = \texttt{in}$ leads to $e(n) > 0$ (and $e(n) \in [-1, 1]$), so

$$|v(n) - e(n)| = 1 - e(n) < 1 = \epsilon.$$

Thus, and repeating the analogous steps for the case $L(n) = \texttt{out}$, if $L$ in the TODF is coherent at $n$, then the respective opinion in AMAD is 1-coherent at $n$, using the estimation function above specified. Reversely, if the opinion in AMAD is 1-coherent, we want to see that $L$ is coherent. Let's see the case $v(n) = 1$ ($L(n) = \texttt{in}$), and conclude that $Pro_L(n) \geq Con_L(n)$. In this case:

$$|v(n) - e(n)| = |1 - e(n)| < 1$$

Which implies that $0 < e(n) < 2$. Knowing that $e(n) \in [-1, 1]$ then $0 < e(n) \leq 1$, but it suffices t know:

$$\frac{1}{|R^+(n)|} \sum_{n_r \in R^+(n)} w(r)v(n_r) > 0 \Longrightarrow \sum_{n_r \in R^+(n)} w(r)v(n_r) > 0$$

That is equivalent to $Pro_L(n) - Con_L(n) > 0$, thus having $Pro_L(n) \geq Con_L(n)$ as we wanted.

### 10.2.3 Translating RM to AMAD

This section translates the RM to AMAD. In this case, the translation of the model and its coherence notion will be straightforward due to the similarity between both models.

- First, we translate the RM structure, the directed relational framework $\langle \mathcal{S}, R, T \rangle$, which in this case is already separated from the opinion. The structure in AMAD will be $\mathcal{S}_{RM} = \langle N_{RM}, R_{RM} \rangle$ where $N_{RM} = \mathcal{S}$ and $R_{RM} = R$. Similar to the previous translation of the TODF, the AMAD's structure does not single out the targets ($T$), though the statements that were the targets in RM are still included as nodes in $N_{RM}$ and its correspondent properties are transferred through the relationships, which are the same.

- The opinions, which are defined in the same format as that of AMAD, are directly translated $\mathcal{O}_{RM} = \mathcal{O}$, and the functions for every agent are defined equally with $I = [-1, 1]$ and $J = [0, 1]$, so $v_i : N \longrightarrow [-1, 1]$ and $w_i : N \longrightarrow [0, 1]$. Thus, the AMAD translation of RM is $AMAD_{RM} = \langle \mathcal{S}_{RM}, \mathcal{O}_{RM} \rangle$.

- The estimation function is defined in the same way, thanks to the similarities of the model:

$$e(n) = \frac{1}{\sum_{r \in R^+(n)} w(r)} \sum_{r \in R^+(n)} w(r)v(n_r).$$

- Finally, since the estimation function is identical for both models, the notion of coherence in $AMAD_{RM}$ is exactly the same as in the RM.

As has been illustrated with the previous adaptations, the AMAD model indeed serves as a general model to represent different types of specific debates in terms of both organisation and interpretation. For those debates that do not separate opinion from structure, such as AF and TODF, AMAD can represent them equivalently, plus making explicit the different types of information that a discussion can hold.

Furthermore, the notion of coherence defined for AMAD is a general definition that can characterise consistency regardless of the model, that is, as long as the estimation function captures the interpretation intended for the debate. This way, several models can be studied based on the same coherence analysis. The abstraction of AMAD is adequate to create general methods that apply to many interpretations of a discussion.

## 10.3 Systematic incoherence

In this section we offer one method to analyse the quality of a debate on AMAD. We use the notion of coherence to detect structural problems in a debate. By analysing the coherence of the participants' opinions, we propose a method to find shared inconsistencies among the participants' that will pinpoint specific parts of a debate that might be deemed problematic. We will say that a node has *systematic incoherence* when a significant number (higher than a threshold) of the agents' opinions are being incoherent at that same node.

In AMAD, given that we can assess the coherence of each opinion, we can count how many participants are coherent on each node. We define a function $i$ below to determine the ratio of incoherent opinions per node. We assume that an estimation function $e$ has been established, according to some semantics, that allows us to analyse the coherence of opinions.

**Definition 10.3.1** (Degree of incoherence)**.** Given an $AMAD = \langle \mathcal{S}, \mathcal{O} \rangle$, where $\mathcal{S} = \langle N, R \rangle$, if $m = |Ag|$ the *degree of incoherence* is defined as:

$$\begin{aligned} i : N &\longrightarrow [0,1] \\ n &\longmapsto 1 - \frac{c(n)}{m}. \end{aligned}$$

where $c(n)$ counts the number of opinions being coherent at a node $n$. Therefore, $\frac{c(n)}{m}$ is the ratio of coherent opinions on node $n$, and reversely, the ratio of incoherence on node $n$ is $i(n)$.

Clearly, $i(n)$ gives a value between $[0, 1]$. A value of $0$ represents that all the agents are coherent in node $n$, and $1$ means that none are coherent. This last case of total incoherence in one specific node might indicate a problem in that specific node. The fact that a significant number of participants are coherent at one specific place cannot be regarded solely as a result of casual inconsistencies of the participants. It should be an indication of a possible structural problem in the discussion, one that caused such shared incoherence.

We can use the degree of incoherence to suggest nodes of the debate that should be further analysed to find if there is a cause of such shared incoherence. However, assuming that we have the degree of incoherence of each node, how do we decide which nodes to analyse first? Many different procedures could be defined to indicate which nodes should be examined first to analyse potential structural problems behind them. Next, we propose one method, *systematic incoherence*, which uses a threshold value to separate those nodes that should be analysed from those that should not.

**Definition 10.3.2** (Systematic incoherence)**.** Given $\lambda \in (0, 1)$ and $i$ the degree of incoherence, we say that a node $n \in N$ has *systematic incoherence* if $i(n) > \lambda$. The $\lambda$ value is called the *incoherence threshold*.

The next example illustrates the previous notions.

**Example 10.3.1.** Assume a debate with 10 participants where 4 agents are being coherent on a node $n_1$ and 7 participants are being coherent on a node $n_2$. The degrees of incoherence of node $n_1$ and $n_2$ are, respectively: $i(n_1) = 1 - \frac{4}{10} = 0.6$ and

$i(n_2) = 1 - \frac{7}{10} = 0.3$. If we set $\lambda = 0.5$, then $n_1$ has systematic incoherence. If we set $\lambda = 0.6$, then none has from systematic incoherence.

The incoherence threshold should be chosen carefully to provide a small set of nodes to examine. Systematic incoherence may indicate where to look in a debate for a problem, but it does not reveal its causes.

At this point, we recall that *coherence in one node* —or incoherence for this matter— arises from the agreementbetween the direct and indirect opinion at that node. Therefore, incoherence reflects an imbalance between the direct and indirect opinions of the node. Suppose incoherence is generalised among the agents so as to cause systematic incoherence. In that case, the imbalance between direct and indirect opinion is also shared, and it must arise from the part they all have in common, the structure of the indirect opinion, namely, the relationships and nodes relating to the indirect opinions. Given that the direct opinion is the unique value from which to compare a representative value for the indirect opinion, the structural problem must lie with the nodes and relationships from where the indirect opinion is extracted. The collection of relationships and descendants attached to that particular node prevents the participants from producing coherent opinions.

Given these circumstances, the cause for a systematic incoherence can arise from:

(i) Information Redundancy. Too many nodes or relationships are causing the node to have redundant information in its dependencies, affecting the indirect opinion of the participants. Redundancy can unbalance the indirect opinion of the participants by giving too much relevance to some information that should weigh less in the indirect opinion, thus preventing coherence in that node.

(ii) Missing Information. The reverse case arises when there are missing nodes or relationships. Information is missing in the debate preventing the participants from giving an opinion that truly reflects their point of view. The participants might issue their opinions on the dependencies of the node considering information that is not represented in the debate, and therefore, an external opinion that is not considered when assessing coherence. This lack of opinion on the relevant information can bias the indirect opinion and thus produce incoherence.

## 10.4 Summary

The following list summarises the contents of this chapter.

- Section 10.2 developed three other examples where specific multi-agent debate models, with specific interpretations, are translated into AMAD. The AF with labellings in Section 10.2.1, the TODF in Section 10.2.2 and the RM in Section 10.2.3. Furthermore, we proved that their respective notions of consistency: respectively, the complete labelling from AF, and the coherence from the TODF or the RM, can also be characterised using the notion of coherence in AMAD. Choosing the right estimation function and $\epsilon$, $\epsilon$-coherence can characterise many other notions of consistency.

- Finally, Section 10.3 introduced a new method to analyse the quality of a debate, Systematic incoherence analysis. This type of analysis uses coherence to detect nodes of the debate where the agents are excessively incoherent and, therefore, may indicate a structural problem in the debate. In particular, we deem as possible that the problem at the node is caused by information redundancy or, the reverse, missing information.

This chapter has contributed to assessing the quality of a multi-agent debate as posed by research question RQ-6. Indeed, this is the aim of the systematic incoherence analysis, which has been conceived to detect structural problems in a debate.

# Chapter 11

# Conclusion and discussion

Part III introduced the Abstract multi-agent debate (AMAD) an its applications in chapters 9 and 10 respectively. This chapter summarises and discusses the research presented in these two chapters.

## 11.1 AMAD and coherence

Chapter 9 introduced the AMAD followed by the coherence notion. The AMAD model captures the essential features that should possess a multi-agent debate model. The structure is formed by nodes and relationships that, respectively, represent generic information shared in the debate and the connections between them. The opinion of the debate is composed of the collection of individual opinions of the agents, their individual opinions over the structure, formed by two functions: the node function to represent the opinion on nodes and the relationship function to represent the opinion on the relationships. Thus, being a new multi-agent debate model, this research positively answers the research question RQ-1.

Separating structure and opinion is fundamental to distinguish between the elements shared by the participants and the individual points of view they can issue on it. In this sense, AMAD extends the Relational Model (RM) that applied the same distinction to increase the participants' expressiveness. By doing this distinction, AMAD contributes to clarifying the nature of the objects forming a debate, not leading to ambiguities caused by the combination of opinion and structural object into one single entity. Furthermore, thanks to its generalisation, AMAD can be used to translate a particular

debate model into its abstract format. In fact, AMAD can even translate those models that do not distinguish between structure and opinion and untie and clarify the two distinct natures. Example 9.2.1 and Section 10.2.1 relative to AF, or Section 10.2.2 relative to the TODF, illustrate this generality of AMAD.

Although AMAD's goal is to generalise different types of models, it is a framework where abstractly study multi-agent discussions, its abstract elements may sometimes hinder such developments. For example, the definition of the functions composing the opinion, which does not specify a certain type of values, or the hypergraph-like form of the relationships might be too general for certain studies. The lack of specificity could obstruct a particular analysis on AMAD that needed a specification on such undetermined objects. Nevertheless, such a problem can be avoided by studying on constraint versions of AMAD that restrict the features as needed —for example, performing an analysis on AMAD assuming the opinion functions to be discrete or by imposing the relationships to be one-to-one.

Coherence is then defined on AMAD to characterise an opinion's consistency in a generalised manner. As well as in the previous part II, it is a positive answer to the research question RQ-2. By using dependencies and modelling the influence that an opinion can produce on another opinion through their connections, coherence defines consistency in terms of the alignment between the direct and indirect opinions of a node. Thus, we capture the notion that an opinion on a node is coherent when the indirect opinion is in line with the direct opinion. To do so, coherence uses a generic function, the estimation function, whose purpose is to combine the values of the indirect opinion. This generic function allows coherence on AMAD to be applicable to many different interpretations of a debate as long as the estimation function is defined according to the semantics intended for each case. Thus, coherence can characterise consistency for many types of discussions. Furthermore, if the estimation function correctly captures the semantics wanted for a specific model, coherence is equivalent to the specific consistency notion of such models. This fact is exemplified in sections 10.2.1, 10.2.2 and 10.2.3 relating to the translations into AMAD of the respective models AF, TODF and RM, thus showing that coherence is an elemental characterisation for consistency in general terms. Thanks to that, coherence can be used to analyse a debate generally to extract conclusions for specific cases, which is key for creating the

Systematic coherence analysis.

We observe that, similarly to the definition of the AMAD, the general definition of coherence can lead to the same issue previously stated: the lack of specification, in this case regarding the estimation function, which can be an obstacle when performing narrow analyses involving coherence. In order to study coherence, an estimation function that can be computed (instead of a generic one) may be needed for particular models.

## 11.2 The AMAD and its applications

Chapter 10 presents a way to apply the AMAD and the generalised coherence defined on it. It shows the ability of AMAD and coherence to provide general methods and analyses that can be applied in many types of debates.

While some minor adaptation may be needed to represent other models, the AF with labellings from [Awad et al. 2015], the TODF from part I and the RM from part II can be expressed using AMAD. In these translations, the notion of coherence, choosing the right estimation function, is a notion clearly capable of characterising the consistency defined in these other models. This capacity of coherence makes the subsequent analysis of the quality of a debate even more useful for its applicability to any model that can be represented in AMAD.

The Systematic incoherence analysis presented in Chapter 10 answers research question RQ-6. It is a new approach to study the quality of a debate in the sense that it uses the participants' opinions to analyse a debate. More specifically, the participants' opinions serve to study the quality of the structure of a debate. Based on the AMAD model, coherence is used to detect the critical places of the debate that potentially suffer from a problem that may affect a participant's ability to produce accurate opinions.

By performing a systematic assessment of the participants' opinions, the analysis can determine which nodes of the structure have an abnormal amount of incoherent opinions. An unusual degree of incoherent opinions associated with one node can indicate a structural problem with that node that negatively affects the participants' opinions. These nodes said to suffer systematic incoherence have to be analysed more deeply to determine the cause of the structural problem and, if possible, solve it.

Two possible causes arise from the systematic incoherence on a node: (i) there is

redundant information attached to the node that is biasing the participants' opinion, or (ii) there is missing information that provokes an incomplete opinion regarding the node.

Systematic incoherence analysis contributes to the area by providing one of the few existing methods that study the quality of a debate in abstract terms. Furthermore, this analysis is a novel approach in the sense that it uses the participants' opinions as a source to analyse a debate. Other approaches, such as [Gómez et al. 2008, Gonzalez-Bailon et al. 2010, Aragón 2019], do not consider that agents' opinions are available to analyse a debate. The approach developed here opens new paths to explore using the opinion for similar purposes.

AMAD is an abstract model on which we can study and generally analyse several interpretations for debates at once.

# Chapter 12

# Conclusion, discussion and future work

This chapter concludes this dissertation. In brief, the contributions can be summarised as follows: we explored three different forms of modelling multi-agent debates; we defined and analysed many methods to aggregate the collective opinion in a discussion, and we explored a generalised method to assess the quality of a debate. Across the different models, aggregation functions and quality analysis, coherence has been a key notion accompanying this work.

The rest of this chapter explores these contributions in more detail in Section 12.1 and provides some pointers to possible future work on the topics concerning this research in Section 12.2.

## 12.1 Discussion

Our research was first motivated to discover those approaches to support e-participation systems and provide collective reasoning methods to them. However, our research evolved into a more abstract and ambitious investigation aiming to embrace more complex and general questions regarding a multi-agent debate. The three-part work presented here illustrates this evolution. Part I develops a simple but practical framework to model a multi-agent debate, the Target-Oriented Discussion Framework, and several opinion aggregation functions to gather the participants' points of view in a debate. Part II goes beyond the previous approach and explores a richer model, the Relational Model

(RM), to improve the expressiveness of the participants in a debate and provide more accurate methods to gather the collective opinion. Finally, part III generalises it all by providing an abstract paradigm, the Abstract multi-agent debate, where to study many forms of multi-agent debates and develop a novel methodology to assess the quality of a debate.

At this point, we can discuss in more detail the answers to the research questions guiding this thesis, as introduced in Chapter 1.

**RQ-1** ***Can we find new models to represent a multi-agent debate?***

The three models developed in this dissertation, TODF, RM and AMAD, offer three separate answers to this question. Each one share similar perspectives on how to conceive a multi-agent debate. In particular, the following features are common in all three models:

- *Formalisation of a complete debate*. The formalisation of each model aims to represent an entire debate rather than provide a detailed process on how to create it. This is, each model presumes complete access to all knowledge and opinions resulting from the participation process in order to study collective reasoning methods. Though some guidelines on how to produce the TODF and the RM were provided, and the quality analysis on AMAD can lead to interferences in the deliberation process, the aim of the models is to represent a whole debate from where to study the collective opinion.
- *Directed and interconnected discussion*. In each model, the pieces of information, represented by different elements (arguments, statements or nodes), are interrelated. Using different types of directed relationships —attack or defence in the TODF, reasoning relationship in the RM, or an abstract relationship in AMAD— each model provides some structure to the debate and represents the connections between the different pieces of information. This feature allows deriving the influence from one object to another, in particular, to use the dependencies of the debate.
- *Participants express their opinions*. Further to structuring the information of the debate, all three models allow the participants to issue their opinions

on the content of the debate. Either by using a labelling system or functions, the participants' opinions are attached to the information structuring the debate. This is how the debate can capture the individual and subjective points of view from the participants, making it a "multi-agent" debate.

- *No rationality assumption*. Though the notion of coherence is used to characterise the consistency of an opinion throughout this research, neither of the models constrains the participants' opinions. The participants are not bound to express their opinions rationally or consistently, for humans can discuss without being rational. It would be a big assumption to meet in real life.

In addition to the previously shared perspective, each model represents a different point of view and therefore is formalised differently. In the following, we present the key elements corresponding to each model:

- *Target oriented discussion framework*

  The TODF is a novel framework that extends the abstract argumentation framework ([Dung 1995]) used by [Awad et al. 2015], which structures a debate by means of abstract objects representing the *arguments* that the participants provide to attack other arguments via an *attack relationship*. The TODF defines the additional *defence relationship* to express the direct support from one argument to another. Furthermore, a *target* argument is set to be the origin of the entire discussion, thus acting as the debate proposal and the discussion's final aim. Therefore, all relationships are constrained to point at the target argument in some manner —i.e. by directly or indirectly pointing at the target argument. The agents' opinions on arguments are qualitatively represented by employing the symbolic representation created in [Caminada 2006], the labelling system. Similar to the work in [Awad et al. 2015], for each participant, a labelling function encodes their opinions on each argument as accepted, rejected, or undecided.

- *Relational model*

Leaving aside abstract arguments and relationships with predefined semantics, such as attack or defence, the RM structures a debate using elemental *statements*, not intended to contain any reasoning, and a *relationship* aiming to represent the reasoning connecting them.

Similarly to the TODF, a set of statements is defined as the *targets* of the debate, and so, the root of the debate structure. Conversely to the TODF, the RM sets its debate to grow from the target statements towards the statements organising the debate.

The RM encodes two distinct types of opinions on the objects of the debate, distinguishing between opinion on statements and opinion on relationships, to represent different natures of subjectivity from the agents. Moreover, each type of opinion is represented by a real-valued function, allowing for a wider expressiveness from the participants than qualitative-based opinions such as those used in the TODF.

- *Abstract multi-agent debate*

  The AMAD extracts the essential features of a multi-agent debate so as to be a model capable of representing many interpretations for a debate, such as the TODF and the RM. AMAD structures a debate in a graph-like structure, common in many models, by means of *abstract nodes* and *abstract relationships* between them. This way, the model does not impose any predefined semantics on the structure of the debate. Besides the structure, AMAD includes the agents' opinions on the debate by means of evaluation functions over the structural objects, nodes and relationships, able to represent several types of opinion. Unlike the TODF and the RM, the AMAD model does not impose a special type of nodes to be the root of the debate, though it can represent debates with this particularity.

**RQ-2** ***Can we find a more flexible notion of consistency for an opinion?***

This research proposes the replacement of the classic notion of rationality or consistency, widely used in the literature (e.g. [Dung 1995, Leite and Martins 2011, Awad et al. 2015]), and exceedingly restrictive, putting forward a more flexible and realistic notion, *coherence*, defined for each model. While coher-

ence in the TODF or the RM is designed specifically for each model, in AMAD, the notion of coherence is intended to be a general characterisation for consistency, applicable to various multi-agent debate models. The complete extension from [Dung 1995] (reinstatement labelling from [Caminada 2006]), the coherence from the TODF and the RM can be captured using this general coherence in AMAD.

**RQ-3** ***Can we use dependencies to aggregate a collective opinion?***

Yes, we can and do use dependencies to obtain a collective opinion in several functions. This research provided means for aggregating a collective opinion from a debate on the TODF and RM models. In both models, we defined several functions to collect the participants' opinions into a single opinion. Departing from other work [Caminada and Pigozzi 2011, Awad et al. 2015, Chen and Endriss 2019], the opinion aggregation functions defined in this research aim to maximise the use of a debate's structure by exploiting the existing dependencies between the objects of the debate. In other words, the aggregation functions introduced here consider the existing relationships between opinions to produce a collective opinion. Thus, the aggregation functions this thesis presented exploit at different degrees the dependencies connecting the opinion resulting in a wide range of aggregation operators.

Regarding the TODF, the Opinion First function, the Support First function and the Balanced Function aggregate the participants' opinions by prioritising the direct opinion, the indirect opinion, or balancing both direct and indirect opinion, respectively.

Regarding the RM, two families of aggregation functions are defined to explore the exploitation of dependencies in different ways. The $\alpha$-Balanced family linearly combines the Direct function —a function that only aggregates the direct opinion— with the Indirect function — a function that only aggregates the indirect opinion. The $\alpha$-Recursive family linearly combines the Direct function with the Recursive function —which aggregates recursively the indirect opinion already computed with the Recursive function.

**RQ-4** ***Can we assess the aggregation functions taking into account the dependencies***

***in a debate?***

To answer this question, for the TODF and RM, we provide a list of properties to assess the proposed opinion aggregation functions. These properties stem from the Social choice literature [List and Pettit 2002].

Furthermore, taking into account the exploitation of dependencies intended for the aggregation functions, several new properties are defined to characterise other behaviours: Endorsed unanimity, to identify when the aggregation respects unanimity on the indirect opinion; familiar monotonicity to characterise monotonicity on aggregators exploiting dependencies, and collective coherence to assess the methods that output coherent opinions as a result of the aggregation. Thus, using these properties, we assess the aggregation functions considering the dependencies, so the answer to the research question is yes.

**RQ-5** ***Does considering the dependencies benefit in the aggregation?***

We exhaustively analysed each aggregation function with respect to the social choice properties they can satisfy. This analysis helps us understand each aggregation function's behaviour and compare them in terms of the social choice properties they fulfil.

The comparisons of aggregation functions, made on both TODF and RM, show that exploiting dependencies benefits the coherence of the collective opinion, which is the most desired property of an aggregation function. In particular, the Balanced function, which merges direct and indirect opinion equally, proves to be the best aggregation function for the TODF. Regarding the RM, in an unconstrained opinion profile —i.e. without assuming coherent profiles or consensus on the acceptance degrees—, the Recursive aggregation functions are the best choice since they can satisfy collective coherence in any scenario. Given this comprehensive analysis that shows how exploiting dependencies benefits the fulfilment of the collective coherence and others, the answer to the research question is yes.

Furthermore, the analysis of the aggregation functions is extended to their performance relating to their computational complexity. This study shows that our implementation of the aggregation functions can handle realistic scenarios.

**RQ-6** ***Can we assess the quality of a debate?***

Yes, we can assess the quality of a debate. Our research proposed means to measure the quality of a debate. Although several features may be analysed regarding the quality of a deliberative process, thanks to the notion of coherence, created in this research as a new way to characterise consistency, part III offered an approach focusing on the possible structural issues that a debate may suffer from.

Within the AMAD model, the Systematic incoherence analysis uses coherence to assess the participants' opinions and detect possible locations in a debate that are suffering from a structural problem. In our approach, we pinpoint problematic nodes that share too many incoherent opinions, which may need to be analysed to rectify the problem.

Overall, two general remarks stand out from this research. First, exemplified by the TODF and the RM: simplicity may be chosen for practical reasons but at the price of losing some accuracy. On the one hand, the TODF models a debate using simple elements (arguments, attack and defence relationships, a labelling system), which permits a more direct application of the model into real participation systems. Discrete opinions are easy to understand and work with but may not be accurate enough for an optimal representation of a human discussion. The RM, on the other hand, constructs a debate using a more complex and richer format (statements, reasonings and two real-valued opinion functions), allowing it to be more optimal for expressiveness and accuracy purposes. However, that may be less intuitive for practical uses in real scenarios. Either way, though, both models share the same elemental configuration, allowing the AMAD to extend both of them. They both represent a debate using connections among basic elements and opinions to express the participants' points of view. Thus, it is not this elemental configuration captured by AMAD that decides the practicality or accuracy of the model.

Second, considering the intrinsic connection of a debate improves coherence and quality. The dependencies in a debate, i.e. the connections that can deliver influence from one element to another, are closely related to the notion of consistency or rationality. Coherence, which in this dissertation characterises the notion of consistency for an opinion, serves to analyse the aggregation functions and even to distinguish those

methods that behave better. Additionally, coherence is the key element for analysing the quality of a debate, detecting structural problems within, and even providing some insight on how to solve them.

## 12.2 Future work

This research offers several paths that can be undertaken to further research in the area of multi-agent debates and collective reasoning.

The TODF, being the model that best relates to the current participation platforms due to its argumentative structure and qualitative opinions, offers many possible paths to explore regarding its application. In addition to implementing a participation system that fits perfectly the TODF model, which would be the most straightforward line of research to undertake, we also can study how to apply the TODF to platforms that use different formats to organise a debate. As an example of this, we have the comparison between the PAM model and the TODF in [Serramià et al. 2019] that shows an experiment applying both models to real debates data from [Decidim Barcelona], which are not in the TODF format. Many challenges can arise trying to apply the TODF to such participatory platforms: how to convert single lists arguments to nested structures, how to translate quantitative opinion into the qualitative labels of the TODF, what kind of relationship relate to non-categorised comments in a discussion, etc. Each platform would lead to different issues to solve in order to represent a debate using the TODF and, therefore, to apply the aggregation functions defined on it.

In its turn, the application of the RM to participation platforms even raises more challenges. Due to its unusual format to store the content of a debate, many issues would arise in such implementation. For instance, how to convert a complex argument into basic statements connected through reasoning relationships; how to translate generalised opinions —such as shared attack and defence relationships— into individual opinions for each participant; how to deal with acceptance degrees when no knowledge in the debate provides insight on how the participants value the reasonings of the model; and many more. A large amount of work can originate from these questions. However, such effort may pay off thanks to the richer format of the RM —for example, the sharp distinction between structure and opinion or the real-valued functions to represent the

opinions— which may clarify the contents of a discussion.

Regarding both the TODF and RM, before any real implementation is done, a study about the majority paradoxes and discursive dilemmas should be addressed in the future. Considering the concept of coherence and the dependencies that the aggregation functions use, this study should have interesting results. In the same way, a strategy-proof analysis of the aggregation functions should be in order. A study to determine how vulnerable an aggregation function is to manipulation would be very useful so we could apply suitable measures to prevent it. In fact, such a study would be closely related to the design of a platform using one of the models due to the many options that a platform can incorporate —such as visualisation of the debate relating to how restricted the information of the debate is to the users, the possibility to change the content already existing in a debate, etc.—, which may prevent or promote manipulation strategies of the debate. Concerning the aggregation process, we could continue exploring new forms to aggregate opinions in both models or even new forms to exploit dependencies in the aggregation process. Also, we could consider new operators incorporating methodologies from other works, such as the WOWA operator (Weighted Ordered Weighted Aggregation)[1], which could be suitable for the RM model.

On the other hand, specifically for the RM, an exciting line of research arises relating to the estimation functions that characterise coherence. The RM builds its structure on relationships representing reasoning steps in the debate. In this research, each reasoning is not intended to be a deduction from a particular logic, hence the general treatment for the estimation function used here, but it would be interesting to consider otherwise and consider a particular logic for this feature. Giving the RM a specific interpretation that restricts the reasoning to be a deduction from a specific logic would enable us to use estimation functions relating to the logic behind the debate and, therefore, offer a notion of coherence associated with the logic system too. As an example, understanding the reasoning in the RM to be fuzzy logic deductions could allow us to define an estimation function in terms of the t-norm[2] used in that logic system. It would be interesting to know if a logic-based coherence notion characterises some notion of consistency al-

[1] Operator used in [Serramià et al. 2019] to aggregate the opinions in the PAM model.

[2] A t-norm is a binary algebraic function representing the conjunction in a fuzzy logic semantics.

ready existing in the specific logic used.

Finally, part III offers more exciting paths to explore in the future. The AMAD model seems capable of being the ground from where to perform many different analyses of a multi-agent debate in general terms. For example, an analysis of the participation rate decay relating to the depth level in the debate —because it seems reasonable to expect that, as we go deeper into the debate structure, the ratio of participants that issued their opinion on the structure would decrease accordingly to some type of progression.

In fact, the Systematic incoherence analysis is a first step on the journey. This last piece of research came up late in this project and was included due to its remarkable value in illustrating how AMAD can be used to analyse a multi-agent debate generally. However, several items must be further examined on this topic. Each possible cause for systematic incoherence should be studied in-depth, and methods to handle these problems should be devised to fix the debate. This line of work seems a promising area to explore, and it will probably be the research path continuing this PhD project.

# Appendix A

# Proofs for the Target oriented discussion framework

In the following we prove the formal results of Section 4.4, summarised in Table 4.1, of the social choice properties that each of the aggregation function introduced in Section 4.3 satisfy. This chapter is divided into four parts, each one devoted to the results of each aggregation function in the following order:

1. Majority function ($M$),
2. Opinion First function ($OF$),
3. Support First function ($SF$),
4. Balanced Function ($BF$).

## A.1 Analysing the Majority function

We next analyse the properties from Section 4.2.1 fulfilled by the majority function. We check first the desirable properties and then the remaining properties.

**Proposition A.1.1.** $M$ satisfies Exhaustive and Coherent domain.

*Proof.* By definition, $M$ accepts every labelling profile as input, thus the domain of $M$ is $\mathcal{D} = \mathcal{L}(TODF)^n$. □

**Proposition A.1.2.** $M$ does not satisfy Collective coherence or Endorsed unanimity.

*Proof.* We prove it by providing the following counterexample.

Let $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, a\rangle$ with argument $b$ attacking target $a$, and with labelling profile $\mathcal{L} = (L)$, where $L(a) = \texttt{in}$ and $L(b) = \texttt{in}$. Computing the majority function over this profile we obtain the same labelling $M(\mathcal{L})(a) = \texttt{in}$ and $M(\mathcal{L})(b) = \texttt{in}$, and hence neither $L$ or $M(\mathcal{L})$ are coherent labellings.

Furthermore, argument $a$ receives full negative support though its final aggregated label is in. Thus, it is clear that $M$ does not fulfil Endorsed unanimity.

□

**Proposition A.1.3.** $M$ satisfies Anonymity and Non-Dictatorship.

The proof is based on the following lemma.

**Lemma A.1.1.** Let be a $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau\rangle$ and $\sigma$ any permutation of the set of agents $Ag = \{ag_1, \ldots, ag_n\}$. Let $\mathcal{L} = (L_1, \cdots, L_n)$ be a labelling profile and $\mathcal{L}' = (L_{\sigma(1)}, \cdots, L_{\sigma(n)})$ the labelling profile resulting from applying the permutation $\sigma$ over $\mathcal{L}$. The following equalities hold: $\texttt{in}_{\mathcal{L}}(a) = \texttt{in}_{\mathcal{L}'}(a)$ and $\texttt{out}_{\mathcal{L}}(a) = \texttt{out}_{\mathcal{L}'}(a)$.

*Proof.* Indeed, on the one hand $\texttt{in}_{\mathcal{L}}(a) = |\{ag_i \in Ag \,|L_i(a) = \texttt{in}\}| = |\{\sigma(ag_i) \in Ag \,|L_{\sigma(i)}(a) = \texttt{in}\}| = \texttt{in}_{\mathcal{L}'}(a)$. On the other hand, $\texttt{out}_{\mathcal{L}}(a) = |\{ag_i \in Ag \,|L_i(a) = \texttt{out}\}| = |\{\sigma(ag_i) \in Ag \,|L_{\sigma(i)}(a) = \texttt{out}\}| = \texttt{out}_{\mathcal{L}'}(a)$. □

Now we are ready to prove proposition A.1.3.

*Proof of proposition A.1.3.* The proof that $M$ satisfies Anonymity, and consequently Non-dictatorship, is a direct consequence of the previous lemma A.1.1. □

Besides Anonymity, $M$ trivially satisfies Unanimity. However, the satisfaction of Endorsed unanimity requires restricting the domain of $M$ to coherent labelling profiles, as we show below.

**Proposition A.1.4.** $M$ satisfies Unanimity.

**Proposition A.1.5.** $M$ satisfies Endorsed unanimity over $Coh_0(TODF)^n$, the class of $0$-coherent labelling profiles.

The proof of the proposition requires the introduction of an auxiliary lemma.

**Lemma A.1.2.** Let $L$ be a labelling giving full positive (resp. negative) support to an argument $a \in \mathcal{A}$. If $L$ is coherent then $L(a) \neq \texttt{out}$ (resp. $L(a) \neq \texttt{in}$). Furthermore, if $L$ is at least 0-coherent, then $L(a) = \texttt{in}$ (resp. $L(a) = \texttt{out}$).

*Proof.* Let $L$ be a labelling with full positive support on argument $a$, i.e, $Pro_L(a) = m$ and $Con_L(a) = 0$, where $m$ is the number of descendants of $a$ ($m = |A(a) \cup D(a)|$). Now we assume that $L$ is coherent. If $L(a) = \texttt{out}$ then, by coherence, $Pro_L(a) \leq Con_L(a)$ which cannot be. So we obtain a contradiction and therefore $L(a) \neq \texttt{out}$. Moreover, suppose that $L$ is 0-coherent, then it cannot be that $L(a) = \texttt{undec}$ because that would mean, by definition of c-coherence, that $|Pro_L(a) - Con_L(a)| \leq 0$, which is not the case. The proof goes analogously for the case with full negative support. □

Following this lemma, we can see that if a labelling $L$ is coherent but not 0-coherent on $a$, we cannot guarantee that $L(a) \neq \texttt{undec}$, which is necessary to prove the property for $M$, as the following proof will exemplify. Now we are ready to prove proposition A.1.5.

*Proof of proposition A.1.5.* Assume full positive support on an argument $a$, i.e, for every $i \in Ag$, $Pro_{L_i}(a) = m$, with $m$ being the number of descendants of $a$ ($m = |A(a) \cup D(a)|$). Analogously, we do the case with full negative support. By the previous lemma we can ensure that for every $i$, $L_i(a) = \texttt{in}$. Therefore, there is unanimity on the direct opinion of $a$, and hence $M(\mathcal{L})(a) = \texttt{in}$. □

In general, note from the previous proofs that for a positive value $c$ smaller than the minimum number of descendants that any argument $a \in \mathcal{A}$ can have, $0 < c < min_{a \in \mathcal{A}} |A(a) \cup D(a)|$, $M$ satisfies Endorsed unanimity over $Coh_c(TODF)^n$.

Interestingly, in what follows, we show that while $M$ satisfies restricted versions of monotonicity properties, it does not satisfy their non-restricted versions.

**Proposition A.1.6.** $M$ satisfies the Binary Monotonicity property, and consequently, Binary familiar monotonicity too.

*Proof.* Let $\mathcal{L}$ and $\mathcal{L}'$ be two labelling profiles satisfying the required hypothesis of Binary monotonicity setting with $l = \texttt{in}$. If $M(\mathcal{L})(a) = \texttt{in}$ we infer that $\texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a)$. Since $L_j(a) \neq L'_j(a) = \texttt{in}$ for the agents $j \in \{i, \ldots, i+k\}$, this entails that

$\texttt{in}_{\mathcal{L}'}(a) > \texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}'}(a)$, and hence $M(\mathcal{L}')(a) = \texttt{in}$. Analogously we prove the case $l = \texttt{out}$. □

**Proposition A.1.7.** $M$ does not satisfy:

(i) Familiar monotonicity,

(ii) therefore neither Monotonicity,

(iii) Supportiveness.

*Proof.* We prove these results through the example displayed in figures A.1 and A.2. Consider the $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, a \rangle$ with two labelling profiles $\mathcal{L} = (L_1, L_2)$ and $\mathcal{L}' = (L_1, L_2')$. The opinion of agent 1 is $L_1(a) = \texttt{in}$, $L_1(b) = \texttt{undec}$ for both profiles, and agent's 2 opinion is $L_2(a) = \texttt{out}$, $L_2(b) = \texttt{undec}$ for the first profile and $L_2'(a) = \texttt{undec}$, $L_2'(b) = \texttt{undec}$ for the second one (see figures A.1(a) and A.1(b)). The collective labellings using function $M$ are $M(\mathcal{L})(a) = \texttt{undec}$, $M(\mathcal{L})(b) = \texttt{undec}$ and $M(\mathcal{L}')(a) = \texttt{in}$, $M(\mathcal{L}')(b) = \texttt{undec}$, as shown in figures A.2(a) and A.2(b) respectively.

(i) Familiar monotonicity. Notice that the result of adding the undec label to the direct opinion of $a$ changes the collective labelling, disproving Familiar monotonicity.

(ii) Monotonicity. It follows directly that if Familiar monotonicity is not satisfied, neither is Monotonicity.

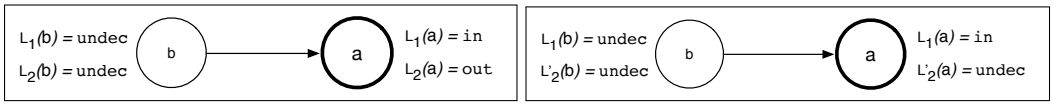
(iii) Supportiveness. Furthermore, we can see that by only using $\mathcal{L}$ and its aggregation (see figures A.1(a) and A.2(a)), Supportiveness is not fulfilled in this counterexample. This is because $M(\mathcal{L})(a) = \texttt{undec}$ though no agent's opinion is undec on $a$.

□

Finally, the satisfaction of Independence is trivial.

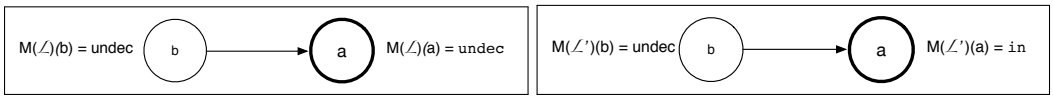
**Proposition A.1.8.** $M$ satisfies Independence.

*Proof.* Direct proof due to the fact that $\texttt{in}_{\mathcal{L}}(a) = \texttt{in}_{\mathcal{L}'}(a)$ and $\texttt{out}_{\mathcal{L}}(a) = \texttt{out}_{\mathcal{L}'}(a)$ for two profiles satisfying the hypothesis. □




(a) (b)

Figure A.1: Counterexample of proposition A.1.7: Agents' labellings, $\mathcal{L}$ on the left and $\mathcal{L}'$ on the right.




(a) (b)

Figure A.2: Counterexample of proposition A.1.7: Aggregated labelling by $M$ of the labelling profiles, $M(\mathcal{L})$ on the left, $M(\mathcal{L}')$ on the right.

## A.2 Analysing the Opinion first function

Following the same scheme used with $M$ here, we study the properties fulfilled by $OF$ starting with the desirable properties.

**Proposition A.2.1.** $OF$ satisfies the Exhaustive domain and Coherent domain properties.

*Proof.* By definition, the function accepts every labelling profile as input. □

**Proposition A.2.2.** $OF$ does not satisfy Collective coherence or Endorsed unanimity.

*Proof.* Given that $OF$ behaves exactly as $M$ when the labelling profile is composed by a single labelling without the `undec` label, the example of proposition A.1.2 also serves here as a counterexample for $OF$. □

**Proposition A.2.3.** $OF$ satisfies Anonymity and Non-Dictatorship.

*Proof.* The proof of Anonymity, and consequently of Non-dictatorship, is straightforward using lemma A.1.1 or the fact that $M$ satisfies them too.

If $OF$ uses the direct opinion, then $OF = M$, which already satisfies Anonymity and Non-dictatorship. If $OF$ uses the indirect opinion, then the collective label cannot depend on the individual labelling because it is computed using the aggregated labelling from the descendant, which has been already obtained using $OF$. □

In the same way as $M$, $OF$ trivially satisfies Unanimity, but satisfying Endorsed unanimity requires restricting $OF$'s domain to coherent labelling profiles, as we show below.

**Proposition A.2.4.** $OF$ satisfies Unanimity.

**Proposition A.2.5.** $OF$ satisfies Endorsed unanimity over $Coh(TODF)$, the class of coherent labelling profiles.

*Proof.* To prove this result we assume that there is full positive support for argument $a$, i.e, for every $i \in Ag$, $Pro_{L_i}(a) = m$, with $m$ being the number of descendants of $a$ ($m = |A(a) \cup D(a)|$), and that the labellings are coherent. Analogously, we can do the case when there is full negative support.

Since the labellings are coherent, by lemma A.1.2, for every agent $i$, $L_i(a) \in \{\texttt{in}, \texttt{undec}\}$, and hence $\texttt{out}_{\mathcal{L}}(a) = 0$. Now we differentiate two cases:

(i) At least one agent has labelled $a$ with `in`. If so, $OF(\mathcal{L})(a) = \texttt{in}$ because $\texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a)$.

(ii) All agents have labelled argument $a$ as `undec`. Then, $\texttt{in}_{\mathcal{L}}(a) = \texttt{out}_{\mathcal{L}}(a) = 0$. In this case, we must consider the aggregated opinion of argument $a$'s descendants. Full positive support on $a$ means that: (i) for each descendant $b_a$ attacking $a$, the agents must have labelled $b_a$ with `out`; and (ii) for each descendant $b_d$ defending $a$, the agents must have labelled $b_d$ with `in`. Therefore, the aggregated opinions will be $OF(\mathcal{L})(b_a) = \texttt{out}$ and $OF(\mathcal{L})(b_d) = \texttt{in}$ for the attacker and the defender respectively. Therefore $Pro_{OF(\mathcal{L})}(a) = m > 0$, and $Pro_{OF(\mathcal{L})}(a) > Con_{OF(\mathcal{L})}(a)$. Finally, since $\texttt{in}_{\mathcal{L}}(a) = \texttt{out}_{\mathcal{L}}(a)$ and $Pro_{OF(\mathcal{L})}(a) > Con_{OF(\mathcal{L})}(a)$, it turns out that $OF(\mathcal{L})(a) = \texttt{in}$, as wanted.

□

Again, similarly to $M$, we show that while $OF$ satisfies the restricted versions of monotonicity properties, it does not satisfy their non-restricted versions.

**Proposition A.2.6.** $OF$ satisfies Binary Monotonicity and, therefore, Binary familiar monotonicity.

*Proof.* Let $\mathcal{L}$ and $\mathcal{L}'$ be two labelling profiles satisfying the required hypothesis of the Binary monotonicity setting the label $l = \texttt{in}$. If $OF(\mathcal{L})(a) = \texttt{in}$, we infer that $\texttt{in}_{\mathcal{L}}(a) \geq \texttt{out}_{\mathcal{L}}(a)$. Since $L_j(a) \neq L'_j(a) = \texttt{in}$ for the agents $ag_j$, $1 \leq j \leq i + k$, it holds that $\texttt{in}_{\mathcal{L}'}(a) > \texttt{in}_{\mathcal{L}}(a) \geq \texttt{out}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}'}(a)$. Hence, $OF(\mathcal{L}')(a) = \texttt{in}$. The proof goes analogously for the case $l = \texttt{out}$. □

**Proposition A.2.7.** $OF$ does not satisfy Familiar monotonicity, Monotonicity, or Supportiveness.

*Proof.* The counterexample of proposition A.1.7 also applies here for $OF$. □

Finally, as we see in the next proposition, $OF$ pays the price for exploiting dependencies. Contrary to $M$, $OF$ uses the indirect opinion, i.e. its dependencies, to decide when there is a tie on the direct opinion, which leads to the loss of Independence.

**Proposition A.2.8.** $OF$ does not satisfy Independence.

*Proof.* To prove this result consider the example in figures A.3 and A.4. Let there be a $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, a \rangle$ with labelling profiles $\mathcal{L} = (L)$ and $\mathcal{L}' = (L')$ such that $L(a) = \texttt{undec} = L'(a)$ and $L(b) = \texttt{in}$, $L'(b) = \texttt{out}$ (figures A.3(a) and A.3(b) show $\mathcal{L}$ and $\mathcal{L}'$ respectively). $OF$ obtains the following labelling for $\mathcal{L}$ and $\mathcal{L}'$: $OF(\mathcal{L})(a) = \texttt{out}$, $OF(\mathcal{L})(b) = \texttt{in}$ (Figure A.4(a)); and $OF(\mathcal{L}')(a) = \texttt{in}$, $OF(\mathcal{L}')(b) = \texttt{out}$ (Figure A.4(b)). Clearly, Independence is not satisfied because $L(a) = L'(a)$, but $OF(\mathcal{L})(a) \neq OF(\mathcal{L}')(a)$.

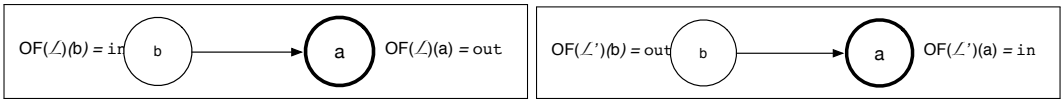

(a) (b)

Figure A.3: Counterexample of proposition A.2.8: Agents' labellings, $\mathcal{L}$ on the left and $\mathcal{L}'$ on the right.

□

## A.3 Analysing the Support first function

Following the same scheme as in the two previous sections, next, we study the social choice properties that $SF$ satisfies.




(a) (b)

Figure A.4: Counterexample of proposition A.2.8: Aggregated labelling by $OF$ of the labelling profiles, $OF(\mathcal{L})$ on the left, $OF(\mathcal{L}')$ on the right.

**Proposition A.3.1.** $SF$ satisfies the Exhaustive domain and Coherent domain properties.

*Proof.* By definition, $SF$ accepts every labelling profile as an input. □

Next, we prove the first positive result regarding Collective coherence. Thus, unlike $M$ and $OF$, $SF$ does satisfy Collective coherence.

**Proposition A.3.2.** $SF$ satisfies Collective coherence.

*Proof.* Let $a$ be an argument such that $SF(\mathcal{L})(a) = \texttt{in}$. From Definition 4.3.3 we know that $Pro_{SF(\mathcal{L})}(a) > Con_{SF(\mathcal{L})}(a)$ or $Pro_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L})}(a)$ and $In_{\mathcal{L}}(a) > Out_{\mathcal{L}}(a)$ holds. Then it follows that $Pro_{SF(\mathcal{L})}(a) \geq Con_{SF(\mathcal{L})}(a)$, and thus $SF(\mathcal{L})$ is coherent. We can analogously check the case when $SF(\mathcal{L})(a) = \texttt{out}$. □

Likewise $M$ and $OF$, $SF$ also satisfies Anonymity and Non-dictatorship.

**Proposition A.3.3.** $SF$ satisfies Anonymity and Non-Dictatorship.

*Proof.* The proof of Anonymity, and consequently Non-dictatorship, is straightforward using lemma A.1.1. □

In the following propositions, we show that $SF$ loses the satisfaction of some monotonicity and unanimity properties with respect to $M$ and $OF$. In fact, $SF$ only manages to preserve the Binary familiar monotonicity. Thus, this can be regarded as the price paid by $SF$ to ensure Collective coherence.

**Proposition A.3.4.** $SF$ satisfies Binary familiar monotonicity.

*Proof.* Let $L$ and $L'$ be two labelling profiles satisfying the hypothesis, and say that $SF(\mathcal{L}) = \texttt{in}$. Since $SF(\mathcal{L}) = \texttt{in}$, we must consider two cases: (i) $Pro_{SF(\mathcal{L})}(a) > Con_{SF(\mathcal{L})}(a)$ or (ii) $Pro_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L})}(a)$ and $\texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a)$. Since both

labelling profiles $\mathcal{L}$ and $\mathcal{L}'$ only differ on the direct opinion over $a$, then we can ensure that $Pro_{SF(\mathcal{L})}(a) = Pro_{SF(\mathcal{L}')}(a)$ and $Con_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L}')}(a)$, i.e, the aggregated indirect support of $a$ does not change.

Regarding case (i), there is no change and $SF(\mathcal{L}')(a) = \texttt{in}$. As to case (ii), we obtain that $\texttt{in}_{\mathcal{L}'}(a) \geq \texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a) \geq \texttt{out}_{\mathcal{L}'}(a)$. Therefore, $SF(\mathcal{L}')(a) = \texttt{in}$.

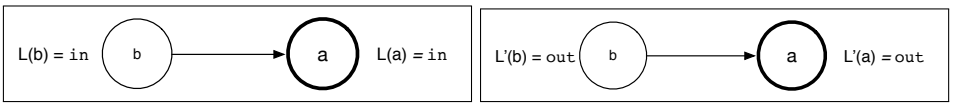
Since analogous reasoning follows when assuming $SF(\mathcal{L})(a) = \texttt{out}$, the Binary familiar monotonicity for $SF$ holds. □

**Proposition A.3.5.** $SF$ does not satisfy Binary Monotonicity nor Unanimity.

*Proof.* For both cases we will employ the same counterexample depicted in Figure A.5. Our counterexample considers a $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, a \rangle$ with labelling profiles $\mathcal{L} = (L)$, $L(a) = \texttt{in} = L(b)$ (Figure A.5(a)), and $\mathcal{L}' = (L')$, $L'(a) = \texttt{out} = L'(b)$ (Figure A.5(b)). The collective labellings obtained by $SF$ for each profile are $SF(\mathcal{L})(a) = \texttt{out}$, $SF(\mathcal{L})(b) = \texttt{in}$ and $SF(\mathcal{L}')(a) = \texttt{in}$, $SF(\mathcal{L}')(b) = \texttt{out}$ (as shown in figures A.6(a) and A.6(b)).

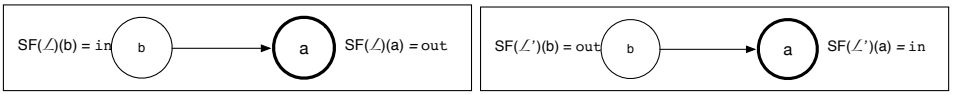
As to Binary monotonicity, our counterexample shows us that although the direct opinion of $a$ increases in favour of $SF(\mathcal{L})(a) = \texttt{out}$, the resulting labelling sets the collective opinion to $SF(\mathcal{L}')(a) = \texttt{in}$.

As to Unanimity, both labelling profiles agree on argument $a$, and hence there is unanimity, and yet the aggregated labels do not agree.




(a) (b)

Figure A.5: Counterexample of proposition A.3.5: Agents' labellings, $\mathcal{L}$ on the left and $\mathcal{L}'$ on the right.




(a) (b)

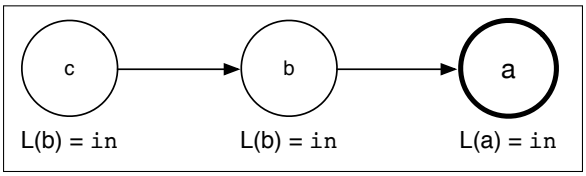
Figure A.6: Counterexample of proposition A.3.5: Aggregated labelling by $SF$ of labelling profiles $\mathcal{L}$ and $\mathcal{L}'$: $SF(\mathcal{L})$ on the left, $SF(\mathcal{L}')$ on the right.

□

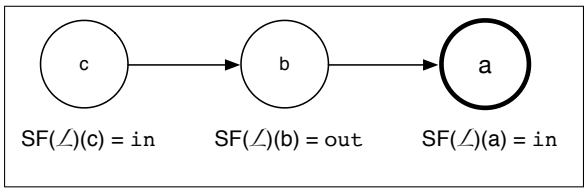
**Proposition A.3.6.** $SF$ does not satisfy Endorsed unanimity.

*Proof.* The proof requires an example of an argument with two levels of descendants, like the one we show in Figure A.7. Consider a $TODF = \langle\{a, b, c\}, \{(b, a), (c, b)\}, \emptyset, a\rangle$ with one labelling profile $\mathcal{L}$ formed by the following labelling: $L(a) = L(b) = L(c) =$ `in` (see Figure A.7(a)). Figure A.7(b) shows the collective labelling obtained by $SF$: $SF(\mathcal{L})(a) = \texttt{in} = SF(\mathcal{L})(c)$ and $SF(\mathcal{L})(b) = \texttt{out}$.

Our example shows that although $a$ has full negative support ($b$ is labelled `in`), the aggregated label for $a$ is accepted (labelled with `in`), contradicting the support.




(a)




(b)

Figure A.7: Counterexample for proposition A.3.6: (a) labelling profile; (b) aggregate labelling obtained by $SF$.

□

**Proposition A.3.7.** $SF$ does not satisfy Supportiveness, Familiar monotonicity, or Monotonicity.

*Proof.* The counterexample for proposition A.1.7 also applies to $SF$. □

Finally, as with $OF$, exploiting indirect opinions means that $SF$ loses the Independence property.

**Proposition A.3.8.** $SF$ does not satisfy Independence.

*Proof.* The counterexample for proposition A.2.8 also applies to $SF$. □

## A.4 Analysing the Balanced function

In what follows, we analyse the properties from Section 4.2.1 that $BF$ fulfils. Following the same scheme, first, we check the desirable properties.

**Proposition A.4.1.** $BF$ satisfies Exhaustive domain and Coherent domain.

*Proof.* It is clear that $BF$ is defined for all labelling profiles and hence, also defined for every coherent labelling. □

Notice that, since $BF$ is defined for all labelling profiles, it is also defined for labelling profiles in $Coh_c(TODF)^n$, namely for labelling profiles whose argument labellings are $c$-coherent. Another achievement of the function $BF$, stated in the following proposition, is the fulfilment of Collective coherence. $BF$ ensures that whatever the input profile, the outcome labelling will be rational.

**Proposition A.4.2.** $BF$ satisfies Collective coherence.

*Proof.* Let $a$ be an argument such that $BF(\mathcal{L})(a) = \texttt{in}$. From Definition 4.3.4 we know that $IO(\mathcal{L})(a) + DO(\mathcal{L})(a) > 0$. Thus, there are three possibilities: (i) $DO(\mathcal{L})(a) = 1$ and $IO(\mathcal{L})(a) = 1$; (ii) $DO(\mathcal{L})(a) = 1$ and $IO(\mathcal{L})(a) = 0$; or (iii) $IO(\mathcal{L})(a) = 0$ and $DO(\mathcal{L})(a) = 1$. Since $IO(\mathcal{L})(a) \geq 0$ in all cases, this implies that $Pro_{BF(\mathcal{L})}(a) \geq Con_{BF(\mathcal{L})}(a)$, and hence $BF$ satisfies the coherence property. The proof goes analogously for the case $BF(\mathcal{L})(a) = \texttt{out}$. □

Now we turn our attention to the Anonymity property and its weaker version: the Non-dictatorship property.

**Proposition A.4.3.** $BF$ satisfies Anonymity, and therefore, Non-Dictatorship.

*Proof.* The proof of Anonymity, and consequently Non-dictatorship, is straightforward using lemma A.1.1.

□

Next, we focus on the unanimity properties. First, we will show that $BF$ does not fulfil the Endorsed unanimity, but it does under some conditions.

**Proposition A.4.4.** $BF$ does not satisfy Endorsed unanimity.

*Proof.* The result can be easily seen with the following example. Let be a $TODF = \langle\{a, b\}, \{(b, a)\}, \emptyset, a\rangle$ with the labelling profile $\mathcal{L}$ formed by only one single labelling, $L(a) = \texttt{in}$ and $L(b) = \texttt{in}$. The collective labelling obtained applying $BF$ to $\mathcal{L}$ is the following: $BF(\mathcal{L})(a) = \texttt{undec}$ and $BF(\mathcal{L})(b) = \texttt{in}$. As we can see, $a$ has full negative support, but the collective label is `undec` instead of `out`.

□

**Proposition A.4.5.** $BF$ satisfies Endorsed unanimity over $Coh_0(TODF)^n$, the class of 0-coherent labelling profiles.

*Proof.* Consider a 0-coherent labelling profile $\mathcal{L} = (L_1, \ldots, L_n)$. Suppose an argument $a$ that has full positive support for all $L_i$, i.e., $Pro_{L_i}(a) = m$ for every $i$, where $m$ is the number of descendants of $a$. Using lemma A.1.2 we have that $L_i(a) = \texttt{in}$ for every $i$. Therefore, $DO(\mathcal{L})(a) = 1$, which in turn implies $BF(\mathcal{L})(a) \in \{\texttt{in}, \texttt{undec}\}$.

Consider $BF(\mathcal{L})(a) = \texttt{in}$. Let there be an $i$ such that $Pro_{L_i}(a) = m$ implies that for every descendant $b$ of $a$, $L_i(b) = \texttt{in}$ if $b \Vdash a$ and $L_i(b) = \texttt{out}$ if $b \mapsto a$. Therefore, for every descendant $b \in \{A(a) \cup D(a)\}$ its indirect support will be $IO(\mathcal{L})(b) = 1$ or $IO(\mathcal{L})(b) = -1$ depending on whether $b$ defends or attacks $a$ respectively. Moreover, this means that for $b$ defending $a$ $BF(\mathcal{L})(b) \in \{\texttt{in}, \texttt{undec}\}$ and for $b$ attacking $BF(\mathcal{L})(b) \in \{\texttt{out}, \texttt{undec}\}$, which in turn implies that $IO(\mathcal{L})(a) \geq 0$. And to finish, if $DO(\mathcal{L})(a) = 1$ and $IO(\mathcal{L})(a) \geq 0$ then $BF(\mathcal{L})(a) = \texttt{in}$, that is what we wanted.

The remaining case to check, supposing the full negative support for every agent, is analogous. □

Notice, however, that $BF$ does not satisfy Unanimity or the Supportiveness properties.

**Proposition A.4.6.** $BF$ does not satisfy Unanimity or Supportiveness.

*Proof.* Figure A.8 graphically represents a $TODF$ that will serve to illustrate our proposition. The $TODF$ contains a target argument $\tau = a$, which is defended by five other arguments $\{a_1, a_2, a_3, a_4, a_5\}$. The $TODF$ involves the argument labellings of three agents, $L_1$, $L_2$, and $L_3$ (see Figure A.8(a)):

(1) $L_1(a) = L_1(a_1) = L_1(a_2) = L_1(a_3) = \texttt{in}$ and $L_1(a_4) = L_1(a_5) = \texttt{out}$,

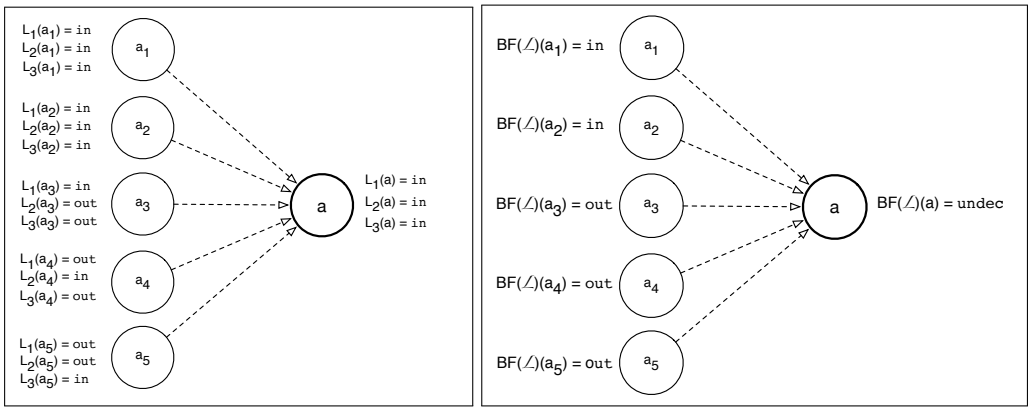
(2) $L_2(a) = L_2(a_1) = L_2(a_2) = L_2(a_4) = \texttt{in}$ and $L_2(a_3) = L_2(a_5) = \texttt{out}$, and

(3) $L_3(a) = L_3(a_1) = L_3(a_2) = L_3(a_5) = \texttt{in}$ and $L_3(a_3) = L_3(a_4) = \texttt{out}$.

Notice that the three agents agree on accepting the target, and hence there is a unanimous opinion on $a$. Figure A.8(b) depicts the resulting labelling when computing the $BF$ function the labelling profile $\mathcal{L}$. Since arguments $a_1$ and $a_2$ are collectively accepted ($BF(\mathcal{L})(a_1) = BF(\mathcal{L})(a_2) = \texttt{in}$) and arguments $a_3, a_4$, and $a_5$ are rejected ($BF(\mathcal{L})(a_3) = BF(\mathcal{L})(a_4) = BF(\mathcal{L})(a_5) = \texttt{out}$), the target is neither accepted nor rejected, $BF(\mathcal{L})(a) = \texttt{undec}$. Thus, $BF$ does not satisfy Unanimity.

As to Supportiveness, it does not hold either. Observe that although the aggregate label of $a$ is undec, no agent has labelled argument $a$ as undec.

□




(a) (b)

Figure A.8: Counterexample of proposition A.4.6; argument labellings (left) and result of the BF function (right).

Finally, we study $BF$'s monotonicity. Although Familiar monotonicity does not hold for $BF$, its weaker version, Binary familiar monotonicity, does hold.

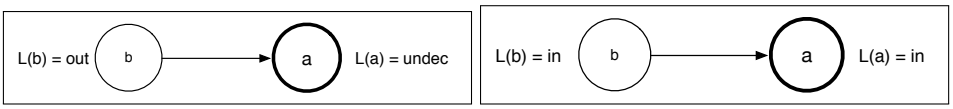
**Proposition A.4.7.** $BF$ does not satisfy Familiar monotonicity or Monotonicity.

*Proof.* The counterexample of proposition A.1.7 also applies for the function $BF$. □

**Proposition A.4.8.** $BF$ does not satisfy Binary Monotonicity.

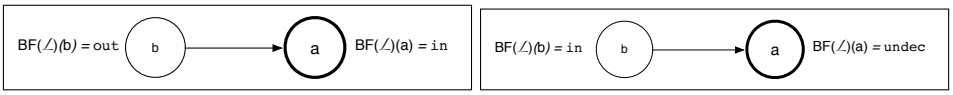
*Proof.* Let consider the following example with a $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, a \rangle$ with two labelling profiles $\mathcal{L} = (L)$ and $\mathcal{L}' = (L')$ with the following properties: $L(a) = \texttt{undec}$, $L(b) = \texttt{out}$ and $L'(a) = \texttt{in}$, $L'(b) = \texttt{in}$. The resulting labellings are $BF(\mathcal{L})(a) = \texttt{in}$, $BF(\mathcal{L})(b) = \texttt{out}$ and $BF(\mathcal{L})(a) = \texttt{out}$, $BF(\mathcal{L})(b) = \texttt{undec}$. As can be seen, $BF(\mathcal{L})(a) = \texttt{in}$ and increasing the labels in, on the second profile,

the collective labelling of $a$ changes to undec due to the changes in the descendant's opinion.




(a) (b)

Figure A.9: Counterexample of proposition A.4.8. On the left $\mathcal{L}$, on the right $\mathcal{L}'$.




(a) (b)

Figure A.10: Counterexample of proposition A.4.8. On the left $BF(\mathcal{L})$, on the right $BF(\mathcal{L}')$.

□

**Proposition A.4.9.** $BF$ satisfies Binary familiar monotonicity.

*Proof.* Let $\mathcal{L}, \mathcal{L}'$ be two labelling profiles satisfying the hypothesis required by the Binary familiar monotonicity property on the argument $a$, and whose collective label on $a$ for $\mathcal{L}$ is $BF(\mathcal{L})(a) = l = \texttt{in}$. Since $L_j(b) = L'_j(b)$ for all $b$ descendant of $a$, we know that $IO(\mathcal{L})(a) = IO(\mathcal{L}')(a)$ because $IO$ only depends on the descendants. Since $BF(\mathcal{L})(a) = \texttt{in}$, we have that $DO(\mathcal{L})(a) \geq 0$. Now, due to $\texttt{in}_{\mathcal{L}}(a) \leq \texttt{in}_{\mathcal{L}'}(a)$ and $\texttt{out}_{\mathcal{L}}(a) \geq \texttt{out}_{\mathcal{L}'}(a)$, we deduce that $DO(\mathcal{L}')(a) \geq DO(\mathcal{L})(a) \geq 0$. From this follows that $DO(\mathcal{L}')(a) + IO(\mathcal{L})(a) \geq DO(\mathcal{L})(a) + IO(\mathcal{L})(a) \geq 1$, and hence $BF(\mathcal{L}')(a) = \texttt{in}$. We can analogously check the case $BF(\mathcal{L})(a) = \texttt{out}$. □

**Proposition A.4.10.** $BF$ does not satisfy Independence.

*Proof.* The counterexample of proposition A.2.8 also applies to this function. □

# Appendix B

# Proofs for the Relational model

In the following, we prove all the formal results presented in Section 7.4 regarding the satisfaction of social choice properties by the opinion aggregation functions introduced in Section 7.3. The section is divided into four parts, one per debate scenario as analysed in Section 7.4.

1. Unconstrained opinion profiles;
2. Constrained opinion profiles: assuming consensus on acceptance degrees;
3. Constrained opinion profiles: assuming coherent profiles; and
4. Constrained opinion profile: assuming consensus on acceptance degrees and coherent profiles.

Furthermore, for each scenario, our results will be grouped by aggregation function in the following order: Direct aggregation, Indirect aggregation, Recursive aggregation, Balanced family aggregation and Recursive family aggregation.

## B.1 Unconstrained opinion profiles

In this section, we analyse the social choice properties fulfilled by the aggregation functions introduced in Section 7.3: assuming unconstrained opinions profiles (*any* opinion profile is deemed to be possible input for the aggregation functions). The results of this section are summarised in Table 7.1 in Section 7.4.1.

**Proposition B.1.1.** The aggregation function $D$ satisfies the following properties:

(i) Exhaustive domain and Coherent domain;

(ii) Anonymity and Non-Dictatorship;

(iii) Monotonicity and Familiar monotonicity;

(iv) Narrow unanimity, Sided unanimity and Weak unanimity; and

(v) Independence.

And does not satisfy:

(vi) Collective coherence; and

(vii) Endorsed unanimity.

*Proof.* (of proposition B.1.1)

(i) Exhaustive domain is straightforward and Collective Domain follows directly.

(ii) Anonymity and Non-Dictatorship. Let $P = (O_1, \ldots, O_n)$ be an opinion profile over a $DRF$ and $\sigma$ a permutation over a set of agents $Ag = \{1, \ldots, n\}$. We must show that $D$ maintains the same collective opinion over the permuted opinion profile $P' = (O_{\sigma(1)}, \ldots, O_{\sigma(n)})$, i.e. that $D(P) = D(P')$. This is the case because the next two equalities hold:

$$v_{D(P)}(s) = \frac{1}{n}\sum_{i=1}^{n} v_i(s) = \frac{1}{n}\sum_{i=1}^{n} v_{\sigma(i)}(s) = v_{D(P')}(a);$$

$$w_{D(P)}(r) = \frac{1}{n}\sum_{i=1}^{n} w_i(r) = \frac{1}{n}\sum_{i=1}^{n} w_{\sigma(i)}(r) = w_{D(P')}(r).$$

Therefore, Anonymity holds and Non-Dictatorship follows from it as we discussed in Section 7.2.2.

(iii) Monotonicity and Familiar monotonicity. Let $s$ be a statement and $P$ and $P'$ two opinion profiles satisfying the Monotonicity assumptions in the definition of the property in Section 7.2.2, i.e. $P = (O_1, \ldots, O_n)$ and $P = (O'_1, \ldots, O'_n)$ are such that $v_i(s) \leq v'_i(s)$ for every agent $i \in \{1, \ldots, n\}$. Then, from the definition of $D$, we obtain the aggregated valuation on $s$ is:

$$v_{D(P)}(s) = \frac{1}{n}\sum_{i=1}^{n} v_i(s) \leq \frac{1}{n}\sum_{i=1}^{n} v'_i(s) = v_{D(P')}(s)$$

Therefore, $D$ satisfies Monotonicity. Hence, from this and lemma 7.2.2, Familiar monotonicity also holds.

(iv) Narrow unanimity, Sided unanimity and Weak unanimity. Let $P = (O_1, \ldots, O_n)$ be an opinion profile over a $DRF$ and a statement $s \in \mathcal{S}$ such that $v_i(s) = \lambda$ for every agent in $Ag = \{1, \ldots, n\}$. The aggregated opinion on $s$ is:

$$v_{D(P)}(s) = \frac{1}{n}\sum_{i=1}^{n} v_i(s) = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda$$

Hence, Narrow unanimity is fulfilled by $D$. As discussed in Section 7.2.2, Weak unanimity follows from Narrow unanimity. Furthermore, according to proposition 7.2.2, Sided unanimity follows from Narrow unanimity and Monotonicity.

(v) Independence follows directly from the fact that $D$ satisfies Monotonicity and from proposition 7.2.3.

(vi) Collective Coherence. To prove that it does not hold, it suffices to find a $DRF$ and an opinion profile for which there is no collective coherence. Thus, consider the example depicted in Figure B.1.

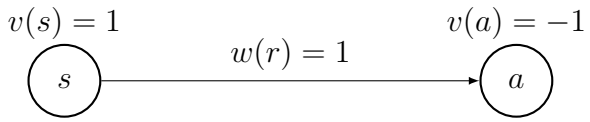



Figure B.1: Counterexample for Collective coherence in proposition B.1.1.

If we check coherence for statement $s$, we obtain that:

$$|v_{D(P)}(s) - e_{D(P)}(s)| = v(s) - v(a) = 2 > \epsilon.$$

for any $\epsilon \in (0, 1)$, and hence collective coherence does not hold for this profile.

(vii) Endorsed unanimity. Using the opinion profile depicted in Figure B.1, we observe that even with full negative support on $s$ (i.e. $v(a) = -1$), the result of the aggregation is the opposite ($v_{D(P)}(s) = 1$). Therefore, this opinion profile also serves as a counterexample to prove that $D$ does not satisfy Endorsed unanimity.

□

**Proposition B.1.2.** The aggregation function $I$ satisfies the following properties:

(i) Exhaustive domain and Coherent domain;

(ii) Anonymity and Non-Dictatorship;

(iii) Endorsed unanimity; and

(iv) Familiar monotonicity.

And does not satisfy:

(v) Collective coherence;

(vi) Narrow unanimity, Sided unanimity and Weak unanimity;

(vii) Monotonicity; and

(viii) Independence.

*Proof.* (of proposition B.1.2)

(i) Exhaustive and Coherent domain are straightforward.

(ii) Anonymity and Non-Dictadorship. Let $P = (O_1, \ldots, O_n)$ be an opinion profile over a $DRF$ and $\sigma$ a permutation over the agent in $Ag = \{1, \ldots, n\}$. We must show that $I$ maintains the same collective opinion over the permuted opinion profile $P' = (O_{\sigma(1)}, \ldots, O_{\sigma(n)})$, i.e. that $I(P) = I(P')$.

For any $i \in \{1, \ldots, n\}$ there is only one $j \in \{1, \ldots, n\}$ such that $\sigma(j) = i$, and hence in terms of expectation functions we know that $e_i = e_{\sigma(j)}$. Using that, we can show that $I(P) = I(P')$ as follows:

$$v_{I(P)}(s) = \frac{1}{n}\sum_{i=1}^{n} e_i(s) = \frac{1}{n}\sum_{i=1}^{n} e_{\sigma(i)}(s) = v_{I(P')}(a);$$

$$w_{I(P)}(r) = \frac{1}{n}\sum_{i=1}^{n} w_i(r) = \frac{1}{n}\sum_{i=1}^{n} w_{\sigma(i)}(r) = w_{I(P')}(r)$$

.

(iii) Endorsed unanimity. Let $s$ be a sentence and $P = (O_1, \ldots, O_n)$ an opinion profile satisfying that $v_i(s_r) = 1$ for any agent $i$ and descendant $s_r \in D(s)$ of sentence $s$. Since the expectation over $s$ is:

$$e_i(s) = \frac{1}{\sum_{r \in R^+(s)} w_i(r)} \sum_{r \in R^+(s)} w_i(r) v_i(s_r) = \frac{1}{\sum_{r \in R^+(s)} w_i(r)} \sum_{r \in R^+(s)} w_i(r) = 1,$$

then the aggregated value for $s$ is:

$$v_{I(P)}(s) = \frac{1}{n} \sum_i e_i(s) = 1.$$

Analogously, if we assume that $v_i(s_r) = -1$ for any agent $i$ and descendant $s_r \in D(s)$ of sentence $s$, we would obtain that $v_{I(P)}(s) = -1$. Since $v_{I(P)}(s) > 0$ when there is full positive support (and $v_{I(P)}(s) < 0$ for negative support), Endorsed unanimity holds.

(iv) Familiar monotonicity. It is straightforward to see that the output of the Indirect aggregation function, which uses an expectation function, depends only on the values of descendants and their relationships. So, it is clear that a different opinion profile maintaining the same values for descendants and their relationships will not change the output of the function.

(v) Collective Coherence. To prove that it does not hold, it suffices to find a $DRF$ and an opinion profile for which there is no collective coherence. Thus, consider the example depicted in Figure B.2. Here $v_{I(P)}(s) = 1$ and $v_{I(P)}(a) = -1 = v_{I(P)}(b)$. Now, if we check coherence for statement $s$, we obtain that $|v_{I(P)}(s) - e_{I(P)}(s)| = 2 > \epsilon$ for any $\epsilon \in (0, 1)$, and hence collective coherence does not hold for this profile.

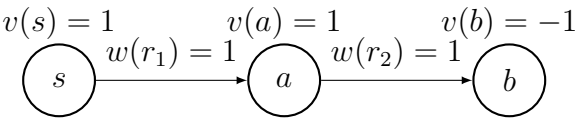



Figure B.2: Counterexample for Collective coherence in proposition B.1.2.

(vi) Narrow unanimity, Sided unanimity and Weak unanimity. Next, we build a $DRF$ and an opinion profile for which Weak unanimity does not hold despite satisfying the assumptions. Consider the example in Figure B.3 with opinion profile $P =$

$(O = (v, w))$. Although $v(s) = 1$, $v_{I(P)}(s) = -1$ instead of greater than 0, and hence $I$ does not satisfy Weak unanimity. As discussed in Section 7.2.2, an aggregation function satisfying either Narrow unanimity or Sided unanimity also satisfies Weak unanimity. Thus, since Weak unanimity does not hold, neither do Narrow unanimity and Sided unanimity.

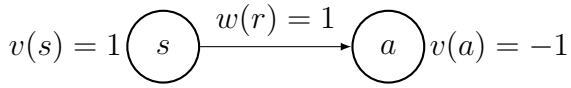



Figure B.3: Counterexample for Narrow, Sided and Weak unanimity in proposition B.1.2.

(vii) Monotonicity. Next we build a $DRF$ and an opinion profile for which Monotonicity does not hold despite satisfying the assumptions. Consider the opinion profile in figures B.4 and B.5 for the same $DRF$. The two profiles $P = (O = (v, w))$ and $P' = (O' = (v', w'))$ only differ on the valuation of $a$: $v(a) = 1$ and $v'(a) = -1$.

Clearly, $x = v(s) \leq v'(s) = x$, thus satisfying the assumptions of monotonicity. However, since the aggregated valuations on $s$ are: $v_{I(P)}(s) = 1$ and $v_{I(P')}(s) = -1$, it does not satisfy that $v_{I(P)} \leq v_{I(P')}$, and hence Monotonicity does not hold.

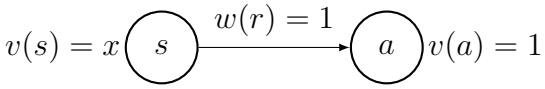



Figure B.4: Original profile in counterexample for Monotonicity in proposition B.1.2.

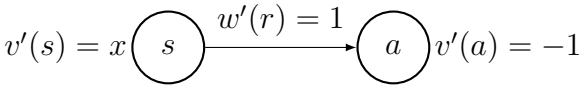



Figure B.5: Modified profile in counterexample for Monotonicity in proposition B.1.2.

(viii) Independence. Next we build a $DRF$ and an opinion profile for which Independence does not hold despite satisfying the assumptions. Consider the opinion profiles $P$ and $P'$ in figures B.6 and B.7. Although $v(s) = v'(s)$ for those profiles, the aggregated valuations on $s$ do not match: $1 = v_{I(P)}(s) \neq v_{I(P')}(s) = 0$. Therefore, Independence does not hold.

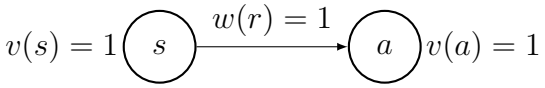



Figure B.6: Original profile in counterexample for Independence in proposition B.1.2.

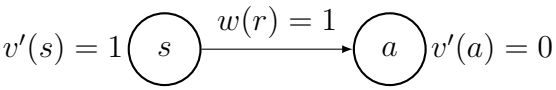



Figure B.7: Modified profile in counterexample for Independence in proposition B.1.2.

□

**Proposition B.1.3.** The aggregation function $R$ satisfies the following properties:

(i) Collective Coherence;

(ii) Exhaustive domain and Coherent domain;

(iii) Anonymity and Non-Dictatorship;

And does not satisfy:

(iv) Narrow unanimity, Sided unanimity and Weak unanimity;

(v) Endorsed unanimity;

(vi) Familiar monotonicity, so neither Monotonicity;

(vii) Independence.

*Proof.* (i) Collective Coherence. Since $v_{R(P)} = e_{R(P)}$, the collective opinion for $R$ is exactly the result of applying the estimation function, and hence collective coherence follows because $|v_{R(P)}(s) - e_{R(P)}(s)| = 0 < \epsilon$ for any $\epsilon \in (0, 1)$ and any sentence $s \in \mathcal{S}$.

(ii) Exhaustive domain and Coherent domain. Straightforward.

(iii) Anonymity and Non-Dictatorship. Let $P = (O_1, \ldots, O_n)$ be an opinion profile over a $DRF$ and $\sigma$ a permutation over the agents in $Ag = \{1, \ldots, n\}$. We must show that $R$ maintains the same collective opinion over the permuted opinion profile $P' = (O_{\sigma(1)}, \ldots, O_{\sigma(n)})$, i.e. that $R(P) = R(P')$.

We consider first the sentences $s \in \mathcal{S}$ with no descendants such that $R^+(s) = \emptyset$. Since these have no descendants, $R$ computes the collective opinion on them using $D$. As shown by proposition B.1.1, since $D$ satisfies Anonymity, it will also hold for $R$ when considering sentences with no descendants. Thus, since these sentences are used at the beginning of a recursive process by $R$, the collective opinion over any sentence will be the same after any permutation. Therefore, Anonymity holds for $R$, and from this Non-Dictatorship.

(iv) Weak, Narrow and Sided unanimity. The example of $DRF$ depicted in Figure B.8 with opinion profile $P = (O = (v, w))$ will be enough to show that $R$ does not satisfy Weak unanimity. Although $v(s) = 1$, and hence the assumptions for Weak unanimity hold, $v_{R(P)}(s) = -1$ influenced by the valuation on $b$. Since $v_{R(P)}(s)$ is not positive, Weak unanimity does not hold for $R$, and consequently neither Side unanimity nor Narrow unanimity.

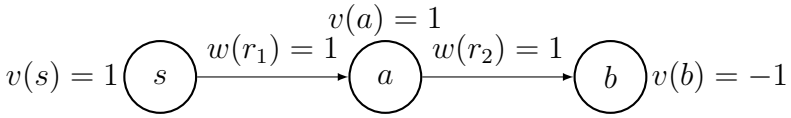



Figure B.8: Counterexample for Weak, Sided and Narrow unanimity in proposition B.1.3.

(v) Endorsed unanimity. Consider again the opinion profile depicted in Figure B.8. Clearly, since $v(a) = 1$, $s$ has full positive support, but $v_{R(P)}(s) = -1$. Since $v_{R(P)}(s)$ is not positive, Endorsed unanimity does not hold.

(vi) Familiar monotonicity and Monotonicity. We build a $DRF$ and two opinion profiles for which Familiar monotonicity does not hold despite satisfying the assumptions. Consider the two opinion profiles $P = (O = (v, w))$ and $P' = (O' = (v', w'))$ depicted in figures B.9 and B.10 respectively. Considering $s$, these two profiles satisfy the assumptions of Familiar monotonicity: $v(s) \leq v'(s)$ and the values on the indirect opinion are the same. However, $P$ and $P'$ differ on the value on $b$: $v(b) = 1$ and $v'(b) = -1$. This leads to a change of value on the aggregated value on $s$. Thus, $v_{R(P)}(s) \not\leq v_{R(P')}(s)$, and $R$ fails at satisfying Familiar monotonicity. By lemma 7.2.2, Monotonicity does not hold either.

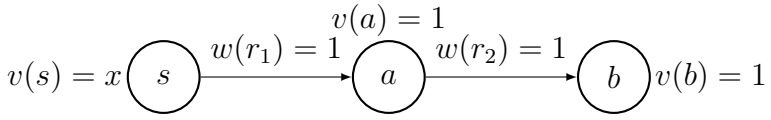



Figure B.9: Initial profile in counterexample for Familiar monotonicity and Monotonicity in proposition B.1.3.

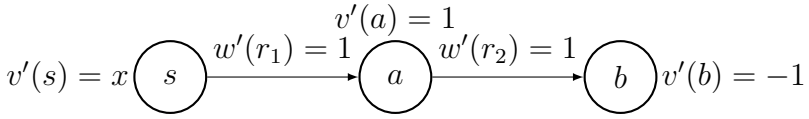



Figure B.10: Modified profile in counterexample for Familiar monotonicity and Monotonicity in proposition B.1.3.

(vii) Independence. Straightforward from the example employed in proposition B.1.2 to prove lack of Independence.

□

Next, we provide the proofs for the analysis of the families of $\alpha$-balanced aggregation functions $\{B_\alpha\}_{\alpha\in(0,1)}$ and $\alpha$-recursive aggregation functions $\{R_\alpha\}_{\alpha\in(0,1)}$. Before that, we first introduce some general lemmas that will be useful to build the proofs of the propositions for both families. To ease notation, these general lemmas that follow consider two generic aggregation functions $F$ and $G$, as well as a generic aggregation function $H = \alpha F + (1-\alpha)G$ instead of $v_H(P) = \alpha v_F(P) + (1-\alpha)v_G(P)$. Hereafter, the following lemmas establish the social properties fulfilled by $H$.

**Lemma B.1.1.** Let $F$ and $G$ be two opinion aggregation functions satisfying Exhaustive domain. For any $\alpha \in (0,1)$, aggregation function $H = \alpha F + (1-\alpha)G$ also satisfies Exhaustive domain.

*Proof.* Straightforward from the fact that both $F$ and $G$ satisfy Exhaustive domain. □

**Lemma B.1.2.** Let $F$ and $G$ two opinion aggregation functions satisfying Anonymity over domain $\mathcal{D}$ . For any $\alpha \in (0,1)$, aggregation function $H = \alpha F + (1-\alpha)G$ also satisfies Anonymity over domain $\mathcal{D}$.

*Proof.* For any given opinion profile $P$ and its permuted profile $P'$, if $F(P) = F(P')$ and $G(P) = G(P')$, then it follows that $H(P) = H(P')$. □

**Lemma B.1.3.** Let $F$ and $G$ two opinion aggregation functions satisfying Familiar monotonicity over domain $\mathcal{D}$. For any $\alpha \in (0,1)$, aggregation function $H = \alpha F + (1-\alpha)G$ also satisfies Familiar monotonicity on domain $\mathcal{D}$.

*Proof.* Let $P = (O_1 = (v_1, w_1), \ldots, O_n = (v_n, w_n))$ and $P' = (O'_1 = (v'_1, w'_1), \ldots, O'_n = (v'_n, w'_n))$ be a profile satisfying the assumptions of familiar montonicity for a statement $s$, i.e. $v_i(s) \leq v'_i(s)$ for any $i$ and for any $r \in R^+(s)$ then $w_i(r) = w'_i(r)$ and $v_i(s_r) = v'_i(s_r)$. Since $F$ and $G$ satisfy Familiar monotonicity, then $v_{F(P)}(s) \leq v_{F(P')}(s)$ and $v_{G(P)}(s) \leq v_{G(P')}(s)$. Thus, since $H = \alpha F + (1-\alpha)G$, it follows directly that $v_{H(P)}(s) \leq v_{H(P')}(s)$, so familiar montonicity holds for $H$. □

**Lemma B.1.4.** Let F and G two opinion aggregation functions satisfying Sided unanimity on domain $\mathcal{D}$. For any $\alpha \in (0,1)$, aggregation function $H = \alpha F + (1-\alpha)G$ also satisfies Sided unanimity on $\mathcal{D}$.

*Proof.* Since Sided unanimity holds for $F$ and $G$, we know that given any opinion profile $P$ of agents $\{1, \ldots, n\}$, i.e. if for any $i \in \{1, \ldots, n\}$ $v_i(s) > 0$ then $v_F(s) > 0$ and $v_G(s) > 0$, and since $v_H = \alpha v_F + (1-\alpha)v_G$, it also follows that $v_H(s) > 0$. Likewise for the negative case, so Sided unanimity holds for $H$. □

**Lemma B.1.5.** Let F and G two opinion aggregation functions satisfying Weak unanimity on the domain $\mathcal{D}$. For any $\alpha \in (0,1)$, aggregation function $H = \alpha F + (1-\alpha)G$ also satisfies Weak unanimity over domain $\mathcal{D}$.

*Proof.* Since Sided unanimity holds for $F$ and $G$, we know that given any opinion profile $P$ of agents $\{1, \ldots, n\}$, for any $i \in \{1, \ldots, n\}$, if $v_i(s) = 1$, then $v_F(s) > 0$ and $v_G(s) > 0$. Since $v_H = \alpha v_F + (1-\alpha)v_G$, it also follows that $v_H(s) > 0$, and hence Weak unanimity holds for $H$. Analogously for the negative case. □

**Lemma B.1.6.** Let F and G two opinion aggregation functions satisfying Endorsed unanimity on domain $\mathcal{D}$. For any $\alpha \in (0,1)$, aggregation function $H = \alpha F + (1-\alpha)G$ also satisfies Endorsed unanimity on $\mathcal{D}$.

*Proof.* Since Endorsed unanimity holds for $F$ and $G$, we know that given any opinion profile $P$ of agents $\{1, \ldots, n\}$, for any $i \in \{1, \ldots, n\}$ and descendant $s_r \in D(s)$ of sentence $s$, if $v_i(s_r) > 1$, then $v_F(s) > 0$ and $v_G(s) > 0$. Since $v_H = \alpha v_F + (1-\alpha)v_G$, it also follows that $v_H(s) > 0$, and hence Endorsed unanimity holds for $H$. Analogously for the full negative support case. □

We are now ready to prove the results for $\alpha$-balanced aggregation functions in $\{B_\alpha\}_{\alpha\in(0,1)}$.

**Proposition B.1.4.** The family of $\alpha$-balanced aggregation functions $\{B_\alpha\}_{\alpha\in(0,1)}$ satisfies the following properties:

(i) Exhaustive domain and Coherent domain;

(ii) Anonymity and Non-Dictatorship;

(iii) Weak unanimity for $\alpha > \frac{1}{2}$;

(iv) Endorsed unanimity for $\alpha < \frac{1}{2}$;

(v) Familiar monotonicity;

and does not satisfy:

(vi) Collective coherence;

(vii) Narrow unanimity, nor Sided unanimity;

(viii) Monotonicity;

(ix) Independence.

*Proof.* (i) Exhaustive domain and Coherent domain follow from propositions B.1.1 and B.1.2, and from lemma B.1.1.

(ii) Anonymity and Non-Dictatorship follow from propositions B.1.1 and B.1.2, and from lemma B.1.2.

(iii) Weak unanimity. Let $P = (O_1, \ldots, O_n)$ be an opinion profile over a $DRF$ for the agents in $Ag = \{1, \ldots, n\}$, and $s \in \mathcal{S}$ a sentence such that $v_i(s) = 1$ for any $i$. By proposition B.1.1, we know that Weak unanimity holds for the Direct aggregation function, and hence $v_{D(P)}(s) = \frac{1}{n}\sum_{i\in Ag} v_i(s) = 1$. Now we turn our attention to $I$, the indirect function. The worst scenario occurs when $v_{I(P)}(s) = -1$ because aggregating this value to $v_{D(P)}(s)$ might prevent that $v_{B_\alpha(P)}(s) > 0$, and thus that Weak unanimity holds. The DRF and an opinion profile depicted in Figure B.11 exemplify this case.

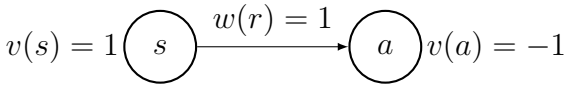



Figure B.11: Worst scenario for Weak and Endorsed unanimity in proposition B.1.4.

Since $v_{D(P)}(s) = 1$ and $v_{I(P)}(s) = -1$, $v_{B_\alpha(P)}(s) = \alpha - (1 - \alpha) = 2\alpha - 1$. Thus, if we set $\alpha$ so that $\alpha > \frac{1}{2}$, then we ensure that $v_{B_\alpha(P)}(s) > 0$, and Weak unanimity holds. The proof is analogous for the negative case of Weak unanimity.

(iv) Endorsed unanimity. Let $s$ be a sentence and $P = (O_1, \ldots, O_n)$ an opinion profile satisfying that $v_i(s_r) = -1$ for any agent $i$ and descendant $s_r \in D(s)$ of sentence $s$. In other words, $s$ has full negative support. It follows that $v_{I(P)}(s) = -1$. Likewise for our proof for Weak unanimity above, we consider the worst case, which would occur when $v_{D(P)}(s) = 1$. Figure B.11 depicts a DRF and single-opinion profile illustrating this case. Since $v_{D(P)}(s) = 1$ and $v_{I(P)}(s) = -1$, $v_{B_\alpha(P)}(s) = \alpha - (1 - \alpha) = 2\alpha - 1$. Thus, if we set $\alpha$ so that $\alpha < \frac{1}{2}$, then we ensure that $v_{B_\alpha(P)}(s) < 0$, and Endorsed unanimity holds. The proof is analogous for the positive case (full positive support) of Endorsed unanimity.

(v) Familiar monotonicity follows from propositions B.1.1 and B.1.2, and from lemma B.1.3.

(vi) Collective coherence.To prove that it does not hold, it suffices to find a $DRF$ and an opinion profile for which there is no collective coherence. Thus, consider the DRF with one-opinion profile depicted below in Figure B.12. Computing the aggregations for the Direct and Indirect functions, we have that $v_{D(P)}(s) = 1$, $v_{D(P)}(a) = 0$, and, $v_{I(P)}(s) = 0$ and $v_{I(P)}(a) = -1$. Therefore, $v_{B_\alpha(P)}(s) = \alpha$ and $v_{B_\alpha(P)}(a) = (-1)(1 - \alpha)$. And hence, the coherence at sentence $s$ is: $|v_{B_\alpha(P)}(s) - e_{B_\alpha(P)}(s)| = |v_{B_\alpha(P)}(s) - v_{B_\alpha(P)}(a)| = 1$. Thus, we conclude that, for any $\epsilon \in (0, 1)$, $\epsilon$-coherence cannot be satisfied regardless of the value of $\alpha$. Therefore, $B_\alpha$ does not satisfy $\epsilon$-coherence.

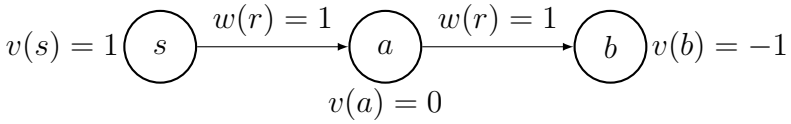



Figure B.12: Counterexample for Collective coherence in proposition B.1.4.

(vii) Sided unanimity, Narrow unanimity. To prove that neither of these properties holds, it suffices to find a $DRF$ and an opinion profile for which there is no Sided unanimity. In particular, we will show that for any $\alpha \in (0, 1)$, we can find a DRF and an opinion profile for which Sided unanimity and Narrow unanimity do not hold. Consider then the DRF with single-opinion profile in Figure B.13, where $x \in (0, 1)$ is such that $0 < x < \frac{1-\alpha}{\alpha}$. Since $v(s) = x > 0$, the assumptions for Sided unanimity hold at sentence $s$. Now, since $v_{D(P)}(s) = x$ and $v_{I(P)}(s) = -1$, it follows that $v_{B_\alpha(P)}(s) = \alpha x + (1-\alpha)(-1) = \alpha x + \alpha - 1 < \alpha\frac{1-\alpha}{\alpha} + \alpha - 1 = 0$. Since $v_{B_\alpha(P)} \not> 0$, Sided unanimity fails at $s$, and as it is single-opinion profile Narrow unanimity fails too. The proof goes analogously for the negative case of Sided unanimity.

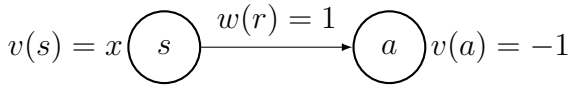



Figure B.13: Counterexample for Sided and Narrow unanimity in proposition B.1.4.

(viii) Monotonicity. It suffices to find a $DRF$ and an opinion profile for which there is no Monotonicity. Consider the two single-opinion profiles $P$ and $P'$ over the very same DRF in figures B.14 and B.15. We will check Monotonicity at sentence $s$, where the conditions for unanimity hold because $v(s) \leq v'(s)$. Computing $B_\alpha$ we obtain that $v_{B_\alpha(P)}(s) = 1$ and $v_{B_\alpha(P')}(s) = 2\alpha - 1$. To fulfil Monotonicity both expressions must satisfy that $v_{B_\alpha(P)}(s) \leq v_{B_\alpha(P')}(s)$, namely that $1 \leq 2\alpha - 1$. This is only possible for $\alpha \geq 1$. Therefore, for any $\alpha \in (0, 1)$ Monotonicity does not hold.

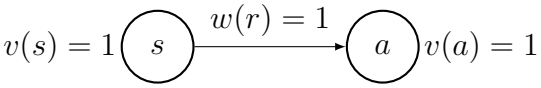



Figure B.14: Initial profile in counterexample for Monotonicity in proposition B.1.4.

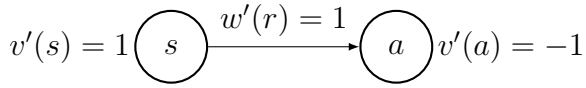



Figure B.15: Modified profile in counterexample for Monotonicity in proposition B.1.4.

$$1 \leq 2\alpha - 1$$

(ix) Independence. For any $\alpha \neq 1$, $B_\alpha$ does not fulfil Independence due to its dependence on $I$.

□

**Proposition B.1.5.** The family of $\alpha$-recursive aggregation functions $\{R_\alpha\}_{\alpha\in(0,1)}$ satisfies the following properties:

(i) Collective Coherence for $\alpha < \frac{\epsilon}{2}$;

(ii) Exhaustive domain and Coherent domain;

(iii) Anonymity and Non-Dictatorship;

(iv) Weak unanimity for $\alpha > \frac{1}{2}$;

and does not satisfy:

(v) Sided unanimity, so neither Narrow unanimity;

(vi) Endorsed unanimity;

(vii) Familiar monotonicity, so neither Monotonicity;

(viii) Independence.

*Proof.* (i) Collective Coherence. Given $\epsilon > 0$ and a DRF, we must prove that $|v_{R_\alpha(P)}(s) - e_{R_\alpha(P)}(s)| < \epsilon$ for any sentence $s \in \mathcal{S}$. First, we develop the difference between valuation and estimation for the collective opinion:

$$
\begin{aligned}
v_{R_\alpha(P)}(s) - e_{R_\alpha(P)}(s) &= v_{R_\alpha(P)}(s) - \frac{\sum_{r\in R^+(s)} w_{R_\alpha(P)}(r) v_{R_\alpha(P)}(s_r)}{\sum_{r\in R^+(s)} w_{R_\alpha(P)}(r)} \\
&= [\alpha v_{D(P)}(s) + (1-\alpha) v_{R(P)}(s)] \\
&\qquad - \frac{\sum_{r\in R^+(s)} w_{D(P)}(r)\big[\alpha v_{D(P)}(s_r) + (1-\alpha) v_{R(P)}(s_r)\big]}{\sum_{r\in R^+(s)} w_{R_\alpha(P)}(r)} \\
&= \alpha\big[v_{D(P)}(s) - \frac{\sum_{r\in R^+(s)} w_{R_\alpha(P)}(r) v_{D(P)}(s_r)}{\sum_{r\in R^+(s)} w_{D(P)}(r)}\big] \\
&\qquad + (1-\alpha)\big[v_{R(P)}(s) - \frac{\sum_{r\in R^+(s)} w_{R(P)}(r) v_{R(P)}(s_r)}{\sum_{r\in R^+(s)} w_{R(P)}(r)}\big] \\
&= \alpha\big[v_{D(P)}(s) - e_{D(P)}(s)\big] + (1-\alpha)\big[v_{R(P)}(s) - e_{R(P)}(s)\big] \\
&= \alpha(v_{D(P)}(s) - e_{D(P)}(s)).
\end{aligned}
$$

Notice that we get rid of $v_{R(P)}(s) - e_{R(P)}(s)$ because is zero. Now the coherence of $R_\alpha$ directly depends on the coherence of the direct aggregation function $D$ and $\alpha$. Figure B.16 depicts a DRF with a single-opinion profile representing a worst-case scenario for $D$ because $v_{D(P)}(s) - e_{D(P)}(s) = 2$. Considering the coherence of $R_\alpha$, we have that $|v_{R_\alpha(P)}(s) - e_{R_\alpha(P)}(s)| = \alpha |v_{D(P)}(s) - e_{D(P)}(s)| \leq 2\alpha$ for any profile $P$. Therefore, we must ensure that $\alpha < \frac{\epsilon}{2}$ so that $|v_{R_\alpha(P)}(s) - e_{R_\alpha(P)}(s)| < \epsilon$ holds for any profile of the domain, and hence Collective coherence holds for $R_\alpha$.

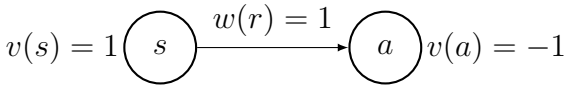



Figure B.16: Worst case scenario for Collective coherence in proposition B.1.5.

(ii) Exhaustive domain and Coherent domain directly follow from propositions B.1.1 and B.1.3 and lemma B.1.1.

(iii) Anonymity and Non-Dictatorship follow directly from propositions B.1.1 and B.1.3 and lemma B.1.2.

(iv) Weak unanimity. To prove Weak unanimity, we can resort to the proof built to prove Weak unanimity for $B_\alpha$ in proposition B.1.4. We simply have to substitute $B_\alpha$ for $R_\alpha$.

(v) Sided unanimity and Narrow unanimity. To prove that neither of these properties holds, it suffices to find a $DRF$ and an opinion profile for which there is no Sided unanimity (nor Narrow unanimity). Consider the DRF and the single-opinion profile depicted in Figure B.17, where $x \in (0,1)$ is such that $0 < x < \frac{1-\alpha}{\alpha}$. Since $x > 0$, the assumption for Sided unanimity holds at $s$. However, $v_{R_\alpha(P)}(s)$ is not positive, since $v_{R_\alpha(P)}(s) = x\alpha - 1 + \alpha < \frac{1-\alpha}{\alpha}\alpha - 1 + \alpha = 0$, and hence Sided (and Narrow) unanimity does not hold.

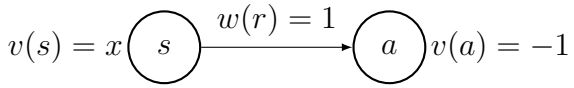



Figure B.17: Counterexample for Sided and Narrow unanimity in proposition B.1.5.

(vi) Endorsed unanimity. Next, we build a DRF and an opinion profile for which Endorsed unanimity does not hold. Consider the DRF and the opinion profile $P$ depicted in Figure B.18. The assumptions for Endorsed unanimity hold at $s$ because $s$ has full negative support. However, $v_{R_\alpha(P)}(s)$ is not negative: since $v_{D(P)}(s) = 1$ and $v_{R(P)}(s) = 1$, we obtain that $v_{R_\alpha(P)}(s) = 1$ for any $\alpha \in (0,1)$. Therefore, Endorsed unanimity does not hold.

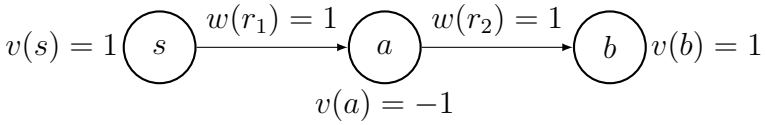



Figure B.18: Counterexample for Endorsed unanimity and original profile for counterexample in Familiar monotonicity and Monotonicity in proposition B.1.5.

(vii) Familiar monotonicity and Monotonicity. We build a $DRF$ and an opinion profile for which Familiar monotonicity does not hold despite satisfying the assumptions. Consider the DRF and single-opinion profile $P$ in Figure B.18 together with another single-opinion profile $P'$ in Figure B.19. Clearly, the assumptions of Familiar monotonicity are fulfilled at $s$ because $v_i(s) \leq v'_i(s)$ and the descendant of $s$ has the same value. However, we will show that $v_{R_\alpha(P)}(s) \leq v_{R_\alpha(P')}(s)$ is not true. For both profiles we have that $v_{D(P)}(s) = v_{D(P')}(s) = 1$, and, $v_{R(P)}(s) = 1$ and $v_{R(P')}(s) = 1 - x$. Thus, for any $\alpha \in (0,1)$: $v_{R_\alpha(P)}(s) = \alpha + (1-\alpha) = 1$

and $v_{R_\alpha(P')}(s) = \alpha + (1-\alpha)(1-x) = 1 - x(1-\alpha) < 1$ for any $x \in (0,1)$. Therefore, $v_{R_\alpha(P)}(s) > v_{R_\alpha(P')}(s)$ and Familiar monotonicity is not satisfied. By lemma 7.2.2, Monotonicity does not hold either.

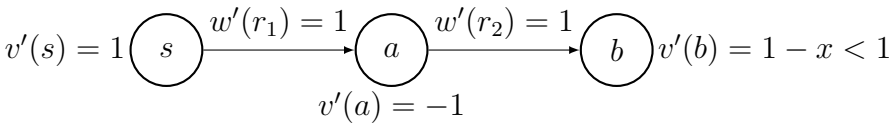



Figure B.19: Modified profile in counterexample for Familiar monotonicity and Monotonicity in proposition B.1.5.

(viii) Independence. Clearly, $R_\alpha$ will not fulfil Independence for any $\alpha \neq 1$ because of its dependence on $R$.

□

## B.2 Constrained opinion profiles: assuming consensus on acceptance degrees

This section relates to Section 7.4.2, where we assume that opinion profiles share consensus on their acceptance degrees on relationships, i.e. for each relationship $r \in R$ of a DRF all the agents agree on their acceptance degrees: $w_i(r) = w_j(r)\ \forall i, j \in Ag$.

In the previous section, each proof and counterexample used to demonstrate that an aggregation function does or does not satisfy a property uses opinion profiles composed by one single agent. Thus, those proofs also serve in this section when assuming consensus on acceptance degrees. For this reason, adding this assumption does not change any of the properties fulfilled by the aggregation functions in the general case (Table 7.1). Therefore there are no further desirable properties gained in this scenario with respect to the more general scenario thoroughly analysed in Section B.1.

## B.3 Constrained opinion profiles: assuming coherent profiles

This section corresponds to the results displayed in Table 7.2 in Section 7.4.3. We prove the results regarding the social choice properties satisfied by the aggregation functions introduced in Section 7.3 when assuming the domain of the aggregation functions to be $\epsilon$-coherent for some $\epsilon \in (0, 1)$. This means that we consider that our aggregation functions take in coherent opinion profiles.

Since, in the previous section, many properties have been proven for the general case, we will not need to prove them again for this more restrictive scenario. For each opinion aggregation function, we will prove only those results regarding social choice properties that change by adding the coherence assumption and disprove again, this time for coherent domains, those properties which are yet not satisfied.

**Proposition B.3.1.** For any $\epsilon \in (0, 1)$, $D$ over an $\epsilon$-coherent domain satisfies Endorsed unanimity.

*Proof.* Let $s$ a statement in a DRF and $R^+(s) \neq \emptyset$ the set of relationships $r$ from $s$ to its descendants $s_r$. Let $P$ be an $\epsilon$-coherent profile for $\epsilon \in (0, 1)$ with full positive support on $s$, i.e. $v_i(s_r) = 1$ for any $i$ and descendant $s_r \in D(s)$. Then:

$$e_i(s) = \frac{1}{\sum_{r \in R^+(s)} w_i(r)} \sum_{r \in R^+(s)} v_i(s_r) w_i(r) = \frac{1}{\sum_{r \in R^+(s)} w_i(r)} \sum_{r \in R^+(s)} w_i(r) = 1$$

By the $\epsilon$-coherence of $P$ we have that:

$$|v_i(s) - e_i(s)| < \epsilon \Longrightarrow v_i(s) > e_i(s) - \epsilon = 1 - \epsilon.$$

Therefore, for any $\epsilon \in (0, 1)$ we can ensure that $v_i(s) > 0$ for any $i$ and the conditions for Sided unanimity hold. Now, since $D$ satisfies Sided unanimity (by proposition B.1.1), we obtain that $v_D(s) > 0$, and hence $D$ fulfils Endorsed unanimity. □

**Proposition B.3.2.** $D$ over a $\delta$-coherent domain, where $\delta \in (0, 1)$, still does not satisfy $\epsilon$-Collective coherence for any $\epsilon \in (0, 1)$.

*Proof.* Consider the DRF and $\delta$-coherent opinion profile $P$ depicted in Figure B.20 and any $\delta \in (0, 1)$. We will show that the collective opinion yield by the direct function for this example is never $\epsilon$-coherent for any $\epsilon \in (0, 1)$.

Clearly, this profile is $\delta$-coherent for any $\delta > 0$. Computing the direct function at $s$ we obtain that: $v_{D(P)}(s) = -1$, $v_{D(P)}(a) = 0$ and $w_{D(P)}(r) = \frac{1}{2}$. Now, if we check collective coherence at $s$, we see that: $|v_{D(P)}(s) - e_{D(P)}(s)| = |-1-0| = 1 > \epsilon$. Thus, since 1 is larger than any $\epsilon$ value that we take in $(0, 1)$, $D$ does not satisfy $\epsilon$-Collective coherence.

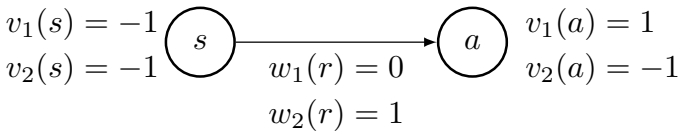



Figure B.20: Counterexample for Collective coherence in proposition B.3.2.

□

**Proposition B.3.3.** For $\epsilon \in (0, 1)$, $I$ over an $\epsilon$-coherent domain satisfies Weak unanimity.

*Proof.* Consider a DRF with a statement $s \in \mathcal{S}$ and $P = (O_1 = (v_1, w_1), \ldots, O_n = (v_n, w_n))$ an opinion profile such that $v_i(s) = 1$ for every $i$. Hence, the conditions for Weak unanimity hold. If the profile $P$ is $\epsilon$-coherent, where $\epsilon \in (0, 1)$, then we can conclude that for any $i$: $1 - \epsilon < e_i(s) < 1 + \epsilon$, being $1 - \epsilon > 0$ for any $\epsilon \in (0, 1)$. Now, computing $v_I$ at $s$ we get:

$$v_{I(P)}(s) = \frac{1}{n} \sum_i e_i(s) > \frac{1}{n} \sum_i 1 - \epsilon > 0$$

Since $v_{I(P)}(s) > 0$, Weak unanimity holds. The proof for the negative case of Weak unanimity is analogous.

□

**Proposition B.3.4.** For any $\delta \in (0, 1)$, $I$ over an $\delta$-coherent domain still does not satisfy $\epsilon$-Collective coherence for any $\epsilon \in (0, 1)$.

*Proof.* To prove that this property does not hold, it suffices to find a $DRF$ and an opinion profile for which there is no $\epsilon$-Collective coherence. Consider the DRF and opinion profile $P$ in Figure B.21. Clearly, opinions $O_1$ and $O_2$ of $P$ are $\delta$-coherent for any $\delta > 0$. Now, we compute the indirect function for all the statement: $v_{I(P)}(s) = \frac{-1}{2}$,

$v_{I(P)}(a) = \frac{1}{2}, v_{I(P)}(b) = 0$, and, $w_{I(P)}(r_1) = \frac{1}{2} = w_{I(P)}(r_2)$. If we check coherence at $s$ we see that:

$$|v_{I(P)}(s) - e_{I(P)}(s)| = |v_{I(P)}(s) - v_{I(P)}(a)| = |\frac{-1}{2} - \frac{1}{2}| = 1 > \epsilon.$$

Thus, since $1$ is bigger that any $\epsilon$ value that we take in $(0, 1)$, $I$ does not satisfy $\epsilon$-Collective coherence.

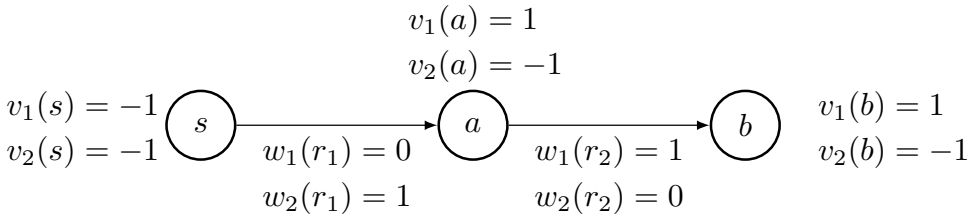



Figure B.21: Counterexample for Collective coherence in proposition B.3.4.

□

**Proposition B.3.5.** For any $\epsilon \in (0, 1)$, $I$ over an $\epsilon$-coherent domain still does not satisfy:

(i) Sided unanimity, and therefore Narrow unanimity;

(ii) Monotonicity.

*Proof.* (i) Sided unanimity and Narrow unanimity. To prove that these properties do not hold, it suffices to find a $DRF$ and an opinion profile for which there is no Sided unanimity. Consider the DRF and one-opinion profile depicted in Figure B.22 such that $\epsilon \in (0, 1)$ and $x, y$ such that $0 < x < y < \epsilon$. The assumptions of Sided unanimity are fulfilled at $s$. We check that the opinion in the profile is $\epsilon$-coherent because $|v(s) - e(s)| = |x - y + \epsilon| < \epsilon$. However, $v_{I(P)}(s) = y - \epsilon < 0$, instead of positive, and hence Sided unanimity is not satisfied. As in previous proofs, as the counterexample is a single opinion, Narrow unanimity does not hold either.

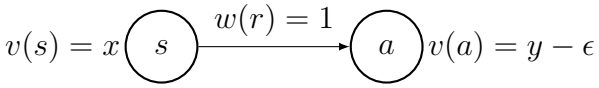



Figure B.22: Counterexample for Sided and Narrow unanimity in proposition B.3.5.

(ii) Monotonicity. To prove that it does not hold, it suffices to find a $DRF$ with opinion profiles satisfying the Monotonicity assumptions are satisfied, and yet Monotonicity is not. In fact, we will create a generic counterexample for any $\epsilon$-coherent domain. Consider the $DRF = \langle \mathcal{S}, R, \tau \rangle$ depicted in Figure B.23 with $\mathcal{S} = \{s, a\}$ and $R = \{r\}$. Also in the figure, let $P = ((v, w)))$ be an opinion profile of one single agent such that $v(s) = x$, $w(r) = 1$, and $v(a) = y >$, where $-1 < y < x < 1$ and $0 < x - y < \epsilon$.

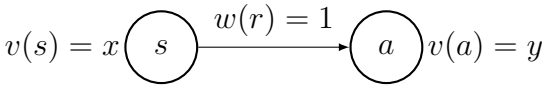



Figure B.23: Initial profile in counterexample for Monotonicity in proposition B.3.5.

Since $v(s) - e_O(s) = v(s) - v(a)$, then the profile $P$ is clearly $\epsilon$-coherent, and hence $P \in \mathbb{C}_\epsilon(DRF)$. We compute the collective opinion using $I$ at $s$ as: $v_{I(P)}(s) = e_O(s) = v(a) = y$.

Now, consider another profile $P' = (O = (v', w))$ over the same DRF, shown in Figure B.24, such that $v'(s) = x + \frac{\omega}{3}$ and $v'(s) = y - \frac{\omega}{3}$, where $\omega > 0$, such that $x - y + \omega \leq \epsilon$, $x + \frac{\omega}{3} \leq 1$ and $y - \frac{\omega}{3} \geq -1$.Clearly $v'(s) > v(s)$ and $P$ is also $\epsilon$-coherent, i.e.:

$$v(s) - e_{O'}(s) = (x + \frac{\omega}{3}) - (y - \frac{\omega}{3}) < \epsilon.$$

Nonetheless, $v_{I(P')}(s) = e_{O'}(s) = y - \frac{\omega}{3} < y$, which means that $v_{I(P)}(s) \not\leq v_{I(P')}(s)$, and hence this example cannot satisfy Monotonicity for any $\epsilon$.

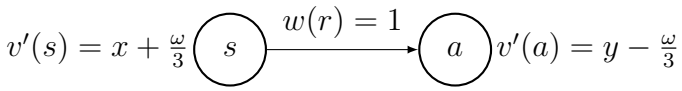



Figure B.24: Modified profile in counterexample for Monotonicity in proposition B.3.5.

□

**Proposition B.3.6.** For any $\epsilon \in (0, 1)$ and considering the domain to be $\epsilon$-coherent, $R$ does not fulfil the following properties:

(i) Weak unanimity, neither Sided nor Narrow unanimity;

(ii) Endorsed unanimity;

(iii) Familiar monotonicity, and therefore Monotonicity.

*Proof.* (i) Weak unanimity, Sided unanimity, Narrow unanimity. To prove that neither of these properties holds, it suffices to build a DRF and opinion profile to show that Weak unanimity does not hold. This is sufficient because Weak unanimity is a weaker case than Sided and Narrow unanimity. Proposition 7.2.1 tells us that Sided and Narrow unanimity will not hold if Weak unanimity does not. Consider the DRF and opinion profile $P = ((v, w))$ in picture B.25 such that $w(r) = 1$ for any relationship $r \in R$, $\epsilon \in (0, 1)$ and $\delta \in (0, \epsilon)$ and $m \in \mathbb{N}$ so that $m\delta \geq 1 > (m-1)\delta$.

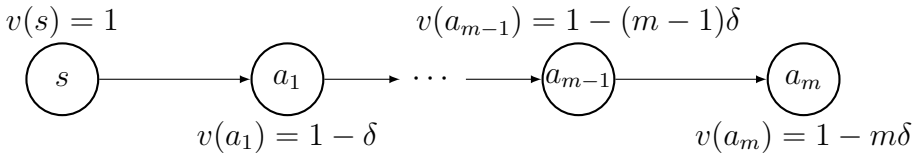



Figure B.25: Counterexample for Weak, Sided and Narrow unanimity in proposition B.3.6.

Clearly, the outcome of the recursive function at each sentence is obtained from the value of the recursive function at the previous sentence, i.e.:

$$v_{R(P)}(a_m) = v_{R(P)}(a_{m-1}) = ... = v_{R(P)}(a_1) = v_{R(P)}(s),$$

which actually is the value $v(a_m) = 1 - m\delta \leq 0$. So, this is an opinion profile $\epsilon$-coherent fulfilling the assumptions of Weak unanimity at sentence $s$ because $v(s) = 1$. However, the value of the recursive function at $s$ is negative. Therefore, $R$ does not fulfil Weak unanimity.

(ii) Endorsed unanimity. We build a DRF and opinion profile to show that Endorsed unanimity does not hold from the example in the previous proof. Figure B.26 shows our example, which extends the one in Figure B.25 with an additional sentence $a$. Since $v(a_i) - v(a_{i-1}) = \delta$, likewise in the proof above, we have an $\epsilon$-coherent opinion profile. Since $v(s) = 1$ the assumption for Endorsed unanimity at $a$ is satisfied, but since $v(a) = 1 - m\delta \leq 0$, Endorsed unanimity does not hold.

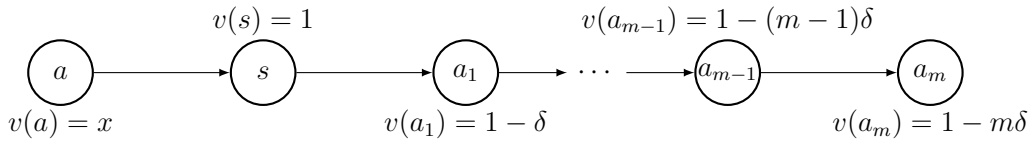



Figure B.26: Counterexample for Endorsed unanimity in proposition B.3.6.

(iii) Familiar monotonicity and Monotonicity. Consider the opinion profiles $P$ and $P'$ over the same DRF depicted in figures B.27 and B.28 respectively. Since $v(s) = v(a) = v(b) = 1$, $P$ is $\epsilon$-coherent. By setting $0 < x < \epsilon$, we also obtain that $P'$ is $\epsilon$-coherent. Therefore, both $P$ and $P'$ are $\epsilon$-coherent and the assumptions for Familiar monotonicity hold at $s$. However, since $1 = v_{R(P)}(s) > v_{R(P')}(s) = 1 - x$, Familiar monotonicity cannot hold. By lemma 7.2.2 Monotonicity does not hold either.

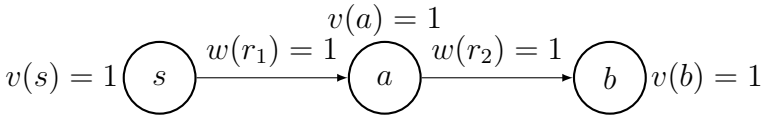



Figure B.27: Initial profile in counterexample for Familiar monotonicity and Monotonicity in proposition B.3.6.

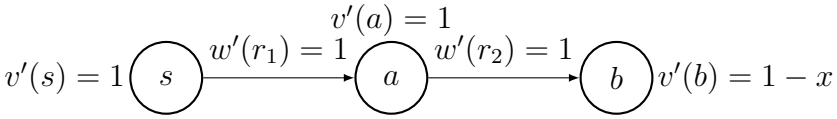



Figure B.28: Modified profile in counterexample for Familiar monotonicity and Monotonicity in proposition B.3.6.

□

**Proposition B.3.7.** For any $\epsilon \in (0, 1)$ and considering the domain to be $\epsilon$-coherent, then the family $\{B_\alpha\}_{\alpha\in(0,1)}$ satisfies:

(i) Weak unanimity; and

(ii) Endorsed unanimity.

*Proof.* (i) Weak unanimity follows from propositions B.1.1, B.3.3 and B.1.5.

(ii) Endorsed unanimity follows from propositions B.1.2, B.3.1 and B.1.6.

□

**Proposition B.3.8.** For any $\delta \in (0,1)$, $B_\alpha$ over a $\delta$-coherent domain still does not satisfy $\epsilon$-Collective coherence for any $\epsilon \in (0,1)$.

*Proof.* First, we show that the $\epsilon$-coherence condition for $B_\alpha$ depends on the functions employed in its definition, namely on $D$ and $I$:

$$|v_{B_\alpha(P)}(s) - e_{B_\alpha(P)}(s)| = \Big| v_{B_\alpha(P)}(s) - \frac{\sum_{r \in R^+(s)} \big(\alpha v_{D(P)}({}_r s) + (1-\alpha) v_{I(P)}(s_r)\big) w_{D(P)}(r)}{\sum_{r \in (R^+(s)} w_{D(P)}(r)} \Big|$$

$$= \Big| \big(\alpha(v_{D(P)}(s) + (1-\alpha)(v_{I(P)}(s)\big) - \big(\alpha e_{D(P)}(s)) + (1-\alpha) e_{I(P)}(s))\big) \Big|$$
$$= \Big| \alpha\big(v_{D(P)}(s) - e_{D(P)}(s)\big) + (1-\alpha)\big(v_{I(P)}(s) - e_{I(P)}(s)\big) \Big|$$

Thus, since $D$ and $I$ do not satisfy $\epsilon$-collective coherence for any $\delta$-coherent profile (by propositions B.3.2 and B.3.4 respectively), neither will $B_\alpha$ satisfy the property for any $\alpha \in (0,1)$. Indeed, consider for instance the DRF and $\delta$-coherent opinion profile $P$, with any $\delta \in (0,1)$, in Figure B.21 as employed in proposition B.3.4. If we compute $\epsilon$-collective coherence for $B_\alpha$ at sentence $s$ we obtain that:

$$|v_{B_\alpha(P)}(s) - e_{B_\alpha(P)}(s)| = \Big| \alpha(-1-0) + (1-\alpha)(-\frac{1}{2} - \frac{1}{2}) \Big| = |-1| = 1 > \epsilon$$

for any $\epsilon \in (0,1)$. So, $B_\alpha$ does not fulfill $\epsilon$-coherence for any $\alpha \in (0,1)$.

□

**Proposition B.3.9.** For any $\epsilon \in (0,1)$ and considering the domain to be $\epsilon$-coherent, $B_\alpha$ does not fulfil the following properties for any $\alpha \in (0,1)$:

(i) Sided unanimity and Narrow unanimity;

(ii) Monotonicity; and

(iii) Independence.

*Proof.* (i) Sided unanimity. It suffices to build a DRF and an opinion profile for which Sided unanimity does not hold for any values of $\alpha$ and $\epsilon$, where $\alpha, \epsilon \in (0, 1)$. The next counterexample serves to see that Narrow unanimity fails too.

Consider the set $A = \{(x, y) \in (0, 1) \quad | \quad 0 < y < \epsilon \text{ and } 0 < x < y - \alpha y\}$. We check first, that this set is actually not empty. For $\alpha \in (0, 1)$, $y - \alpha y > 0$, thus $y > \alpha y > 0$. So for $y \in (0, \epsilon)$, there are $x \in (0, 1)$ satisfying $x < y - \alpha y$.

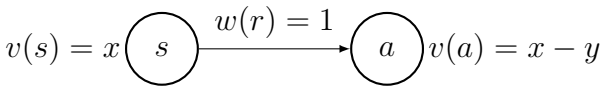



Figure B.29: Counterexample for Sided and Narrow unanimity in proposition B.3.9.

Now, we consider the DRF and opinion profile depicted in Figure B.29 where $x$ and $y$ are values from $A$, namely $(x, y) \in A$. Since $|v(s) - e(s)| = |x - (x - y)| = |y| < \epsilon$, the opinion profile in the figure is $\epsilon$-coherent, and satisfies the assumptions for Sided unanimity at $s$ because $v(s) = x > 0$. However,

$$\begin{aligned} v_{B_\alpha(P)}(s) &= \alpha v_{D(P)}(s) + (1 - \alpha) v_{I(P)}(s) \\ &= \alpha x + (1 - \alpha)(x - y) \\ &= x - y + \alpha y < 0 \end{aligned}$$

since $(x, y) \in A$. So, clearly this example shows that Sided unanimity does not hold for the family $B_\alpha$ in an $\epsilon$-coherent profile.

(iii) Independence. For any $\alpha \in (0, 1)$, $B_\alpha$ does not fulfil Independence due to its dependence on $I$.

(ii) Monotonicity. Straightforward from the fact that $B_\alpha$ does not fulfil Independence for any $\alpha \in (0, 1)$ and from proposition 7.2.3.

□

**Proposition B.3.10.** For any $\delta \in (0, 1)$, $R_\alpha$ over a $\delta$-coherent domain satisfies $\epsilon$-Collective coherence for $\alpha \leq \frac{\epsilon}{2}$.

*Proof.* As seen before in proposition B.1.3, the collective coherence of $R_\alpha$ entirely depends on the collective coherence of $D$, i.e.:

$$|v_{R_\alpha(P)}(s) - e_{R_\alpha(P)}(s)| = \alpha|v_{D(P)}(s) - e_{D(P)}(s)|$$

Thus, finding the worst-case scenario for $D$ will give us the condition on $\alpha$ that ensures that $R_\alpha$ satisfies $\epsilon$-collective coherence for any $\epsilon$. Next, we consider an example showing that $|v_{D(P)}(s) - e_{D(P)}(s)|$ can be as close to 2 as wanted, depending on the number of agents.

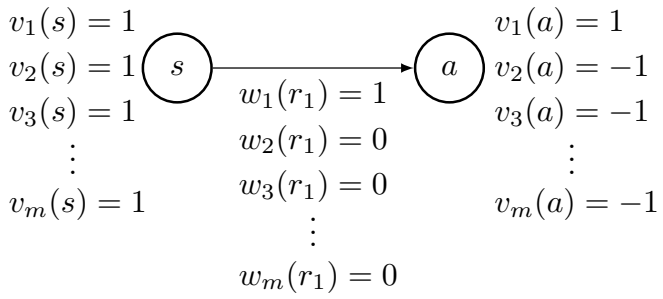



Figure B.30: Worst case scenario for Collective coherence in proposition B.3.10.

Let $P$ be the $\delta$-coherent opinion profile over a DRF depicted in Figure B.30, for any $\delta \in (0, 1)$. For any $i > 1$: $v_i(s) = 1$, $v_i(a) = -1$ and $w_i(r_1) = 0$; whereas $v_1(s) = 1$, $v_1(a) = 1$ and $w_1(r_1) = 1$. We check the condition for collective coherence at $s$ to find that:

$$|v_{R_\alpha(P)}(s) - e_{R_\alpha(P)}(s)| = \alpha(1 + \frac{m-2}{m}) < \epsilon$$

if $\alpha < \frac{\epsilon}{1+\frac{m-2}{m}}$. Thus, by taking $\alpha < \frac{\epsilon}{2} < \frac{\epsilon}{1+\frac{m-2}{m}}$ we ensure that $R_\alpha$ satisfies $\epsilon$-coherence for the worst case. Therefore, for any $\delta$-coherent opinion profile, $\delta \in (0, 1)$, choosing $\alpha < \frac{\epsilon}{2}$ will ensure that $R_\alpha$ satisfies $\epsilon$-collective coherence for any $\epsilon \in (0, 1)$.

□

**Proposition B.3.11.** Let $\epsilon \in (0, 1)$ such that the domain of $R_\alpha$ is an $\epsilon$-coherent domain, then the family $\{R_\alpha\}_{\alpha\in(0,1)}$ satisfies:

(i) Weak unanimity for $\alpha > \frac{1}{2}$; and

(ii) Endorsed unanimity for $\alpha > \frac{1}{2-\epsilon}$.

*Proof.* (i) Weak unanimity. Consider a DRF with sentences $\mathcal{S}$, $P$ an opinion profile over the DRF and $s \in \mathcal{S}$ a sentence such that $v_i(s) = 1$ for any agent $i$. We know

that

$$v_{D(P)}(s) = \frac{1}{n}\sum_{i \in Ag} v_i(s) = 1.$$

Now we turn our attention to $R$, the Recursive function. We consider the worst scenario for $R_\alpha$, which happens when $v(s) = 1$ and $v_{R(P)}(s) = -1$. The DRF and profile depicted in Figure B.8 above shows that, in fact, this scenario exists with $v_{D(P)}(s) = 1$ and $v_{R(P)}(s) = -1$, and hence $v_{R_\alpha(P)}(s) = \alpha + (1-\alpha)(-1) = 2\alpha - 1$. To fulfil Weak unanimity, we need that $v_{R_\alpha(P)}(s) > 0$ holds, but we also know that $v_{R_\alpha(P)}(s) \geq 2\alpha - 1$. Therefore, we can guarantee Weak unanimity by choosing $\alpha > \frac{1}{2}$. The proof for the negative case of Weak unanimity goes analogously.

(ii) Endorsed unanimity. To prove this property, we will build a customised DRF and opinion profile to show the worst case that we can find when fulfilling the assumptions of Endorsed unanimity.

Consider a DRF and let $P = (O_1 = (v_1, w_1), \ldots, O_n = (v_n, w_n))$ be an $\epsilon$-coherent profile with full positive support on statement $s \in \mathcal{S}$.

First, we consider the worst case where $v_{R(P)}(s) = -1$ can be achieved when $s$ has full positive support. Figure B.31 depicts a $DRF$ and an opinion profile illustrating this situation.

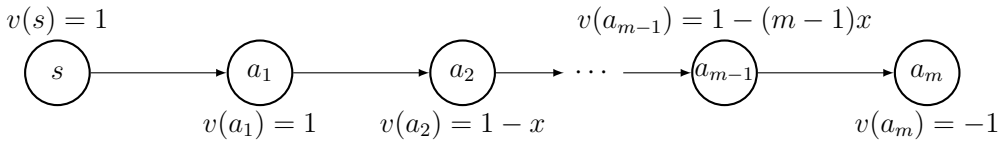



Figure B.31: Worst case scenario for Endorsed unanimity in proposition B.3.11.

By choosing $0 < x < \epsilon$ and $m \in \mathbb{N}$ such that $mx > 2 \geq (m-1)x$, this example shows an $\epsilon$-coherent profile where $v(a_1) = 1$ (full positive support) and $v_R(s) = -1$. Next, we move to the general setting considered by the proof, an opinion profile with $n$ agents, knowing that the worst case for this property is possible. Since $v_i(s_r) = 1$ for any descendant $s_r \in D(s)$ and any agent $i$, the estimation function on $s$ will be $e_i(s) = 1$ for any agent. Therefore, from the

coherence condition at $s$ we conclude that

$$1 - \epsilon < v_i(s) < 1 + \epsilon.$$

Consider that for every $i$, $v_i(s) = 1 - \delta_i$ such that $0 \leq \delta_i < \epsilon$. This clearly satisfies the previous inequality. Now we take $\delta = \max_i\{\delta_1, \ldots, \delta_n\}$ to create a new opinion profile $P' = (O'_1 = (v'_1, w_1), \ldots, O'_n = (v'_n, w_n))$ such that $v_i(a) = v'_i(a)$ for $a \in \mathcal{S} \setminus \{s\}$ and $v'_i(s) = 1 - \delta$ for any $i$. Then, since $D$ fulfils Monotonicity and Narrow unanimity, we know that $v_{D(P')}(s) \leq v_{D(P)}(s)$ and $v_{D(P')}(s) = 1 - \delta$ respectively. And, from the example in Figure B.31 we know that for any $\epsilon$-coherent opinion profile $v_{R(P)}(s) \geq -1$. Therefore,

$$\begin{aligned} v_{R_\alpha(P)}(s) &= \alpha v_{D(P)}(s) + (1-\alpha) v_{R(P)}(s) \\ &\geq \alpha v_{D(P')}(s) + (1-\alpha)(-1) \\ &= (1-\delta)\alpha - (1-\alpha) = (2-\delta)\alpha - 1 \end{aligned}$$

So, if we set $\alpha \in (0,1)$ so that $(2-\delta)\alpha - 1 > 0$, $R_\alpha$ will satisfy Endorsed unanimity. Since $\delta < \epsilon$, as close as possible, imposing $\alpha \geq \frac{1}{2-\epsilon} > \frac{1}{2-\delta}$ the property is satisfied.

□

**Proposition B.3.12.** For any $\epsilon \in (0,1)$, and considering the domain to be $\epsilon$-coherent, $R_\alpha$ does not fulfil the following properties for any $\alpha \in (0,1)$:

(i) Sided unanimity, and therefore Narrow unanimity;

(ii) Familiar monotonicity, and therefore Monotonicity; and

(iii) Independence.

*Proof.* (i) Sided unanimity and Narrow unanimity. It suffices to build a DRF and an opinion profile for which Sided unanimity does not hold for any $\alpha, \epsilon \in (0,1)$.

Consider the DRF and opinion profile $P$ depicted in Figure B.32, where: $x \in (0,1)$ is such that $0 < x < \frac{1-\alpha}{\alpha}$, $0 < \delta < \epsilon$; $m \in \mathbb{N}$ satisfies $(m-1)\delta \leq 1+ < m\delta$; and for any $r \in R$, $w(r) = 1$.

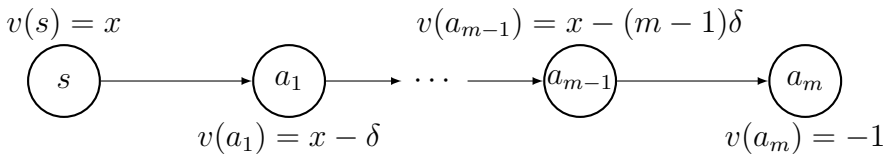



Figure B.32: Counterexample for Sided and Narrow unanimity in proposition B.3.12.

Clearly, $P$ is an $\epsilon$-coherent because $|v(a_m) - e(a_m)| = 0$, and for any $i < m$, $|v(a_i) - e(a_i)| = v(a_i) - v(a_{i+1}) = \delta < \epsilon$, and $|v(s) - e(s)| = x - x + \delta < \epsilon$. Furthermore, $P$ satisfies the assumptions of Sided unanimity at $s$ since $v(s) = x > 0$. It is straightforward to see that $v_{D(P)}(s) = x$ and $v_{R(P)}(s) = v_{R(P)}(a_m) = -1$. Hence, $v_{R\alpha(P)}(s) = x\alpha + (1-\alpha)(-1) = x\alpha + \alpha - 1$. But since $x < \frac{1-\alpha}{\alpha}$, we conclude that $v_{R\alpha(P)}(s) < \frac{1-\alpha}{\alpha}\alpha + \alpha - 1 = 0$. This proves that Sided unanimity is not fulfilled, and as it is a single-opinion profile Narrow unanimity fails too. We can proceed analogously for the negative case of Sided unanimity.

(ii) Familiar monotonicity. The counterexample employed in proposition B.3.6 to show that Familiar monotonicity does not hold for $R$ serves here as well to prove that $R_\alpha$ does not satisfy Familiar monotonicity for any $\alpha \in (0, 1)$. From opinion profiles $P$ and $P'$ depicted in figures B.27 and B.28 respectively, we extract that $v_{D(P)}(s) = v_{D(P')}(s) = 1$ and $1 = v_{R(P)}(s) > v_{R(P')}(s) = 1 - x$. Therefore, it follows that $v_{R_\alpha(P)}(s) > v_{R_\alpha(P')}(s)$, hence proving that Familiar monotonicity does not hold.

(iii) Independence. For any $\alpha \in (0, 1)$, function $R_\alpha$ does not fulfil Independence due to its dependence on $R$.

□

## B.4 Constrained opinion profiles: assuming consensus on acceptance degrees and coherent profiles

Next, we show the results of our fourth and last scenario. We now assume that opinion profiles are both $\epsilon$-coherent, for some $\epsilon \in (0, 1)$, and agree on their acceptance degrees over relationships. The results that follow are summarised in Table 7.3 in Section 7.4.4.

As in previous sections, next we only prove per aggregation function those properties that either were partially satisfied or not satisfied at all in previous scenarios but do hold in this new scenario. We do not prove those properties for which the proofs in the previous sections serve as well for this scenario.

**Proposition B.4.1.** Let be a $DRF$ and an opinion profile $P = (O_1, \ldots, O_n)$. For any $s \in \mathcal{S}$, assume that for each $r \in R^+(s)$ $w_i(r) = \lambda_r \in (0, 1]$ for any $i$, then:

(i) For any $\epsilon \in (0, 1)$, if $0 < \delta \leq \epsilon$ and the domain $\mathcal{D}$ is $\delta$-coherent then $D(P)$ is $\epsilon$-coherent, so satisfies $\epsilon$-Collective coherence.

(ii) For any $\epsilon \in (0, 1)$, if $0 < \delta \leq \epsilon$ and the domain $\mathcal{D}$ is $\delta$-coherent then $I(P)$ is $\epsilon$-coherent, so satisfies $\epsilon$-Collective coherence.

(iii) For any $\epsilon \in (0, 1)$, if $0 < \delta \leq \epsilon$ and the domain $\mathcal{D}$ is $\delta$-coherent then $B_\alpha(P)$ is $\epsilon$-coherent for any $\alpha \in (0, 1)$, so satisfies $\epsilon$-Collective coherence.

(iv) For any $\epsilon \in (0, 1)$, if $0 < \delta \leq \epsilon$ and the domain $\mathcal{D}$ is $\delta$-coherent then $R_\alpha(P)$ is $\epsilon$-coherent for any $\alpha \in (0, 1)$, so satisfies $\epsilon$-Collective coherence.

*Proof.* (i) Collective Coherence of $D$. Let $s \in \mathcal{S}$. We assume that for any $i$, $|v_i(s) - e_i(s)| < \delta \leq \epsilon$. Next we calculate the coherence condition for $D$ at sentence $s$:

$$
\begin{aligned}
|v_{D(P)}(s) - e_{D(P)}(s)| \;=\; & \Bigg| \frac{1}{n} \sum_i v_i(s) \\
& - \frac{1}{\sum_{r \in R^+(s)} w_{D(P)}(r)} \sum_{r \in R^+(s)} w_{D(P)}(r) v_{D(P)}(s_r) \Bigg| \\
= \; & \left| \frac{1}{n} \sum_i v_i(s) - \frac{1}{\sum_{r \in R^+(s)} \lambda_r} \sum_{r \in R^+(s)} \lambda_r \big( \frac{1}{n} \sum_i v_i(s_r) \big) \right| \\
= \; & \left| \frac{1}{n} \sum_i \Big( v_i(s) - \frac{1}{\sum_{r \in R^+(s)} w_i(r)} \sum_{r \in R^+(s)} w_i(r) v_i(s_r) \Big) \right| \\
= \; & \left| \frac{1}{n} \sum_i \Big( v_i(s) - e_i(s) \Big) \right| \\
\leq \; & \frac{1}{n} \sum_i \Big| v_i(s) - e_i(s) \Big|
\end{aligned}
$$

Thus, by $\delta$-coherence of the domain we obtain that:

$$|v_{D(P)}(s) - e_{D(P)}(s)| \leq \frac{1}{n}\sum_i \Big|v_i(s) - e_i(s)\Big| < \frac{1}{n}\sum_i \delta \leq \epsilon$$

This proves that the collective opinion by $D$ is $\epsilon$-coherent.

(ii) Collective Coherence of $I$. We prove collective coherence for $I$ similarly to the proof above for $D$. Let $s \in \mathcal{S}$. We assume that for any $i$, $|v_i(s) - e_i(s)| < \delta \leq \epsilon$. We compute the condition for the collective coherence of $I$ at sentence $s$ as follows:

$$\begin{aligned}
|v_{I(P)}(s) - e_{I(P)}(s)| \;&=\; \Bigg|\frac{1}{n}\sum_i e_i(s) \\
&\qquad - \frac{1}{\sum_{r\in R^+(s)} w_{I(P)}(r)} \sum_{r\in R^+(s)} w_{I(P)}(r) v_{I(P)}(s_r)\Bigg| \\
&= \left|\frac{1}{n}\sum_i e_i(s) - \frac{1}{\sum_{r\in R^+(s)} \lambda_r} \sum_{r\in R^+(s)} \lambda_r \big(\frac{1}{n}\sum_i e_i(s_r)\big)\right| \\
&= \left|\frac{1}{n}\sum_i \Big(e_i(s) - \frac{1}{\sum_{r\in R^+(s)} w_i(r)} \sum_{r\in R^+(s)} w_i(r) e_i(s_r)\Big)\right| \\
&= \left|\frac{1}{n}\sum_i \frac{\sum_{r\in R^+(s)} w_i(r) v_i(s_r) - \sum_{r\in R^+(s)} w_i(r) e_i(s_r)}{\sum_{r\in R^+(s)} w_i(r)}\right| \\
&= \left|\frac{1}{n}\sum_i \sum_{r\in R^+(s)} \frac{w_i(r)\Big(v_i(s_r) - e_i(s_r)\Big)}{\sum_{r\in R^+(s)} w_i(r)}\right|
\end{aligned}$$

So, by $\delta$-coherence of the domain, we obtain that:

$$|v_{D(P)}(s) - e_{D(P)}(s)| \;\leq\; \frac{1}{n}\sum_i \sum_{r\in R^+(s)} \frac{w_i(r)\Big|v_i(s_r) - e_i(s_r)\Big|}{\sum_{r\in R^+(s)} w_i(r)}$$

$$< \frac{1}{n}\sum_i \frac{\sum_{r\in R^+(s)} w_i(r)\,\delta}{\sum_{r\in R^+(s)} w_i(r)} \;=\; \frac{1}{n}\sum_i \delta \;=\; \delta \leq \epsilon$$

This proves that the collective opinion by $I$ is $\epsilon$-coherent.

(iii) Collective Coherence of $B_\alpha$. We have just proven that $D$ and $I$ satisfy $\epsilon$-collective coherence assuming consensus on acceptance degrees and a $\delta$-coherent domain with $\delta < \epsilon$. It directly follows that for any $\alpha \in (0, 1)$, then $B_\alpha$ on a $\delta$-coherent domain also satisfies $\epsilon$-collective coherence.

(iv) Collective Coherence of $R_\alpha$. We have proven that $D$ satisfies $\epsilon$-collective coherence assuming consensus on acceptance degrees and a $\delta$-coherent domain. $\epsilon$-collective coherence also holds for $R$ under the same assumptions following proposition B.1.3 (see collective coherence for $R$). Hence, it follows that for any $\alpha \in (0, 1)$, $R_\alpha$ on a $\delta$-coherent domain also satisfies $\epsilon$-collective coherence.

□

# Bibliography


Amgoud, L. and Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215. 55

Amgoud, L., Cayrol, C., Lagasquie-Schiex, M.-C., and Livet, P. (2008). On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093. 21, 41, 42

Amgoud, L., Maudet, N., and Parsons, S. (2000). Modelling dialogues using argumentation. In *Proceedings of the 4th International Conference on MultiAgent Systems*, pages 31–38. IEEE. 41

Appgree (Last visited 2021-04-3-11-30). `http://www.appgree.com/`. 16, 38

Aragón, P. (2019). *Characterizing online participation in civic technologies*. PhD thesis, Universitat Pompeu Fabra. Departament de Tecnologies de la Informació i les Comunicacions. 50, 176

Arrow, K. (1963). *Social Choice and Individual Values*. Cowles Foundation Monograph Series. Yale University Press. 17

Arrow, K. J. and Maskin, E. S. (2012). *Social Choice and Individual Values: Third edition*. Yale University press. 44

Arrow, K. J., Sen, A., and Suzumura, K. (2002). *Handbook of Social Choice and Welfare*, volume 1 of *Handbooks in Economics*. North Holland/Elsevier. 51

Awad, E., Bonnefon, J.-F., Caminada, M., Malone, T. W., and Rahwan, I. (2017). Experimental assessment of aggregation principles in argumentation-enabled collective intelligence. *ACM Transactions on Internet Technology*, 17(3). 45

Awad, E., Booth, R., Tohmé, F., and Rahwan, I. (2015). Judgment aggregation in multi-agent argumentation. *Journal of Logic and Computation*, 27(1):227–259. 17, 18, 19, 21, 22, 23, 30, 33, 34, 35, 41, 42, 43, 44, 45, 46, 47, 48, 51, 55, 56, 60, 63, 67, 70, 71, 72, 73, 74, 75, 76, 81, 84, 91, 99, 100, 101, 111, 120, 121, 122, 123, 124, 150, 154, 155, 156, 158, 160, 161, 162, 163, 164, 175, 179, 180, 181

Bachtiger, A., Shikano, S., Pedrini, S., and Ryser, M. (2009). Measuring Deliberation 2.0: Standards, Discourse Types, and Sequenzialization. *European Consortium for Political Research General Conference*. 48

Baroni, P., Caminada, M., and Giacomin, M. (2011). An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410. 22, 41, 42, 43, 48, 60, 63, 111

Baroni, P. and Giacomin, M. (2009). Semantics of abstract argument systems. In Simari, G. and Rahwan, I., editors, *Argumentation in Artificial Intelligence*, pages 25–44. Springer US. 40, 41

Baroni, P., Romano, M., Toni, F., Aurisicchio, M., and Bertanza, G. (2015). Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49. 39

Bench-Capon, T. J. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448. 41, 43

Benn, N. and Macintosh, A. (2011). Argument visualization for eparticipation: towards a research agenda and prototype tool. In *International Conference on Electronic Participation*, pages 60–73. Springer. 39

Besnard, P. and Hunter, A. (2001). A logic-based theory of deductive arguments . *Artificial Intelligence*, 128(1-2):203–235. 40, 41, 42

Better Reykjavík (Last visited 2021-04-03-11-30). `http://reykjavik.is/en/better-reykjavik-0`. 16, 37, 55

Bochman, A. (2003). Collective argumentation and disjunctive logic programming. *Journal of Logic and Computation*, 13(3):405–428. 42

Bodanza, G., Tohmé, F., and Auday, M. (2017). Collective argumentation: A survey of aggregation issues around argumentation frameworks. *Argument & Computation*, 8(1):1–34. 45

Cabrio, E. and Villata, S. (2013). A natural language bipolar argumentation approach to support users in online debate interactions. *Argument & Computation*, 4(3):209–230. 39

Caminada, M. (2006). On the issue of reinstatement in argumentation. In *European Workshop on Logics in Artificial Intelligence*, pages 111–123. Springer. 18, 21, 30, 32, 33, 41, 43, 51, 60, 67, 91, 154, 158, 159, 161, 164, 179, 181

Caminada, M. and Pigozzi, G. (2011). On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102. 23, 42, 45, 46, 47, 48, 51, 181

Caminada, M. W. and Gabbay, D. M. (2009). A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145. 60

Carr, C. S. (2003). Using computer supported argument visualization to teach legal argumentation. In *Visualizing argumentation: Software tools for collaborative and educational sense-making*, pages 75–96. Springer. 38

Cayrol, C. and Lagasquie-Schiex, M. (2005). Gradual valuation for bipolar argumentation frameworks. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 366–377. Springer. 21, 41

Cayrol, C. and Lagasquie-Schiex, M.-C. (2005). On the acceptability of arguments in bipolar argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 378–389. Springer. 21, 41, 42

Chen, W. and Endriss, U. (2019). Preservation of semantic properties in collective argumentation: The case of aggregating abstract argumentation frameworks. *Artificial Intelligence*, 269:27–48. 23, 45, 46, 47, 48, 51, 181

City of Helsinki (Last visited 2021-04-03-11-30). https://www.hel.fi/helsinki/en/administration/participate/channels/participation-model/. 16, 38

Consider.it (Last visited 2021-04-03-11-30). https://consider.it/. 16, 38

Consul (Last visited 2021-04-03-11-30). http://consulproject.org/en/. 38, 40, 51

Coste-Marquis, S., Devred, C., Konieczny, S., Lagasquie-Schiex, M.-C., and Marquis, P. (2007). On the merging of Dung's argumentation systems. *Artificial Intelligence*, 171(10):730–753. 41, 44, 51, 55, 99, 162

De Vries, R., Stanczyk, A. E., Ryan, K. A., and Kim, S. Y. (2011). A framework for assessing the quality of democratic deliberation: Enhancing deliberation as a tool for bioethics. *Journal of Empirical Research on Human Research Ethics*. 48

Decide Madrid (Last visited 2021-04-03-11-30). https://decide.madrid.es/. 16

Decidim Barcelona (Last visited 2021-04-03-11-30). https://www.decidim.barcelona/. 11, 16, 28, 37, 38, 39, 40, 51, 55, 93, 184

Dietrich, F. (2007). A generalised model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565. 71, 120

Dietrich, F. and List, C. (2007). Strategy-proof judgment aggregation. *Economics & Philosophy*, 23(3):269–300. 48

Dietrich, F. and Mongin, P. (2010). The premiss-based approach to judgment aggregation. *Journal of Economic Theory*, 145(2):562–582. 47, 84

Dung, P. M. (1993). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. In *Proceedings of the 13th International Joint Conference on Artifical Intelligence - Volume 2*, IJCAI'93, page 852–857, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc. 63

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*,

77(2):321–357. 18, 21, 22, 26, 30, 31, 32, 33, 40, 41, 42, 43, 45, 51, 55, 56, 63, 67, 91, 99, 101, 150, 152, 158, 159, 160, 161, 162, 163, 164, 179, 180, 181

Dung, P. M., Kowalski, R. A., and Toni, F. (2006). Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170(2):114–159. 40

Dunne, P. E., Hunter, A., McBurney, P., Parsons, S., and Wooldridge, M. (2011). Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486. 42, 43, 100

Endriss, U. and Grandi, U. (2017). Graph aggregation. *Artificial Intelligence*, 245:86–114. 45, 46, 47

Endriss, U. and Moulin, H. (2016). Judgment aggregation. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, *Handbook of Computational Social Choice*, pages 399–426. Cambridge University Press. 17, 46, 51

Fox, J., Barber, D., and Bardhan, K. (1980). Alternatives to Bayes? A quantitative comparison with rule-based diagnosis. *Methods of Information in Medicine*, 19(04):210–215. 40

Fox, J., Krause, P., and EIvang-Gøransson, M. (1993). Argumentation as a general framework for uncertain reasoning. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 428–434. Elsevier. 40

Freeman, S. (2000). Deliberative democracy: A sympathetic comment. *Philosophy & Public Affairs*, 29(4):371–418. 36

Friess, D. and Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3):319–339. 24, 35, 36, 44, 48

Fung, A. and Wright, E. O. (2001). Deepening democracy: Innovations in empowered participatory governance. *Politics & Society*, 29(1):5–41. 16

Gaertner, W. (2009). *A Primer in Social Choice Theory: Revised edition*. Oxford University Press. 44

Gallo, G., Longo, G., Pallottino, S., and Nguyen, S. (1993). Directed hypergraphs and applications. *Discrete Applied Mathematics*, 42(2-3):177–201. 142

Ganzer, J., Criado, N., Lopez-Sanchez, M., Rodriguez-Aguilar, J. A., and Parsons, S. (2017). Collective decision making library. `https://bitbucket.org/jariiia/argumentation-for-collective-decision-making/src/master/`. 87

Ganzer-Ripoll, J., Criado, N., Lopez-Sanchez, M., Parsons, S., and Rodriguez-Aguilar, J. A. (2019). Combining social choice theory and argumentation: Enabling collective decision making. *Group Decision and Negotiation*, 28(1):127–173. 99, 100, 101

García, A. J. and Simari, G. (2004). Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(1):95–138. 40, 99

Gómez, V., Kaltenbrunner, A., and López, V. (2008). Statistical analysis of the social network and discussion threads in slashdot. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, page 645–654, New York, NY, USA. Association for Computing Machinery. 49, 50, 51, 176

Gonzalez-Bailon, S., Kaltenbrunner, A., and Banchs, R. E. (2010). The structure of political discussion networks: A model for the analysis of online deliberation. *Journal of Information Technology*, 25(2):230–243. 49, 50, 51, 176

Grandi, U., Lorini, E., and Perrussel, L. (2015). Propositional opinion diffusion. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '15, page 989–997, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems. 62

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*. 50

Iandoli, L., Quinto, I., De Liddo, A., and Shum, S. B. (2014). Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. *International Journal of Human-Computer Studies*, 72(3):298–319. 39

Jackson, S. K. and Kuehn, K. M. (2016). Open source, social activism and "necessary trade-offs" in the digital enclosure: A case study of platform co-operative, Loomio.org. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 14(2):413–427. 38

Joseph, S. and Prakken, H. (2009). Coherence-driven argumentation to norm consensus. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 58–67. 42

Kahn, A. B. (1962). Topological sorting of large networks. *Communications of the ACM*, 5(11):558–562. 86

Kialo (Last visited 2021-04-03-11-30). `https://www.kialo.com/`. 38, 93

Klein, M. (2012). Enabling large-scale deliberation using attention-mediation metrics. *Computer Supported Cooperative Work*, 21(4-5):449–473. 18, 39, 40, 51, 60, 93

Klein, M. and Convertino, G. (2015). A roadmap for open innovation systems. *Journal of Social Media for Organizations*, 2(1):1. 39

Krause, P., Ambler, S., Elvang-Goransson, M., and Fox, J. (1995). A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11(1):113–131. 40

Landemore, H. and Page, S. E. (2015). Deliberation and disagreement: Problem solving, prediction, and positive dissensus. *Politics, Philosophy and Economics*, 14(3):229–254. 36

Lang, J., Slavkovik, M., and Vesic, S. (2016). Agenda separability in judgment aggregation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, pages 1016–1022. 47, 84

Leite, J. and Martins, J. (2011). Social abstract argumentation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, volume 11, pages 2287–2292. 17, 18, 41, 42, 46, 47, 55, 99, 100, 180

List, C. (2018). Democratic deliberation and social choice. *The Oxford Handbook of Deliberative Democracy*, page 463. 17, 35, 36, 44, 45

List, C. and Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18(01):89–110. 17, 23, 70, 119, 182

Loui, R. P. (1987). Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3(3):100–106. 40

Manosevitch, E., Steinfeld, N., and Lev-On, A. (2014). Promoting online deliberation quality: cognitive cues matter. *Information, Communication & Society*, 17(10):1177–1195. 49, 50

McBurney, P. (2002). *Rational Interaction*. PhD thesis, Department of Computer Science, University of Liverpool. 41

McBurney, P. and Parsons, S. (2009). Dialogue games for agent argumentation. In *Argumentation in artificial intelligence*, pages 261–280. Springer. 41

McGuire, R., Birnbaum, L., and Flowers, M. (1981). Opportunistic processing in arguments. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 58–60. 40

Menéame.net (Last visited 2021-04-03-11-30). `https://www.meneame.net/`. 50

Modgil, S. and Caminada, M. (2009). Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer. 40

Modgil, S. and Prakken, H. (2013). A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397. 40, 41, 99

Mongin, P. (2008). Factoring out the impossibility of logical aggregation. *Journal of Economic Theory*, 141(1):100–113. 47, 84

New York City Participatory Budgeting (Last visited 2021-04-03-11-30). `https://pbnyc.participatorybudgeting.org/budgets`. 38

Parlement & Citoyens (Last visited 2021-04-03-11-30). `https://parlement-et-citoyens.fr/`. 16, 37

Parsons, S. (1997). Normative argumentation and qualitative probability. In Gabbay, D. M., Kruse, R., Nonnengart, A., and Ohlbach, H. J., editors, *Qualitative and Quantitative Practical Reasoning*, volume 1244, pages 466–480, Berlin, Heidelberg. Springer. 40

Pigozzi, G., Slavkovik, M., and van der Torre, L. (2008). Independence in judgment aggregation. In *Proceedings of the Ninth International Meeting of the Society for Social Choice and Welfare*. 47, 84

Prakken, H. (2005). A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 85–94. 41

Prakken, H. and Sartor, G. (1997). Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7(1-2):25–75. 40

Rago, A. and Toni, F. (2017). Quantitative argumentation debates with votes for opinion polling. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 369–385. Springer. 18, 42, 46, 48, 51, 101

Rahwan, I. and Simari, G. R., editors (2009). *Argumentation in Artificial Intelligence*, volume 47. Springer. 17, 40

Rahwan, I. and Tohmé, F. (2010). Collective argument evaluation as judgement aggregation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, volume 1 of *AAMAS '10*, pages 417–424, Toronto, Canada. International Foundation for Autonomous Agents and Multiagent Systems. 44

Rajendran, P., Bollegala, D., and Parsons, S. (2016). Assessing weight of opinion by aggregating coalitions of arguments. In *Proceedings of the 6th International Conference on Computational Models of Argument*, pages 431–438. 39

Reed, C. and Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979. 39

Rinner, C. (2006). Argumentation mapping in collaborative spatial decision making. In *Collaborative geographic information systems*, pages 85–102. IGI Global. 39, 40

Rodriguez-Aguilar, J. A., Serramià, M., and Lopez-Sanchez, M. (2016). Aggregation operators to support collective reasoning. In Torra, V., Narukawa, Y., Navarro-Arribas, G., and Yañez, C., editors, *Modeling Decisions for Artificial Intelligence*, pages 3–14, Cham. Springer International Publishing. 39, 41, 42, 44, 46

Serramià, M., Ganzer, J., Lopez-Sanchez, M., Rodriguez-Aguilar, J. A., Criado, N., Parsons, S., Escobar, P., and Fernández, M. (2019). Citizen support aggregation methods for participatory platforms. In *Frontiers in Artificial Intelligence and Applications: Artificial Intelligence Research and Development.*, volume 319, pages 9–18. IOS Press. 39, 163, 184, 185

Shum, S. B. (2003). The roots of computer supported argument visualization. In *Visualizing argumentation: Software tools for collaborative and educational sense-making*, pages 3–24. Springer. 38

Steenbergen, M., Bächtiger, A., Spörndli, M., and Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics*, 1:21–48. 49, 50

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation*, 3(1). 49, 50

Suthers, D., Weiner, A., Connelly, J., and Paolucci, M. (1995). Belvedere: Engaging students in critical discussion of science and public policy issues. In *Proceedings of the 7th World Conference on Artificial Intelligence in Education*, pages 266–273. Washington, DC. 38

Sycara, K. P. (1990). Persuasive argumentation in negotiation. *Theory and decision*, 28(3):203–242. 41

Thagard, P. (2002). *Coherence in thought and action*. MIT Press. 72, 112, 121

Trénel, M. (2004). Measuring the Deliberativeness of Online Discussions. Coding Scheme 2.2. *Report, Berlin: Social Science Research Centre*. 49, 50

Van Gelder, T. (2003). Enhancing deliberation through computer supported argument visualization. In *Visualizing argumentation: Software tools for collaborative and educational sense-making*, pages 97–115. Springer. 38

Verheij, B. (1995). Accrual of arguments in defeasible argumentation. In *Proceedings of the Second Dutch/German Workshop on Nonmonotonic Reasoning*, pages 217–224, Utrecht. 41

Vreeswijk, G. A. (1997). Abstract argumentation systems. *Artificial intelligence*, 90(1-2):225–279. 40

Walton, D. and Krabbe, E. C. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press. 41