CrossMark

Combining Social Choice Theory and Argumentation: Enabling Collective Decision Making

Jordi Ganzer-Ripoll[1] · Natalia Criado[1] · Maite Lopez-Sanchez[2] ·
Simon Parsons[1] · Juan A. Rodriguez-Aguilar[3]

© The Author(s) 2018

Abstract

Argumentation-based debates are mechanisms that a group can use to resolve conflicting opinions and hence reach agreement. They have many potential applications in on-line communities and other open environments. In this paper, we provide computational infrastructure to support argumentation-based debates, in particular focusing on the problem of how participants in a debate can reach agreement about the outcome of the debate, given all the statements that have been made. Our approach makes it possible to represent arguments that are put forward by the participants in a debate, allows both positive and negative relationships between the arguments to be represented, and makes it possible for participants to express opinions about both the arguments and the outcome of the debate. Our main contribution is to provide a novel method—indeed the first method—for computing the collective decision that emerges from the combination of a set of arguments and a set of opinions about whether the arguments hold or not. To do this, we carry out a formal investigation of a family of aggregation functions. This family starts with a function that is firmly rooted in the social choice literature, and is extended with functions that are more oriented towards the use of argumentation. We prove that to ensure that the collective decision is coherent, a property that we think is essential, an aggregation function needs to take into account the dependencies between arguments. We also provide an empirical analysis of the performance of our approach to reaching a collective decision, showing that a collective decision can be reached for debates, of the size that one currently finds online, in reasonable time.

Keywords Argumentation · Social choice · Collective decision making · On-line debates · Social intelligence

✉ Juan A. Rodriguez-Aguilar
jar@iiia.csic.es

Extended author information available on the last page of the article

🍃 Springer

1 Introduction

Across the world there is an increased interest in the process of using technology to provide a route for the greater engagement of citizens in the governance of their communities. The idea is that thanks to the use of so-called *e-governance* or, more specifically, *e-participation* systems, individuals will be able to have a greater say in the way that they are governed (Weerakkody and Reddick 2012), and decisions taken by their elected representations will more closely reflect the views of the represented (Dawes 2008). For example, both Barcelona and Reykjavík municipalities have opened up some aspects of their policy making to citizens through the use of online portals. These portals, Decidim Barcelona (2017) and Better Reykjavík (2017), respectively, allow individual citizens to put forward policy proposals, state their support for proposals made by their peers, and debate the pros and cons of proposals made by themselves and by others.

Decidim and Better Reykjavík, by their restriction to the cities in question, focus on local issues and have limited reach into the broader communities. However, there is no reason why such experiments in participatory democracy have to be geographically limited. Indeed, the ambitious Parlement et Citoyens project in France (2017) aims to scale this kind of activity up to the national level, allowing citizens and deputies to collaboratively draft proposals for legislation.

All these efforts can be seen as extensions of earlier e-governance efforts, such as the UK's online petition site (Petitions 2017), along the lines of online collaboration platforms intended to support small group discussions. The UK petition site allows citizens and residents to request that parliament consider a topic. It is a one way mechanism—if a petition collects 10,000 signatures, then the government responds. If a petition collects 100,000 signatures (as did the recent petition to deny Donald Trump a state visit), then parliament debates the issue. However, there is no ability for citizens to directly discuss issues amongst themselves, nor is there a facility to engage in debate with lawmakers. It is this ability to engage in structured debates that, for us, is the key element in Decidim, Better Reykjavík and Parlement et Citoyens. The structure comes from an initial proposal, which participants can then subject to scrutiny, offering arguments for and against the central issue. This is a structure that they have in common with sites like Quoners (2017) and `consider.it` (2017) which are not tied to a particular institution, and tools like the Deliberatorium (Klein 2012, 2017).

The work described in this paper is inspired by these kinds of e-participation systems, and explores computational mechanisms for evaluating the output of these systems. In other words, we define and formally analyze a computational mechanism that can take the output of an e-participation system—a number of arguments about a proposal and opinions about whether or not those arguments have value—and establish what the balance of opinion is. Indeed, the scenario that we investigate is somewhat more general than that supported by the systems mentioned above. That is because existing e-participation systems are limited to either providing a list of arguments for a proposal and a list of argument against, so that there are no relationships between the list of arguments, or providing a forum-like setting where arguments are structured in a tree. In contrast, we allow for a more general discussion.

In particular, in this paper we do the following:

– We introduce a novel, formal framework—which we call a *target-oriented discussion framework*—that can support discussions about whether some proposal should be accepted or not. (The proposal is the "target".) This allows participants in a debate to put forward arguments for and against the target, and to indicate the relationships between those arguments. (If argument *a* is in favour of the target and argument *b* is against it, then *a* and *b* may conflict, and our framework allows this to be recorded).

– We provide the means for individuals to express their opinions about the proposal and the arguments that have been put forward in our framework—both arguments that they have put forward and arguments put forward by others—along with a method that can assess whether or not the set of opinions as a whole is reasonable. This notion of reasonableness is established by the formal notion of a *coherent labelling*. A coherent labelling can be thought of as a relaxed variant of the standard argumentation notion of a complete labelling (Baroni et al. 2011) which provides further flexibility in expressing opinions.[1]

– We introduce and formally evaluate a family of aggregation functions that take a set of arguments and opinions about those arguments and return the collective decision about the target. We investigate the properties of these aggregation functions by borrowing from, and extending, the adaptation of classical properties from social-choice theory to the domain of argumentation that was carried out in Awad et al. (2017). We find that the aggregation functions that we introduce span a range of properties—summarised in Table 3—illustrating the trade-offs between those properties. We prove that two of the functions guarantee that the outcome satisfies the property of *coherent collective rationality*, meaning that they generate a coherent labelling, so that the collective opinion from the discussion is coherent—and this is the case even when aggregating individual opinions that may not be coherent.

As the terminology we have used so far suggests, our work draws both on *argumentation theory* and *social choice theory*. We consider that the statements made by participants in discussion are arguments. That is they are statements that include a conclusion and the reason why that conclusion is considered to hold. (However, like much work in argumentation (Dung 1995; Modgil and Caminada 2009), we will deal with them at a purely abstract level, meaning that we treat them as if they were atomic objects with no internal structure.) As in standard argumentation theory (Dung 1995), we consider that arguments can conflict, that they *attack* each other, for example if their conclusions are contradictory, but we also consider that arguments may *defend* one another, as in Cayrol and Lagasquie-Schiex (2005b). Unlike most work in argumentation theory, we are not directly concerned with computing the acceptability of the arguments in the discussion using some version of the standard semantics (Baroni et al. 2011; Baroni and Giacomin 2009). Rather, we are interested in establishing which arguments the participants in the discussion consider to hold,[2] and doing this is where the ideas from social choice theory come in.

---

[1] As we point out below, this additional flexibility arises because we do not require participants in debates to conform to the rules for labelling developed in argumentation. We want to allow debaters to be *irrational* in this sense, when they assign labels to arguments.

[2] Naturally, since we are establishing which arguments hold on the basis of the views of the participants in the discussion, rather than the rational computation of one of the standard semantics, the results we get

We consider the problem of merging opinions about arguments in a debate to be an instance of collective decision-making as studied in social-choice theory. Given a set of agents, and a set of arguments about a topic, where each agent may have an opinion about whether or not the arguments hold, we are interested in how the agents can, as a group, reach a decision about the topic. To deal with this situation, we consider a family of novel functions which aggregate the opinions of the agents to compute the overall opinion on the topic. Therefore, our work tackles the same problem that Awad et al. posed in Awad et al. (2017), and we also encode agents' opinions about arguments using a set of non-binary labellings just as in Awad et al. (2017). However, our work takes an important step beyond (Awad et al. 2017) since we establish the overall labelling over the set of arguments without assuming independence between arguments as Awad et al. (2017) does (and as the judgement aggregation and preference aggregation literature in general does). Instead we adhere here to the important observation made in Awad et al. (2017), where the authors note that assuming independence is questionable because (as noted above), it is natural for a set of arguments to have support and attack relationships between them. Since these relationships exist, it is logical for the process of merging opinions about arguments to take these relationships between the arguments into consideration in some form. That is what two of our aggregation functions do, guaranteeing that the resulting aggregated opinion is *coherent*, that is, in some sense, free of contradictions. Coherence, in the sense we use the term, is a weaker condition than the conflict-freeness that is the standard minimum condition in argumentation theory, and our future work will look to establish whether other aggregation functions can attain a conflict-free set of opinions.

From a social choice perspective, it is important to note that unlike in the literature on judgement aggregation and preference aggregation, we do not impose the condition that opinions satisfy any property from which our aggregation functions can benefit to guarantee collective rationality. The rationale for this is clear, and the same as the rationale for not insisting on any conditions on the set of labellings that are given to the arguments by participants—since the agents that are involved in the discussions are humans, we cannot assume that they will have rational opinions. That is because we know that humans are frequently inconsistent when expressing their opinions, often contradicting themselves. We therefore believe that not assuming rationality is essential for our aggregation operators to be capable of being used in realistic settings.

Organisation. The paper is structured as follows. Section 2 surveys related work. The next two sections characterise, Sect. 3, and formalise, Sect. 4, our novel multi-agent discussion framework. Then, Sect. 5 details both the decision problem that we study and the desired properties of aggregation functions; Sect. 6 introduces a family of novel aggregation functions and studies their social-choice properties; Sect. 7 provides an algorithm for computing the collective decision of a discussion framework using the functions from Sect. 6 and tests how long it takes to compute a collective decision for

---

will differ from those that the standard semantics will generate. One might therefore consider the results irrational in some sense. However, as we discuss at length below, they are rational in the sense of (some aspects of) social choice theory. And we believe that highlighting the differences between the results as established by argumentation and the results as established by social choice theory, as our work will do in the long term, will be to the benefit of both those who study argumentation and those who study social choice theory.

realistically-sized discussion frameworks; and Sect. 8 draws conclusions and plans future research.

2 Related Work

We identify several broad research areas with connections to the topics that we discuss in this paper. These include tools for online discussion, computational argumentation, and social choice theory.

2.1 Tools for Online Discussion

As mentioned above, we see this work as being inspired by work on online discussion forums such as Decidim Barcelona (2017), Better Reykjavík (2017) and Parlement et Citoyens (2017), where participants can carry out a structured discussion around some topic, typically a policy proposal. These particular tools just allow participants to offer arguments for and against a proposal, and only in the context of a specific institution. Other approaches have extended the scope of these tools. One direction is in developing tools that are not tied to a specific institution. In this category of non-institutional tools we find Quoners (2017) and `consider.it` (2017), which we mentioned above, and Appgree (2017) and Baoqu (2017) where the main focus is on scalability—making the systems fit for use by large numbers of participants. Another direction is that of allowing participants to do more than just comment. Here we have the example of Jackson and Kuehn (2016) and Loomio (2017), where participants can both comment on proposals, albeit in an unstructured way, and also vote on them. What distinguishes our work from all of these approaches is that we aim to support discussions that are more than just structured—they are argument-based, and we take the interaction between the arguments into account.[3]

There are other approaches that allow for structured argument-based discussions. Most notable here is Klein's work on the Deliberatorium (Klein 2012; Klein and Convertino 2015) which allows for the presentation of arguments and their interactions. The Deliberatorium is part of a long line of work that allows to structure reasoning about complex scenarios in terms of arguments for and against options. Other work in this line is (Carr 2003; Reed and Rowe 2004; Suthers et al. 1995; Van Gelder 2003), where the focus is more on drawing the relationships between arguments as a means of helping people understand the scenarios. Our work differs from these approaches in its attempt to provide computational methods to summarise the information that has been put forward. In other words, our focus is on using the results of debate as input to a computational process, rather than providing support for the debate itself. In that sense our work could be viewed as a post-processing stage that could be applied in conjunction with any of the tools to support structured discussion.

---

[3] Although we acknowledge that we do not deal with the interactions as comprehensively as all work in argumentation does.

2.2 Computational Argumentation

Computational argumentation (Rahwan and Simari 2009) has a lengthy history within artificial intelligence. At the time of writing, it is hard to overstate the influence of the work of Dung (1995) which both introduced the idea of studying argumentation at the abstract level, that is without considering the structures from which arguments are constructed, and the idea of using argumentation as a way of establishing a consistent viewpoint from an inconsistent set of data.[4] Dung (1995) provided a number of methods—"semantics" as they are called—to extract a consistent set of arguments from a set of arguments that conflict with one another, and inspired much subsequent work on abstract argumentation systems (Baroni and Giacomin 2009; Modgil and Caminada 2009; Vreeswijk 1997). This includes work on bipolar argumentation (Amgoud et al. 2008; Cayrol and Lagasquie-Schiex 2005b), which includes a "support" relation between arguments.

However, work on abstract argumentation is only one aspect of work on argumentation. In fact, Dung (1995) was pre-dated by work that looked at decision-making as a process of putting forward reasons—arguments, though they were not called that at the time—for and against particular conclusion (Fox and Bardhan 1980). This approach was then refined into systems of argumentation such as (Fox et al. 1993) and (Krause et al. 1995). Such systems were precursors of work where the internal structure of an argument is important: logic-based argumentation (Besnard and Hunter 2001), assumption-based argumentation (Dung et al. 2006) and structured argumentation systems such as ASPIC+ (Modgil and Prakken 2013), and DeLP (García and Simari 2004). In these more subtle forms of argumentation, the focus is often still on establishing consistency—the difference with abstract argumentation is just that they don't consider arguments as primitive objects, rather arguments are constructed from sentences in some language.

Another line of work in computational argumentation, separate from that on establishing consistency, is that on argument *accrual* (Besnard and Hunter 2001; Cayrol and Lagasquie-Schiex 2005a; Fox and Bardhan 1980; Prakken 2005; Verheij 1995). Accrual involves the "summing up" of arguments, with the idea of establishing the strongest argument, sometimes in the face of arguments for and against some option, typically with the aim of being able to decide between some alternatives.

The work mentioned above uses argumentation as a mechanism for a single entity to come to a conclusion. However, as Sycara (1990), Walton and Krabbe (1995) and others have pointed out, argumentation is also a natural mechanism for multiple entities to use to reach consensus on some point. As a result, argumentation has been used (Amgoud et al. 2000; McBurney and Parsons 2009) in multiagent systems as a mechanism for *rational interaction* (McBurney 2002) for a particular meaning of "rational". That is "rational" in the sense that each stage in the interaction is supported by well-founded reasons. Here we build upon this prior work in rational interaction. Our approach allows agents to put forth arguments about some topic under discussion, be them either in favour or against the topic.

---

[4] Though this latter aspect clearly has a debt to (Lin and Shoham 1989).

Our work connects to several of these themes in argumentation. First, since we are interested in arguments from a number of participants, our work is clearly related to the use of argumentation in multiagent interaction. As we will see below, just like (Amgoud et al. 2000) and subsequent work, we assume a particular protocol for arguments to be posed, and we are interested in being able to compute the outcome of a discussion taking into account the arguments put forward by multiple participants.

Second, our work connects with the idea of argumentation as a means of extracting a coherent view from a number of conflicting arguments. One, commonly used, approach to doing this is the labelling approach (Baroni et al. 2011), which attaches the labels `in` (for arguments that should be accepted), `out` (for arguments that should be rejected) and `undec` (where the status cannot be decided). In this paper we borrow the idea of the labelling, but rather than finding one or more consistent labellings from the relations between arguments, we allow participants to indicate which labels they think apply to which arguments. In other words, we take participants' votes on what label should apply to what argument as input and from them compute a consensus labelling, where the consensus labelling need only be, in our terminology, "coherent". Coherency is formally defined in Sect. 4.3, but informally we can say that a labelling is coherent if every argument that is labelled `in` has more more arguments for it than against it, and if every argument that is labelled `out` has more arguments against it than for it. Coherency is a weaker requirement than the consistency applied by the standard approaches in argumentation (Baroni et al. 2011), thus setting our work apart from that on merging argumentation systems, for example (Coste-Marquis et al. 2007), which looks to construct one or more consistent merged labellings from several different consistent labellings.

Considering the input labellings as votes from human participants places our work in close relation to that of social argumentation (Leite and Martins 2011), though, unlike that work we take as input votes on the status of arguments rather than votes on the strength of the relation between arguments.

2.3 Social Choice Theory

Given a 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 choice that appropriately reflects the agents' individual preferences (Aziz et al. 2017). 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 the 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; etc.), 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 Sen 2002).

The work in this paper is in the vein of Awad et al. (2017). There, the authors pose the very same problem that we tackle here: *given a topic under discussion and a set of*

agents expressing their individual opinions about the arguments in the discussion, how can the agents reach a collectively rational decision? Likewise Awad et al., we consider that reaching a collective decision is a *judgement aggregation problem* in which the aggregation of opinions must satisfy desirable social choice properties. A further similarity with (Awad et al. 2017) stems from the way we encode opinions (subjective evaluations). Indeed, notice that while in judgement aggregation each proposition may take on one of two values (True or False), here, when aggregating labellings, each argument can take on one of three values (*in*, out, and undec). Therefore, aggregating labellings, as we do it here, has more in common with non-binary evaluations (Dokow and Holzman 2010).

Notwithstanding the similarities with Awad et al. (2017), there are several important differences with respect to that work and to the judgement aggregation literature as a whole. First of all, and very importantly, we do not assume independence between arguments as a fundamental postulate as is the case in Awad et al. (2017). As admitted by Awad et al., the necessity of independence is questionable because of the dependencies between arguments that come already encoded in the form of relationships such as attack. Despite that, they opt to stick with independence to keep open the possibility of proving strategy-proofness. Thus, they follow the usual methodology in judgement aggregation, though they do not establish the relation between independence and strategy-proofness. Indeed, independence is a fundamental property in the judgement aggregation literature because of its theoretical value in proving strategy-proofness and strategic manipulation. If the independence criterion is not satisfied, then the function aggregating judgements is not immune to strategic manipulation (Dietrich and List 2007). However, independence is not always upheld. On the one hand, from a theoretical point of view, independence is regarded as too strong a property, since, together with mild further conditions, it implies dictatorship (Lang et al. 2016). Furthermore, it is also considered as not very plausible (Mongin 2008). Hence, the theoretical and computational benefits of relaxing independence have been subject of much research (see e.g. Dietrich and Mongin 2010; Lang et al. 2016; Mongin 2008; Pigozzi et al. 2008).

Against this background, and given that *dependencies do exist between arguments*, our work departs from and goes beyond (Awad et al. 2017) by dropping independence. Thus, the aggregation functions that we introduce in this paper exploit *dependencies* between arguments and combine agents' opinions to yield an aggregated opinion. To the best of our knowledge, we are the first to take this step in a multi-agent argumentation context[5].

A second, major difference has to do with the approach chosen to achieve collective rationality. Here we focus on *designing novel aggregation functions* that exploit dependencies between arguments to ensure collective coherence. Instead, Awad et al. are concerned with characterising the restrictions that are necessary so that the plurality rule, a well-known voting function in the literature, produces collectively rational outcomes. Notice that Awad et al. study social choice properties satisfied by the plurality rule by adapting various classical social-choice theoretic properties (see e.g. Arrow

---

[5] While the work of Coste-Marquis et al. (2007) and others on merging argumentation systems could be considered work on combining opinions in multi-agent argumentation this work is based on methods form argumentation theory. We come at it from a different perspective.

and Sen 2002; Arrow et al. 2010) to the argumentation domain. Here we borrow some of those properties to study our aggregation operators. Nonetheless, since some of those properties assume independence and we do not, we define further social choice properties that take into account dependencies between arguments.

More recently, in the intersection of social choice theory and argumentation, we find the interesting work in Rago and Toni (2017). Similarly to our work, 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). Nonetheless QuAD-V does not allow arguments to be attackers and defenders at the same time. Although Rago and Toni (2017) propose the QuAD-V algorithm to determine a collective decision from multiple opinions by exploiting the dependencies between arguments, their goal is rather different from ours. Thus, they focus on the debate procedure (opinion polling) to ensure that, at the end of the debate, the agents contribute with individually rational opinions, a weaker version of our notion of coherent labelling. Instead, our focus is the design of aggregation functions that satisfy desirable social choice properties, particularly collective rationality (strict rationality in Rago and Toni's terms), without requiring agents' individual rationality. Along this line, notice also that the social choice properties of the QuAD-V algorithm are not investigated.

Finally, notice that unlike the literature on judgement aggregation and preference aggregation, in this paper we will not impose any particular properties on opinions from which our aggregation operators can benefit to guarantee collective rationality. Note that this is the case, for instance, for some aggregation functions in the judgement aggregation literature. For example, among distance-based aggregators, the *Kemeny* rule (Endriss and Moulin 2016) only considers consistent judgement sets, and hence disregards those which are not, and premise-based aggregators (Endriss and Moulin 2016) typically make assumptions on the agenda to guarantee consistency and completeness. In contrast to that literature, some of the aggregation operators introduced in this paper guarantee collective rationality independently of opinions' properties. As discussed above, the rationale for this is clear: we must disregard rationality when humans are involved in debates because their opinions may show contradictions and inconsistencies.

3 Introducing our Discussion Framework

Overall, we consider a situation where several individuals try to reach some consensus on a given issue. We refer to this issue or topic as the discussion *target*. During the discussion process, individuals provide arguments in favour or against this topic (or other arguments) in an orderly manner. Notice that although the example used throughout this paper considers a norm as the topic under discussion, this need not be the case. Indeed, we can imagine any of the dialogues discussed in Parsons et al. (2003), for example, to be discussions about a target which is the subject of the first statement to be made in the dialogue. Putting forward arguments, which may either be directed towards the target, or to arguments that have previously been put forward, is one way in which participants in the discussion can make their points of view known. In addition, participants are able to express their opinion on the target as well as by

indicating which arguments they find acceptable or not. Next Sect. 4 is devoted to formalise this setting, which, as described in Ganzer-Ripoll et al. (2017a), we name it *target-oriented argumentation framework*.

We admit this is clearly a rather restricted notion of a discussion, not least because of the restriction to a single target, and many real-world discussions would not be encompassed by it. Furthermore, as we shall see in Sect. 4, what we formalise is even simpler, because we insist that any statement made after the target is an argument, and this argument has to relate to the target and/or to previous arguments. However, despite this simplicity, a target-oriented discussion allows more complex discussions than any of the existing, implemented, discussion frameworks discussed above. Thus, while we may need to extend the target-oriented discussion framework if we want to capture the full richness of human discussions, what we have here is already a considerable step beyond what currently exists.

Within a target-oriented argumentation framework, we distinguish two relationships between arguments: one argument can be *for* another argument, or it can be *against* another argument. These possible relationships between arguments are those discussed in Besnard and Hunter (2001)[6] Notice that *for* and *against* relationships are binary and directed. Moreover, they are mutually exclusive. In addition, in order to allow participants in a discussion to show their opinion of existing arguments, we make use of the notion of *labels* for arguments. Whereas in standard argumentation, labels are derived from the structure of the set of arguments (Baroni et al. 2011), in our approach a set of labels are assigned by each participant in the discussion. Every label is either in, out or undec. Participants assign an in label to the target or an argument in order to indicate that they accept it. Conversely, they assign out to signal rejection. Finally, an undec label denotes undecision, which may be related to two different situations. Firstly, this label can be used to indicate a participant is doubtful about whether to opt for one of the two options (i.e., in or out). Secondly, uncertainty may also derive from situations where participants simply miss the opportunity of assigning a label (or, in other words, providing their opinion about the target or an argument). Such situations seem to be rather realistic in human debates, as we can hardly expect participants will label absolutely all discussion elements.

Once participants have allocated labels to the target and arguments, we have a number of sets of labels. In order to reach a consensus on whether the target is accepted or rejected, we need to aggregate the sets of labels. This is the main contribution of this paper, investigating how to aggregate all the legitimate and subjective opinions of the participants, expressed as labellings, into a single collective labelling. Once we can aggregate all labellings, then we will be able to assess whether participants as a whole accept, reject, or fail to reach a clear decision about the topic (i.e., the target) under discussion.

In establishing suitable aggregation functions, we have to take into account that we are dealing with human providers of labels, and so cannot expect that the labels are assigned in a rational manner—contradictions or inconsistencies in assigning labellings may occur when expressing opinions. Despite allowing individual labellings

---

[6] The notions of "for" and "against" in Besnard and Hunter (2001) are tied to argument conclusions because this work deeps down into the level of argument structure. However, since we deal at the argument abstract level, we intend to capture the same notions through argument relations.

Table 1 The opinions of the neighbours in the discussion about the street cleaning norm

		Arguments			
		N	a_1	a_2	a_3
Agents	Ag_1	✗	✓	✓	✓
	Ag_2	✓	✗	✗	✓
	Ag_3	✓	✗	✓	✓

to be irrational, we still aim at designing aggregation functions that are able to combine these "imperfect" individual labellings into a "reasonable" agreed opinion. We intuitively characterise "reasonable" by the notion of *coherent labelling* (Ganzer-Ripoll et al. 2017a) and other desirable properties. The next section formally introduces the notion of coherence and subsequent sections study how it is possible to define some aggregation functions that will yield a single aggregated labelling that satisfies several desirable properties which include coherence.

Having introduced the concept of a target-oriented discussion framework, we introduce a simple example that will allow us to illustrate some of the ideas in the paper.

Example 1 (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. for cleaning leaves in the street". Thus, they pose 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. Making explicit both these arguments and their relations 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 Ag_1 in first row in Table 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 (Ag_2 in second row in Table 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 (Ag_3 in third row in Table 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.

Given this situation, the question that arises, and which this paper answers, is: should the neighbours agree to accept this street cleaning norm? or, in other words: how should they aggregate their individual opinions into a consensual one?

4 The Target-Oriented Discussion Framework

The debate between neighbours in the previous section exemplifies the key concepts of our discussion framework. In this framework, a norm N is the target of the debate between multiple agents. Agents can put forward arguments relating to the target or to other arguments and can express their opinions on those arguments together

with the target of the debate. In this section, in addition to introducing formally the key concepts of our framework, we define opinions not presenting inconsistencies as *coherent*. In particular, we introduce the *target-oriented discussion framework* in Sect. 4.1, the agent's labelling representing the agent's opinions in Sect. 4.2, and our *coherence* notion in Sect. 4.3.

4.1 Formalisation of the Target-Oriented Discussion Framework

We aim to define a formal framework capturing both for and against relations between arguments. In this sense our work has some similarities with work in bipolar argumentation frameworks (Amgoud et al. 2008; Cayrol and Lagasquie-Schiex 2005b) and work on argument accrual (Besnard and Hunter 2001; Prakken 2005; Verheij 1995). The motivation behind including arguments for the target and for other arguments is given by novel works concerning humans participating in large-scale argumentation frameworks. (e.g. Klein 2012; Klein and Convertino 2015). These works allow human participants to express both for and against relationships between arguments. Within our framework, we aim to provide that expressiveness.[7] Our desire to capture human uses of argumentation also explains many of the differences between our system and those in the literature—this was explored in more detail in Sect. 2. In what follows we use the term "attack" express the existence of an "against" relationship between two arguments, as is common in the argumentation literature. We also use the term "defence" to express the existence of a "for" relationship between two arguments. We do not use the term "support" for this positive relation between arguments to stress the difference between our work and bipolar argumentation frameworks.[8]

To simplify the formal analysis, we provide some restrictions on the way that a discussion unfolds. In other words, we insist that discussions follow a particular *protocol* with the following steps:

1. One agent puts forward the target of the discussion.
   While any agent is allowed to start a discussion by putting forward a target, only one target is allowed per discussion.
2. Any agent is then allowed to put forward an argument in favour of, or against, the target and/or any arguments that have already been put forward.
   This process continues until no agent has any further arguments to put forward.
3. Agents express their opinions about whether the arguments that have been put forward hold, or whether those arguments do not hold by assigning in, out or undec labels to the arguments.
   Agents are not required to have an opinion about whether every argument holds or not—they are allowed to not express an opinion about any given argument—but

---

[7] Thus while the "for" relations between arguments can be converted into "against" relations (if *a* if for *b* then it is against *c* which is against *a*), this does not appear to be what people naturally do, and since we aim to support natural human argument, we work directly with "for" and "against" relations.

[8] In bipolar argumentation systems, the support relations are mainly used to infer additional "extended" attack relations which are fed into a standard argumentation semantics rather than being weighted up, accrual-style, with attack relations as they are in our work.

Fig. 1 Representation of an
attack relationship $b \mapsto a$ and a
defence relationship $d \Vdash c$
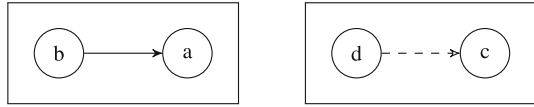
any agent can express an opinion about any argument. Only one opinion can be
expressed about any given argument by any one agent.

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

Before defining the structure that results from the first two steps of this process, we
define a more general structure, the *discussion framework*.

Definition 1 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 $\Vdash \subseteq \mathcal{A} \times \mathcal{A}$ are disjoint attack and defence
relationships (i.e., $\mapsto \cap \Vdash = \emptyset$). We represent that argument $b \in A$ attacks argument
$a \in A$ as $b \mapsto a$, and that b defends a as $b \Vdash a$.

A discussion framework can be also modelled as a graph whose nodes represent the
arguments and whose edges represent either attack or defence relationships between
arguments. Figure 1 shows the graphical representation of attack and defence rela-
tionships.

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

Definition 2 Let $DF = \langle \mathcal{A}, \mapsto, \Vdash \rangle$ be a discussion framework and $a \in A$ one of its
arguments. We say that an argument $b \in \mathcal{A}$ is a *descendant* of a if there is a finite set
of arguments $\{c_1, \ldots, c_r\} \subseteq \mathcal{A}$ 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 our notion of descendant, next we formalise a *target-oriented discussion frame-
work* as having a target argument (e.g., a norm or proposal) as the main focus of the
discussion.

Definition 3 A *target-oriented discussion framework* $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$ is a
discussion framework satisfying the following properties: (i) for every argument $a \in
\mathcal{A}$, a is not a descendant of itself; and (ii) there is an argument $\tau \in \mathcal{A}$, called the *target*,
such that for all $a \in \mathcal{A} \setminus \{\tau\}$, a is a descendant of τ.

Observation 1 *The previous definitions allow us to identify some properties to further
characterise a target-oriented discussion framework:*

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

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

Considering the previous definitions and observation, we can also infer the proposition below.

Proposition 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.*   □

Note the similarity between the graph structure of a $TODF$ and the way that (Proietti 2017) models debates using bipolar argumentation frameworks.

We can formalise the protocol for constructing a $TODF$ as follows:

Definition 4 Given a discussion framework $DF = \langle \mathcal{A}, \mapsto, \Vdash \rangle$, the discussion framework $DF' = \langle \mathcal{A}', \mapsto', \Vdash' \rangle$ is constructed *target-first* if it is constructed according to the following rules:

1. If $\mathcal{A} = \emptyset$ then $\mathcal{A}' = \{\tau\}$ for some argument τ. τ is the target of DF'.
2. If $\mathcal{A} \neq \emptyset$ then $\mathcal{A}' = \mathcal{A} \cup \{b\}$ for $b \notin \mathcal{A}$ and there are subsets of arguments $\mathcal{A}_A, \mathcal{A}_D \subseteq \mathcal{A}$ with $\mathcal{A}_A \cup \mathcal{A}_D \neq \emptyset$ and $\mathcal{A}_A \cap \mathcal{A}_D = \emptyset$, such that $b \mapsto a$ for each $a \in \mathcal{A}_A$ and $b \Vdash c$ for each $c \in \mathcal{A}_D$, and hence $\mapsto' = \mapsto \cup \bigcup_{a \in \mathcal{A}_A} \{(b, a)\}$ and $\Vdash' = \mapsto \cup \bigcup_{c \in \mathcal{A}_D} \{(b, c)\}$.

Note that in rule 2, b must attack or defend at least one argument in $\mathcal{A}$ and can attack or defend multiple arguments, but cannot attack and defend at the same time the very same argument.

The following proposition can be directly derived from Definitions 3 and 4.

Proposition 2 *Any discussion framework DF constructed target-first will be a target-oriented discussion framework.*

It is possible to construct a target-oriented discussion framework in a way that is not target-first, but in so far as we consider the construction of a discussion framework we will only consider target-first construction. Doing so not only ensures that the discussion framework is of a form that is easy to analyse—because it is acyclic—but it also fits with the way, sketched informally above, that existing discussion frameworks are used in practice.

Example 2 (A formalization of the neighbourhood discussion) Figure 2a depicts the neighbours' target-oriented discussion framework. The nodes in the graph represent the set of arguments $\mathcal{A} = \{N, a_1, a_2, a_3\}$ in the example of Sect. 3, where N is the street cleaning norm, and a_1, a_2, a_3 are the rest of arguments. Thus, N, the norm under discussion, is taken to be the target τ in our $TODF$. 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.
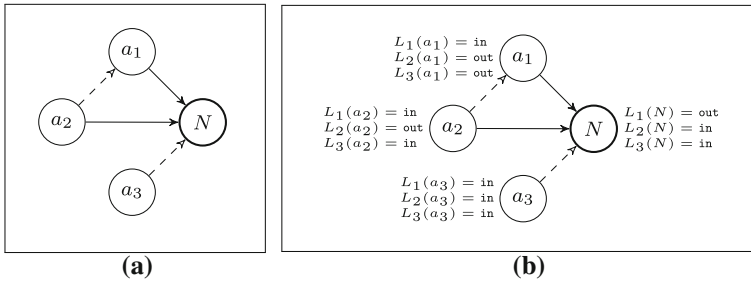
Fig. 2 The example of the neighbours as a $TODF$. **a** associated graph to $TODF$. **b** $TODF$ together with labellings

4.2 Argument Labellings

Once the notion of target-oriented discussion framework has been formalised, in this section we introduce the agent's opinions, what we call *argument labellings*. In terms of the four step protocol given above, this corresponds to step 3. Recall that step 3 involves agents expressing their opinions about the arguments in the discussion framework. Here we consider that each such opinion corresponds to a *labelling* in the sense of Baroni et al. (2011), Caminada (2006) and Caminada and Gabbay (2009). 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 doesn't have an opinion as to whether they are in or out. Besides expressing uncertainty the undec label represents the lack of an opinion. This feature is specially relevant in large-scale debates. As can be seen in Klein (2012), participants usually give their opinion about those arguments of their interest, but we cannot expect them to provide their opinions about all arguments posed within the context of a discussion.

Definition 5 (*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 $L : \mathcal{A} \longrightarrow \{$in, out, undec$\}$ that maps each argument of $\mathcal{A}$ to one of the following labels: in (accepted), out (rejected), or undec (undecidable).

We note as $Ag = \{ag_1, \ldots, ag_n\}$ the set of agents taking part in a $TODF$, and as L_i the labelling encoding the opinion of agent $ag_i \in Ag$. We will put together the opinions of all the agents participating in an argumentation as follows.

Definition 6 (*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 ag_i. A *labelling profile* is a tuple $\mathcal{L} = (L_1, \ldots, L_n)$.

Example 3 (The opinions of the neighbours) Figure 2b graphically depicts Alan's, Bart's, and Cathy's labellings (noted as L_1, L_2, L_3 respectively), representing their opinion about the $TODF$ illustrated by Fig. 2a.

4.3 Coherent Argument Labellings

As pointed out in Awad et al. (2017), there are several ways in which a labelling over an argument structure can be evaluated. In Awad et al. (2017), the authors use the notion of *complete* labelling (Baroni et al. 2011). A complete labelling requires that an argument is labelled `in` iff all the arguments which attack it are labelled `out`; and an argument is labelled `out` iff at least one of the arguments that attack it is labelled `in`. The idea of a complete labelling starts with Dung (1995), and reflects the idea that a rational agent will label arguments consistently—thus an argument can only be accepted (`in`) if all of its attackers are not accepted (`out`) and so on. We believe that the restrictions imposed by complete labelling conditions are not suitable for human participation systems. Instead, we impose less conditions for a labelling to be classified as reasonable or coherent. Hence, given an argument a we contrast the opinions about the argument, named *direct opinion*, with the opinions about its immediate descendants, what we call *indirect opinion*, and look for ways in which these may be made somewhat consistent.

Consider the neighbours' example in Fig. 2b, given argument N, we take into consideration its assigned labels , i.e., its direct opinion $L_1(N)$, $L_2(N)$, and $L_3(N)$; and the labels assigned to its descendants (a_1, a_2, and a_3), i.e., its indirect opinion. Similarly for argument a_1, its direct opinion is formed by the labels assigned to a_1 and its indirect opinion is determined by the labels of its defending argument a_2.

Then, the labelling over an argument will be *coherent* if its indirect opinion agrees with its direct opinion. In other words, when the *majority* of labels in its indirect opinion are in line with its direct label. In the following, the formalization of the notion the coherent labelling is proposed.

First, given an argument a we define its set of attacking arguments $A(a) = \{b \in \mathcal{A} | b \mapsto a\}$; and its set of defending arguments $D(a) = \{c \in \mathcal{A} | c \Vdash a\}$. Hence, the labels attached to the arguments in $A(a) \cup D(a)$ form the indirect opinion of a.

Let L be a labelling and S a set of arguments, we denote the number of arguments accepted in S as $\text{in}_L(S) = |\{b \in S \,|L(b) = \text{in}\}|$ and the number of rejected arguments as $\text{out}_L(S) = |\{b \in S \,|L(b) = \text{out}\}|$. Given this notation, we can consider the number of accepted defending arguments of a as $\text{in}_L(D(a))$ and the number of rejected defending arguments as $\text{out}_L(D(a))$. Similarly, the number of accepted and rejected attacking arguments respectively is represented by $\text{in}_L(A(a))$ and $\text{out}_L(A(a))$, respectively. We define the *positive* and *negative support* of the indirect opinion about an argument below.

Definition 7 (*Positive support*) Let $a \in \mathcal{A}$ be an argument and L a labelling on $\mathcal{A}$. We define the *positive (pro) support* of a as: $Pro_L(a) = \text{in}_L(D(a)) + \text{out}_L(A(a))$. If $Pro_L(a) = |A(a) \cup D(a)|$ we say that a receives *full positive support* from L.

Definition 8 (*Negative support*) Let $a \in \mathcal{A}$ be an argument and L a labelling on $\mathcal{A}$. We define the *negative (con) support* of a as: $Con_L(a) = \text{in}_L(A(a)) + \text{out}_L(D(a))$. If $Con_L(a) = |A(a) \cup D(a)|$ we say that a receives *full negative support* from L.

Observe that the positive support of an argument merges the accepted defending arguments with the rejected attacking ones, whereas the negative support merges the

Table 2 The coherence of the labellings from the neighbourhood 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
Cohe rence	✓			✓			✗		

accepted attacking arguments with the rejected defending ones. Table 2 illustrates the positive and negative support for the arguments involved in the neighbour's example.

We now introduce our notion of coherence by combining the positive and negative support of an argument. We consider that a labelling is *coherent* if for each argument the next conditions are fulfilled: (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, is labelled out, then its negative support has to be higher than positive support

Definition 9 (*Coherence*) Given a $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$, a *coherent labelling* is a total function $L : \mathcal{A} \to \{\text{in}, \text{out}, \text{undec}\}$ such that for all $a \in \mathcal{A}$ with $A(a) \cup D(a) \neq \emptyset$: (1) if $L(a) = \text{in}$ then $Pro_L(a) \geq Con_L(a)$; and (2) if $L(a) = \text{out}$ then $Pro_L(a) \leq Con_L(a)$.

To finish, we define a more general notion of coherence, a stronger one, taking into account what is the difference between the positive and negative support.

Definition 10 (*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 $L : \mathcal{A} \to \{\text{in}, \text{out}, \text{undec}\}$ such that for all $a \in \mathcal{A}$ with $A(a) \cup D(a) \neq \emptyset$: (i) if $L(a) = \text{in}$ then $Pro_L(a) > Con_L(a) + c$; (ii) if $L(a) = \text{out}$ then $Pro_L(a) + c < Con_L(a)$; and (iii) if $L(a) = \text{undec}$ then $|Pro_L(a) - Con_L(a)| \leq c$.

Note that the weakest form of c-coherence, 0-coherence, is exactly the coherence of Definition 9.

Let $TODF$, we will note the class of all the argument labellings of $TODF$ 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 $c \in \mathbb{N}$.

Example 4 Now we apply this definition to the example in Fig. 2b. Table 2 shows that while labellings L_1 and L_2 are coherent, L_3 is not. L_3 is not coherent because the labelling is not coherent for 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)). Just L_1, L_2 belong to the

subclass of its coherent argument labellings $Coh(TODF)$. Moreover, L_1 and L_2 are *0-coherent*.

We have now provided the machinery for agents to express their opinions about the arguments in a target-oriented argumentation framework, and so have all we need to support step 3 of the protocol in Sect. 4.1.

5 The Aggregation Problem

As stated above, our goal is to help agents reach a collective decision on the acceptance or rejection of a target. This corresponds to step 4 of the protocol in Sect. 4.1. In Sect. 5.1 we cast our goal as a *judgement aggregation* (List and Pettit 2002) problem that is solved by having a set of agents collectively decide how to label a target-oriented argumentation framework. We propose to solve such problem using an aggregation function that provides a label for the target and the arguments. Although labellings can be aggregated in different ways, here we follow (Awad et al. 2017) in requiring the outcome of an aggregation must be *fair*. In particular, Sect. 5.2 defines a set of properties to analyse different aggregation functions.

5.1 Collective Labelling

First, we define our notion of discussion problem by putting together a TODF and the individual labellings of the agents involved in a discussion.

Definition 11 (*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 agents, and $TODF$ is a target-oriented discussion framework.

In our example, the labelling discussion problem is $\mathcal{LDP} = \langle \{ag_1, ag_2, ag_3\}, TODF \rangle$. Our goal is to aggregate the individuals' labellings in a $\mathcal{LDP}$ to produce a labelling that represents the collective opinion in the discussion. Again, in our example, that would amount to aggregating L_1, L_2, L_3 into a single labelling.

Definition 12 (*Aggregation function*) Given a labelling discussion problem $\langle Ag, TODF \rangle$, a function $F : \mathcal{D} \longrightarrow \mathbf{L}(TODF)$, where $\mathcal{D} \subseteq \mathbf{L}(TODF)^n$, is called an *aggregation function* for the discussion problem.

In short, an aggregation function F outputs a single labelling from the opinions of the agents contained in a labelling profile. The resulting single labelling encodes the collective decision over the target and the arguments.

Definition 13 (*Decision over a target*) Given an aggregation function F for a labelling discussion problem $\langle Ag, TODF \rangle$ and a labelling profile $\mathcal{L}$, the label $F(\mathcal{L})(\tau)$ stands for the *decision over the target* of the $TODF$.

5.2 Social Choice Properties

Social choice theory provides a collection of formal properties that make it possible to characterise aggregation methods in terms of outcome fairness (Dietrich 2007). Based on Awad et al. (2017), here we formally adapt some of these properties to characterise the desirable properties of an aggregation function in terms of both the arguments in a discussion framework and the collective decision output by the function. Besides these adapted properties, we define some novel properties to characterise aggregation functions with regard to: our coherence notion; and the consideration of dependencies between arguments, recall that our work is the first to relax the limiting assumption of argument independence in the context of collective decisions.

The first two properties characterise aggregation functions in terms of the labellings that they can take as input. In particular, we first adapt from Awad et al. (2017) the notion of exhaustive domain to characterise aggregation functions defined for any labelling profile; and, then, we modify this property to consider if a function is at least defined for coherent labelling profiles.

> **Exhaustive Domain (ED)** (Awad et al. 2017). An aggregation function F satisfies ED if its domain is $\mathcal{D} = \mathbf{L}(TODF)^n$, namely if the function can operate over 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 also define *collective coherence* as a property characterising aggregation functions that produce coherent outcomes.

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

We consider CC as the most important property to satisfy by an aggregation function. Notice that an aggregation function fails at satisfying collective coherence when it is not able to produce a coherent labelling. This is the case when there is a contradiction between the collective label (direct opinion) and the collective indirect opinion for some argument. Such contradiction may pose a threat to the acceptability of collective decisions (Thagard 2002). Notice that collective coherence is the counterpart of the *collective rationality* property defined in Awad et al. (2017). There, Awad et al. require that the outcome of aggregating labellings is a *complete* labelling. As argued in Sect. 4.3, our notion of coherence can be viewed as a relaxation of the notion of completeness. Hence, collective coherence can be regarded as our relaxation of collective rationality.

Within a discussion, the opinions of all the agents involved must be considered equally significant. Anonymity is a social choice property that captures such requirement.

> **Anonymity (A)** (Awad et al. 2017). Let $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile in $\mathcal{D}$, σ a permutation over Ag, and $\mathcal{L}' = (L_{\sigma(1)}, \ldots, L_{\sigma(n)})$ the labelling profile resulting from applying σ 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 rest of the agents. Notice that since non-dictatorship follows directly from the satisfaction of anonymity, the former is a weaker version of the latter.

Non-Dictatorship (ND) (Awad et al. 2017). An aggregation function F satisfies ND if no agent $ag_i \in Ag$ satisfies that $F(\mathcal{L}) = L_i$ for every labelling profile $\mathcal{L} \in \mathcal{D}$.

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, we define two unanimity properties that take into account the relationships between the arguments in the target-oriented discussion framework. In particular, we adapt the notion of unanimity to express a different desirable property: if all agents share the very same direct opinion on an argument, we demand that the collective opinion is in line with such agreed opinion. We name this property *direct unanimity* to reflect that only direct opinions are taken into account. Then, we expand the notion of unanimity to consider the dependencies between the arguments and study the cases when there is unanimity in the indirect opinions. In particular, *endorsed unanimity* is the counterpart of direct unanimity for indirect opinions: if there is an unanimous indirect opinion for an argument, the collective opinion for the argument must be in line with it.

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 $L_i \in \mathcal{L}$, where $l \in \{\text{in}, \text{out}, \text{undec}\}$, then $F(\mathcal{L})(a) = l$ holds.

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

In addition to unanimity, we also consider a complementary property, namely supportiveness. This requires that an aggregation function does not label an argument with a label that has not been employed by any agent.

Supportiveness (S) (Awad et al. 2017). An aggregation function F satisfies S if for every argument $a \in A$ and for all labelling profile $\mathcal{L} = (L_1, \ldots, L_n)$, $\mathcal{L} \in \mathcal{D}$, we can find some agent $ag_i \in Ag$ for which $F(\mathcal{L})(a) = L_i(a)$ holds.

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 labelling, then this collective labelling should remain the same. Here we adapt monotonicity and in-out-monotonicity properties from Awad et al. (2017). Unlike monotonicity, in-out-monotonicity (we prefer the name *binary monotonicity*) only considers the in and out labels.[9]

---

[9] Ensuring the monotonicity of undec is particularly hard.

Monotonicity (M) (Awad et al. 2017). Let $l \in \{\text{in}, \text{out}, \text{undec}\}$ 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 $i, \ldots, i + k$. We say that an aggregation function F is monotonic if $L_j(a) \neq l$ while $L'_j(a) = l$ for all $j \in \{i, \ldots, i + k\}$, then $F(\mathcal{L})(a) = l$ implies that $F(\mathcal{L}')(a) = l$.

Binary Monotonicity (BM) (Awad et al. 2017) Let $l \in \{\text{in}, \text{out}\}$ 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 $i, \ldots, i + k$. We say that an aggregation function F is binary monotonic if $L_j(a) \neq l$ while $L'_j(a) = l$ for all $j \in \{i, \ldots, i + k\}$, then $F(\mathcal{L})(a) = l$ implies that $F(\mathcal{L}')(a) = l$.

We expand the notion of monotonicity with two novel properties that, unlike the notion of monotonicity presented in Awad et al. (2017), consider the opinions of an argument's descendants. The first of these novel properties, which we call *familiar monotonicity*,[10] determines that when the direct support for the collective labelling of an argument increases, the collective labelling 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 labelling might change after the opinions on its descendants are changed. The second property that we propose is the binary version of familiar monotonicity.

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

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

The previous notions of monotonicity are related:

Proposition 3 *If an aggregation function is monotonic (respectively binary monotonic), then it satisfies familiar monotonicity (respectively binary familiar monotonicity).*

Proof The proof is straightforward because the satisfaction of the hypothesis required by familiar monotonicity (resp. binary familiar monotonicity) implies the satisfaction of the hypothesis required by monotonicity (resp. binary monotonicity). □

---

[10] We call it "familiar" because it captures the monotonicity of the descendants—the family—of an argument.

Finally, the notion of independence (Awad et al. 2017) 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. Independence is not a desirable property; but we include it here for completeness.

> **Independence (I)** (Awad et al. 2017). Let be two profiles $\mathcal{L} = (L_1, \ldots, L_n)$ and $\mathcal{L}' = (L'_1, \ldots, ' L_n)$, such that $\mathcal{L}, \mathcal{L}' \in \mathcal{D}$; and $a \in \mathcal{A}$ an argument, such that for all agents $i \in \{1, \ldots, n\}$ $L_i(a) = L'_i(a)$. An aggregation function F satisfies I if $F(\mathcal{L})(a) = F(\mathcal{L}')(a)$.

In this section we have listed a set of properties to characterise aggregation functions. However, it is important to note that not all of them are equal. For a multi-party argumentation-based discussion we believe that Collective Coherence is the most important property. If an aggregation function is collectively coherent, then it extracts a coherent labelling, regardless of the coherency of the individual opinions being aggregated. Along with collective coherence, we also consider that aggregation functions should satisfy the two domain related properties—Exhaustive Domain and Coherent Domain (where we would prefer exhaustive domain to allow wider applicability)— and the usual social choice properties of Anonymity and (if that is not possible) Non-Dictatorship. We also consider that aggregation functions should be monotonic, and, given that we want to capture dependencies, Familiar Montonicity (binary or otherwise) is then desirable. Unanimity is also important, but we consider that this is less important, since we can imagine cases in which unanimity is not satisfied to achieve more important properties such as coherence.[11]

We do not think that aggregation functions should satisfy the remaining properties, namely Monotonicity (binary or otherwise), Supportiveness and Independence.[12] We only include them in order to provide a complete characterisation of aggregation functions.

6 Designing Aggregation Functions to Enact Collective Decision Making

The purpose of this section is to design aggregation functions that calculate the collective labelling for a labelling discussion problem and, thus, the decision over a target. With this aim, notice that in Sect. 1 we observed that independence cannot be considered as a reasonable assumption, and hence our aggregation functions should aim at exploiting dependencies between arguments. At this point, the question is how to exploit dependencies, which fundamentally amounts to deciding how to exploit indirect opinions when computing the aggregated labelling for a given argument. This motivates the design in this section of a family of aggregation functions that exploit

---

[11] For example, when everyone has voted that a contradictory pair of arguments—e.g., that taxes should be cut to improve the economy and that the budget should be balanced in order to improve the economy—should both be `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`. We realise that there are real life groups, such as the current Republican caucus in the US Congress, which would prefer unanimity to collective coherence in such cases.

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

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 they are valuable to the same degree. Besides introducing such functions in Sects. 6.1.2, 6.1.3, and 6.1.4 below, we also investigate the social choice properties that each one satisfies. Thereafter, in Sect. 6.2 we compare the satisfied social choice properties per aggregation function to elucidate the aggregation function that best exploits indirect opinions. Before that, and for the sake of completeness, this section starts, in 6.1.1, by introducing an aggregation function that completely disregards indirect opinions: the so-called majority rule. This will allow us to analyse, as part of our discussion in Sect. 6.2, the benefits and drawbacks, in social choice terms, of exploiting indirect opinions.

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

6.1 Defining Aggregation Functions

6.1.1 Disregarding Dependencies: A Majority Rule

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

Definition 14 (*Majority function*) Given a labelling profile $\mathcal{L}$, the *majority function* for any argument a is defined as:

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

Example 5 (Majority rule in the neighbourhood discussion) Following the neighbours' example, we use the majority function to compute the collective labels of each arguments. See the Fig. 3 that graphically represents the collective labelling obtained. For arguments a_2, a_3 and N there are more in's than out's opinions, 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.

6.1.2 Exploiting Dependencies: Prioritising Direct Opinions

The next function to study, the so-called *opinion first function (OF)*, is a variation of the majority function that exploits dependencies, but prioritising direct opinions

Fig. 3 Neighbourhood discussion: aggregated labellings (and decision over target N) computed by M
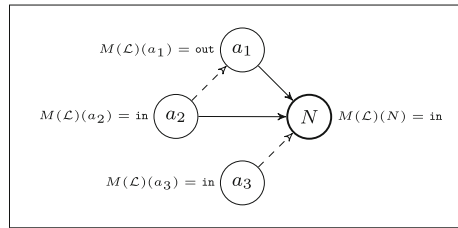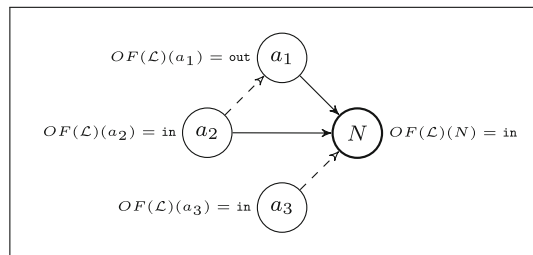

Fig. 4 Neighbourhood discussion: aggregated labellings (and decision over target N) computed by OF

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 indirect opinions to resolve the tie, if possible. Formally,

Definition 15 (*Opinion First Function*) Given a labelling profile $\mathcal{L}$, the *opinion first* function for any argument a is calculated as:

$$OF(\mathcal{L})(a) = \begin{cases} \text{in,} & in_{\mathcal{L}}(a) > out_{\mathcal{L}}(a) \\ \text{in,} & in_{\mathcal{L}}(a) = out_{\mathcal{L}}(a) \text{ and } Pro_{OF(\mathcal{L})}(a) > Con_{OF(\mathcal{L})}(a) \\ \text{out,} & in_{\mathcal{L}}(a) < out_{\mathcal{L}}(a) \\ \text{out,} & in_{\mathcal{L}}(a) = out_{\mathcal{L}}(a) \text{ and } Pro_{OF(\mathcal{L})}(a) < Con_{OF(\mathcal{L})}(a) \\ \text{undec,} & otherwise \end{cases}$$
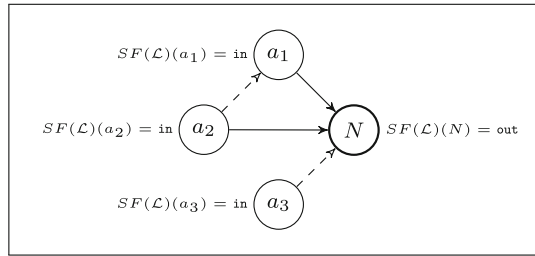
Example 6 Figure 4 shows the collective label produced by OF for each argument in the neighbours' example. Since there are no ties for any argument, OF behaves like M, and so its collective labelling accepts a_2, a_3 and N, and rejects a_1.

6.1.3 Exploiting Dependencies: Prioritising Indirect Opinions

As a counterpart of OF, next we define and study the so-called *Support First function (SF)*, which prioritises indirect opinions over direct opinions. SF considers first indirect opinions to obtain an aggregated opinion on an argument. If using indirect opinions leads to a tie, then SF uses direct opinions to resolve the tie, if possible. Formally,

Definition 16 (*Support First Function*) Given a labelling profile $\mathcal{L}$, the *support first* function for any argument a is calculated as:

Fig. 5 Neighbourhood discussion: aggregated labellings (and decision over target N) computed by SF

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

Example 7 Figure 5 shows the collective label produced by SF for each argument in the neighbours' example. 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, $SF(\mathcal{L})(a_2) = SF(\mathcal{L})(a_3) = \text{in}$. As to argument a_1, SF first considers the collective labelling of a_2, that is in, and thus $SF(\mathcal{L})(a_1) = \text{in}$. Finally, target N is attacked by arguments a_1, a_2, both with collective label in, and defended by argument a_3 with label in. Therefore, the indirect collective support of N is against N, and hence SF rejects it, namely $SF(\mathcal{L})(N) = \text{out}$.

6.1.4 Exploiting Dependencies: Combining Direct and Indirect Opinions

Finally, after studying functions giving priority to either direct opinions, OF, or indirect opinions, SF, in what follows we design an intermediate function balancing both. With this aim, we introduce 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 17 (*Balanced function*) Given a labelling profile $\mathcal{L}$, the *balanced function* over $\mathcal{L}$ for any argument a calculated as:

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

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

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

Fig. 6 Neighbourhood discussion: aggregated labellings (and decision over target N) computed by BF
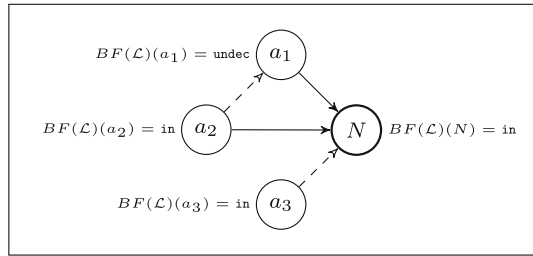
$$DO(\mathcal{L})(a) = \begin{cases} 1, & \mathrm{in}_{\mathcal{L}}(a) > \mathrm{out}_{\mathcal{L}}(a) \\ 0, & \mathrm{in}_{\mathcal{L}}(a) = \mathrm{out}_{\mathcal{L}}(a) \\ -1, & \mathrm{in}_{\mathcal{L}}(a) < \mathrm{out}_{\mathcal{L}}(a) \end{cases}$$

Example 8 (Neighbourhood discussion) Figure 6 shows the aggregated opinion and the decision over the target for our 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., $BF(\mathcal{L})(N) = \mathrm{in}$) and the norm is accepted.

6.2 Comparing Aggregation Functions

This section compares the results obtained by the aggregation functions proposed in the previous sections. The results themselves are contained in "Appendix A" and we refer the reader who wants to understand the formal propositions and proofs of the properties fulfilled by each function to that "Appendix". Here, we just use those results to compare the performance of the functions. We make this comparison with the aid of Table 3, which shows the social choice properties fulfilled by each aggregation function. Table 3 splits social choice properties into two groups: those identified as desirable in Sect. 5.2 when exploiting dependencies between arguments, and those that are not so relevant for our purposes but are typically referred to in the social choice literature. Recall that, as stated in the previous sections, the most important property is *collective coherence* because it ensures the rationality of the outcome of an aggregation function. Table 3 shows us the relation 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 3: 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 3 we draw 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 by means of the undec label.

Table 3 Comparison of social choice properties fulfilled by the aggregation functions that exploit dependencies versus the majority rule. Symbol code: ✓ means fully satisfied; (✓) represents satisfied under some assumptions; and ✗ stands for unsatisfied

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	✓	✗	✗	✗

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, notice that 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.

– *Exploiting indirect opinions preserves anonymity.* Notice that our aggregation functions exploiting indirect opinions (OF, BF, and SF) only consider different volumes 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 a price for exploiting dependencies, namely:

– *The exploitation of indirect opinions impacts the satisfaction of unanimity and monotonicity properties.* Notice that as we move from left to right in Table 3, the unanimity and monotonicity properties become less satisfied, 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 less the number of 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 in this paper) stronger than independence. However, note that we do not regard this observation as a negative result. Recall from our discussion in Sect. 2.3 that (Awad et al. 2017) consider the necessity of independence questionable (because of the existing dependencies between arguments), while the literature considers independence as too strong and not very plausible.

At this point, given the above-mentioned pros and cons regarding the exploitation of dependencies, we are ready to identify what we believe is 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, it fails at satisfying unanimity and monotonicity properties. BF sits between OF and SF.

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 kind of labelling profile.* Implicitly, in our approach 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 undo ties (when the number of acceptances equals the number of rejections) that would occur in the absence of 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 we have introduced, since they all fulfil the exhaustive and coherent domain properties.
– *The introduction of uncertainty negatively affects monotonicity properties.* The use of the undec label may cause the lack of a "positive" or "negative" decision regarding the acceptance of an argument. This fact impacts directly on the satisfaction of the monotonicity properties, and hence the need for weaker versions such as binary monotonicity and binary familiar monotonicity.

Besides the general observations compiled above, Table 3 is also valuable to help us individually analyse each of the aggregation functions introduced in this paper:

– We have shown that M does not satisfy our most important property, collective coherence. Therefore, M does not ensure the coherence of the labelling obtained as a collective decision, and therefore it might contain irrational sets of argument labellings. Despite this fact, the majority function satisfies many of the other desired social choice properties without any restrictions, with the exception of 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. Finally, the non-exploitation of dependencies guarantees the satisfaction of the independence property, but due to the undec label resulting from a tie, it 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 in order to hold. Nonetheless, OF still fails, just like M, to satisfy collective coherence, and hence we cannot ensure the rationality of 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 our 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 with respect to OF. Furthermore, SF also is further from satisfying endorsed unanimity than OF, since SF does not satisfy endorsed unanimity even when we impose some kind of coherence on agents' individual 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 Sect. 5.2, including collective coherence. However, note that BF only satisfies endorsed unanimity in case of 0-coherent labellings. However, notice that SF did 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 as desirable.

7 Computational Analysis

The purpose of this section is twofold. First, given a labelling discussion problem, in Sect. 7.1, we detail an algorithm for computing a collective decision on its target. Thereafter, we empirically analyse the use of that algorithm to solve real-world collective decision problems.

7.1 Computing the Decision Over a Target

Consider a discussion framework, $TODF = \langle \mathcal{A}, \mapsto, \Vdash, \tau \rangle$, with a target τ for which we aim at computing a collective label. Thus, we required 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).

Now, observe that according to Proposition 1, the graph associated to the TODF is a DAG. 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 embed the computation of the collective labels for the arguments and the target of a discussion framework into a topological sorting algorithm. From this follows that the computation of the collective label for the target 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 τ of a discussion framework $TODF$ from a profile $\mathcal{L}$ and aggregation function F. In Ganzer-Ripoll et al. (2017b) we provide a public implementation of Algorithm 1 together with all the functions introduced in Sect. 6.

Algorithm 1 Compute collective decision

1: **function** COMPUTECOLLECTIVEDECISION(G_{TODF}, τ, F, $\mathcal{L}$)
2:    $PendingArguments \leftarrow$ arguments with no descendants (neither attacked nor defended)
3:    **while** $PendingArguments$ is not empty **do**
4:        $b \leftarrow$ remove argument from $PendingArguments$
5:        $F(\mathcal{L})(b) \leftarrow$ 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** $F(\mathcal{L})(\tau)$ ▷ Collective label for target τ

7.2 Empirical Analysis

This subsection empirically analyses the time required by our implementation of Algorithm 1 to compute collective decisions. Our purpose is to investigate whether our approach to collective decision making can be used in practice.

In order to do this, we took as a reference Parlement et Citoyens (2017), which, as mentioned above, is a well established e-government participation site that enables French citizens to participate in the development of laws by making, debating and voting for law proposals. Table 4 provides some details of the 12 consultations that had been completed as of November 2017 (the time of writing). The first column gives the topic of the proposal, and illustrates that these policy consultations are conducted over a wide range of different topics ranging from sustainability to migrants and even modification of the constitution. Each consultation is structured in chapters grouping a number of articles. The number of articles for a proposal is given in the second column. For each individual article, citizens can provide pro and con arguments. The total number of arguments for each proposal is given in the third column. The fourth ("arg/art") column shows the ratio of arguments per article for each of these proposals. Thus, there are 6.2 arguments per article on average. The fifth column shows the number of people who participated per proposal. The sixth column, labelled "arg%" in Table 4, provides the number of arguments as a percentage of the number of participants. The average is 2.84%, meaning that the average participant makes around 3 arguments on each consultation they participate in.[13] Moreover, participants can vote for both the arguments and the article proposals. The last column in Table 4 shows the aggregated

---

[13] Note that there is a high variability in the number or arguments per participant. The standard deviation over the values in Table 4 is 3.08.

Table 4 Completed consultations at Parlement et Citoyens (2017) as of November 2017

Consultation	Articles	Arguments	arg/art	Participants	arg%	Votes
Biodiversity	355	1126	3.17	9334	0.03	51, 516
Restore confidence (a)	240	834	3.48	1109	0.31	10, 523
Restore confidence (b)	292	2102	7.20	773	0.93	9806
On-line consultation	71	435	6.13	659	0.93	5203
Pesticides	97	1106	11.40	499	2.28	4714
Nation protection	56	400	7.14	1636	0.44	8776
Useful prison	77	722	9.38	286	3.28	2687
Local democracy	93	313	3.37	33	10.20	2156
Collaborative economy	69	235	3.41	145	2.35	779
Constitution modification	30	236	7.87	179	4.39	1636
Open data	31	263	8.48	118	7.19	1153
Migrants	25	86	3.44	194	1.77	1399
Average	119.67	654.83	6.20	1247.08	2.84	8362.33

The average number of arguments per participant is 2.84% (in our experiments this number ranges from 0.1 to 50%)

number of votes. Finally, it is also worth mentioning that other participation sites, such as Quoners (2017) or Decide Madrid (2017), are characterised by a similar proportion of arguments to proposals.

Using this real world case scenario as a reference, we artificially generated discussion frameworks where arguments are the nodes of a directed acyclic graph and the edges represent the relationship between the arguments. Given a number of arguments, the graph representing a discussion is a directed acyclic random graph with a probability of 0.5 of creating an edge between any two nodes. Given an edge between two nodes, there is a probability of 0.5 that the edge represents an attack between the arguments, and a probability of 0.5 that it represents support between the arguments. The directions of the relationships between arguments are also randomly determined during the generation of the directed acyclic graph.

Given a discussion framework, we then generated labellings to compose a labelling profile. Each labelling within a profile is built by assigning a random label to each argument in the directed acyclic graph representing the discussion framework. Hence, randomly generated labellings are not guaranteed to be coherent. Despite that, recall that two of our aggregation functions do ensure collective coherence of the resulting decision.

All the computations of collective decisions four our artificially generated discussion frameworks were performed on an Ubuntu 16.04 box with an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz and 4 cores. Furthermore, our experiments set F to be BF, since we consider it to be the best-in-class aggregation function. Computation times with OF, and SF are expected to be similar because they both aggregate direct and indirect opinions just as BF does.
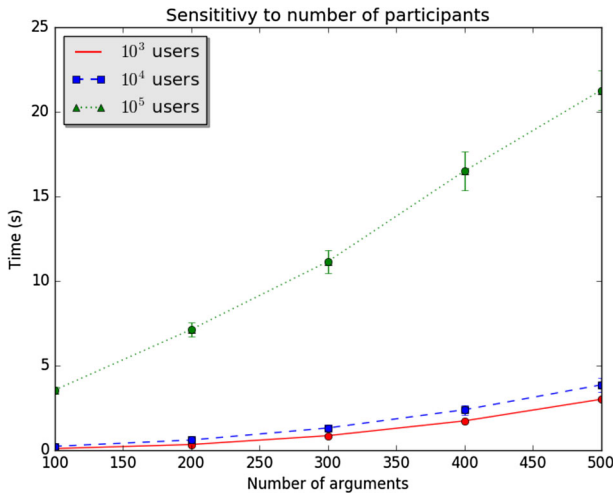
Fig. 7 Time to compute collective decisions as arguments and participants (profile) grow

To establish how the time required to compute a collective decision varies as the number of arguments and participants grows, we looked at discussions with 100–500 arguments, and 10^3–10^5 labellings (representing participants in a discussion). Comparing with Table 4, we can see that these numbers go far beyond the number of arguments and participants involved in debates in actual-world platforms. For each pair of number of arguments and number of participants we generated 100 artificial debates as described above, and computed the collective decision for each debate.

Figure 7 shows the average time (in seconds) required to compute a collective decision as the number of arguments and participants varies. From the figure we learn that we can employ our approach to compute collective decisions in only a few seconds for debates involving as many as 500 arguments and 10^4 participants. Computing collective decisions for 10^5 becomes more demanding, but still all the scenarios can be solved in a a few tens of seconds.

Following the above analysis we investigated the sensitivity of the time required to compute a collective decision to the density of connections between arguments in a discussion framework. For this second study we set the number of participants to 10^3—in the middle of the range we studied before—and varied the number of arguments, between 100 and 500, and the probability for edge creation in the random graph representing the discussion framework, was set to a value from {0.25, 0.5, 0.75}. This allowed us to generate artificial debates with low, medium, and high density of connections between arguments, corresponding to probabilities 0.25, 0.5, and 0.75 respectively. Figure 8 clearly shows that the time to compute a collective decision is affected by density of connections between arguments as the number of arguments grows: the larger the number of arguments, the larger the impact of density of connections between arguments on computational time. Moreover, the larger the density, the more costly to compute a collective decision.
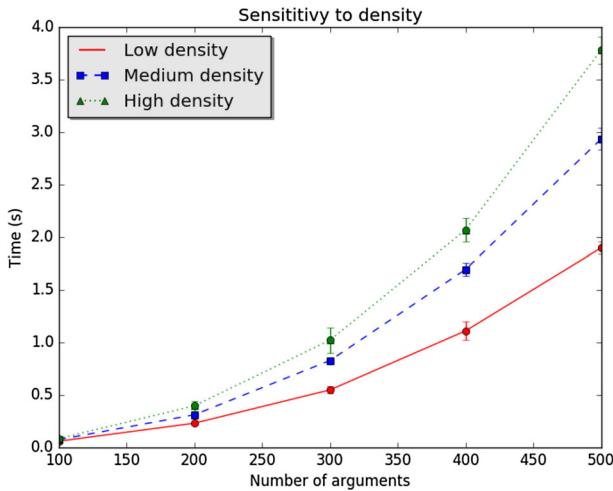
Fig. 8 Time to compute collective decisions as the density of connections between arguments grows

At this point, notice that even the lowest density of connections considered in Fig. 8 goes far beyond the densities that we find in actual-world debates. For instance, when considering a low density scenario (by setting the probability for edge creation to 0.25), each argument receives $\sim 12\%$ attacks and defences (on average) of the total number of arguments. That means that in a debate with 500 arguments, each argument would be related (on average) to other ~ 60 arguments. This already configures a rather dense debate compared to actual-world debates, where humans add new arguments by relating them to a few arguments. In fact, as we have previously mentioned, many participation sites just allow to pose arguments in favour or against a proposal (i.e., specifying a single relationship for each new argument).

To summarise, given that we have evaluated configurations of artificial debates whose scale goes far beyond those of actual-world participation systems, we consider that our approach has the potential to be used in practice, even the largest existing scenarios.

8 Conclusions and Future Work

Within the context of participation and governance domains, collective decision making is taking advantage of the application of technical systems to support on-line debates. In this context, we advance the state of the art by assuming humans may not participate in a rational manner, since they may express their opinions about a topic or an argument in inconsistent or contradictory ways. Furthermore, we also consider uncertainty as an inherent part of any debate, be it because participants signal that they do not hold a clear opinion about certain topics or because they do not express any opinion at all. All things considered, when tasked with computing a collective

decision, we must look to produce an outcome that is *coherent*, namely one that is free of contradictions. Doing that has been the main goal in this paper.

In particular, our approach to solving the above-mentioned collective decision problem makes several contributions. First, we have proposed a mechanism to support debates. More precisely, we have introduced a novel *multi-agent argumentation framework* aimed to articulate a discussion on a given targeted topic or proposal. Thanks to our framework, participants in a debate can express arguments for and against this proposal, indicate the relationships between arguments, and express their opinions about arguments. Furthermore, our framework makes it possible to determine whether a participant's opinion is reasonable (coherent) or not.

Second, we enrich our *multi-agent argumentation framework* with a novel set of aggregation functions that operationalise of the combination individual opinions. This operationalisation results in a consensual decision over the topic under discussion. Since, as argued in Sect. 1, independence cannot be considered as a reasonable assumption when dealing with arguments, we have designed a family of aggregation functions capable of exploiting dependencies between arguments in different ways. We proved that two of those functions guarantee the *coherent collective rationality* of the outcome. And this is the case for *any* sort of labelling profile, namely even those in which participants' opinions are not individually coherent. We also studied several social-choice properties of our aggregation functions, inspired by the work in Awad et al. (2017), where classical properties from social choice are also checked in argumentation settings. Our study produced insights into the design of an aggregation function and the price paid to ensure coherence and handle uncertainty. We showed that either 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 properties that are satisfied. In the end, we observed that the balanced aggregation function, which treats direct and indirect opinions equally, is the one that provided the best trade-off between exploiting direct and indirect opinions. As to uncertainty management, although the introduction of uncertainty favours the general treatment of any kind of labelling profile, it negatively affects monotonicity properties.

Overall, the contributions in this paper break new ground in bringing together the fields of argumentation and computational social choice. We believe that the intersection of these two fields is a sweet spot in which to base the investigation of principled debate-based systems.

In future work, we will particularly focus on three directions. First, we will look at generalising the argumentation framework to allow it to capture more natural debates than the rather simple discussions that are currently captured. Here we see the framework from McBurney and Parsons (2002) as a suitable starting place. Second, we plan to enrich the expressiveness of our target-oriented discussion framework so that opinions about arguments can be expressed not as"accepted", "rejected", and "don't know", as is currently the case, but instead as a number, indicating the degree to which the argument is accepted. This will obviously require the design of another family of aggregation functions. Third, we plan to investigate the type of interfaces required

by humans to participate in target-oriented discussions without being overwhelmed by their complexity. In this line, the works by Gabbriellini and Torroni (2015), Klein (2012), Sklar et al. (2016) appear as promising pointers to the direction we should take.

A Formal Proofs and Results

In this section we detail the results of the aggregation functions, defined in Sect. 6, proving all the properties summarized in Table 3.

First, in Sect. A.1, taking advantage of the similarities between the proposed aggregation functions, we prove some generic results applicable to all of them. Second, in the rest of sections of this appendix, we prove the remaining properties per aggregation function.

A.1 General Results

The next proposition groups the properties equally fulfilled by all aggregation functions.

Proposition 4 *The Majority function M, the Opinion First function OF, the Support First function SF and the Balanced function BF:*
Satisfy the following properties:

  (i) *Exhaustive and Coherent Domain;*
 (ii) *Anonymity and, hence, Non-Dictatorship.*

Do not satisfy the following properties:

 (iii) *Familiar monotonicity and, therefore, Monotonicity;*
 (iv) *Supportiveness.*

Proof of Proposition 4 (i) Exhaustive and Coherent Domain.
   By the definitions in Sect. 6, the functions defined compute an outcome using the cardinal number of different kinds of finite sets whose elements are not restricted by properties of the labellings. Hence, whatever the labelling profiles are (coherent or not), these sets can be computed, and, therefore, the functions can compute an outcome.
  (ii) Anonymity and Non-Dictatorship.
   Consider a TODF, let $Ag = \{ag_1, \ldots, ag_n\}$ be the set of agents involved and $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile. Let σ be a permutation of the set of agents and $\mathcal{L}' = (L_{\sigma(1)}, \ldots, L_{\sigma(n)})$ the labelling profile resulting

from applying the permutation σ over $\mathcal{L}$. Let F be an aggregation function in $\{M, OF, SF, BF\}$.

First of all, let's consider any argument a and see that the following equalities hold: $\text{in}_{\mathcal{L}}(a) = \text{in}_{\mathcal{L}'}(a)$ and $\text{out}_{\mathcal{L}}(a) = \text{out}_{\mathcal{L}'}(a)$. On the one hand $\text{in}_{\mathcal{L}}(a) = |\{ag_i \in Ag \,|L_i(a) = \text{in}\}| = |\{\sigma(ag_i) \in Ag \,|L_{\sigma(i)}(a) = \text{in}\}| = \text{in}_{\mathcal{L}'}(a)$. On the other hand, $\text{out}_{\mathcal{L}}(a) = |\{ag_i \in Ag \,|L_i(a) = \text{out}\}| = |\{\sigma(ag_i) \in Ag \,|L_{\sigma(i)}(a) = \text{out}\}| = \text{out}_{\mathcal{L}'}(a)$. Hence, the direct opinion of an argument is not affected by any permutation of the agents.

Regarding indirect opinions, note that for any argument its indirect opinion depends on the direct and indirect opinion of its descendants. For any argument a without descendants, $Pro_{F(\mathcal{L})}(a) = Pro_{F(\mathcal{L}')}(a) = 0$ and $Con_{F(\mathcal{L})}(a) = Con_{F(\mathcal{L}')}(a) = 0$. For any argument b only having as descendants arguments without descendants, its indirect opinion will be determined by the direct opinion on their descendants, which will not change by the permutation and, hence, $Pro_{F(\mathcal{L})}(b) = Pro_{F(\mathcal{L}')}(b)$ and $Con_{F(\mathcal{L})}(b) = Con_{F(\mathcal{L}')}(b)$. Applying this reasoning recursively, we conclude that the indirect opinion of any argument is not affected by the permutation of the agents.

As $F(\mathcal{L})(a)$ and $F(\mathcal{L}')(a)$ are determined by the indirect ($Con_{F(\mathcal{L})}(a) = Con_{F(\mathcal{L}')}(a)$ and $Pro_{F(\mathcal{L})}(a) = Pro_{F(\mathcal{L}')}(a)$) and direct opinion ($\text{in}_{\mathcal{L}}(a) = \text{in}_{\mathcal{L}'}(a)$ and $\text{out}_{\mathcal{L}}(a) = \text{out}_{\mathcal{L}'}(a)$) of argument a and these are not affected by permutations, we can conclude that $F(\mathcal{L})(a) = F(\mathcal{L}')(a)$.

Hence, it is deduced that any aggregation function based on the direct and indirect opinion to extract an aggregated labelling, such as M, OF, SF or BF, fulfils anonymity, and consequently, non-dictatorship.

(iii) Familiar Monotonicity and Monotonicity.

We prove this by the counter-example displayed in Fig. 9. Consider a $TODF = \langle\{a, b\}, \{(b, a)\}, \emptyset, \{a\}\rangle$ with two labelling profiles $\mathcal{L} = (L_1, L_2)$ and $\mathcal{L}' = (L_1, L_2')$. The opinion of agent 1 is the same in both profiles $L_1(a) = \text{in}$, $L_1(b) = \text{undec}$, whereas the opinion of agent 2 is $L_2(a) = \text{out}$, $L_2(b) = \text{undec}$ for the first profile and $L_2'(a) = \text{undec}$, $L_2'(b) = \text{undec}$ for the second one (see Fig. 9a, b).

Let F stand for any of the aggregation functions $F \in \{M, OF, SF, BF\}$, then the aggregated labellings are $F(\mathcal{L})(a) = \text{undec}$, $F(\mathcal{L})(b) = \text{undec}$ and $F(\mathcal{L}')(a) = \text{in}$, $F(\mathcal{L}')(b) = \text{undec}$, as shown in Fig. 9c and d, respectively. Notice that changing the direct opinion of agent 2 on a from out to undec changes the aggregated labelling, $F(\mathcal{L}')(a) = \text{in} \neq \text{undec} = F(\mathcal{L})(a)$, disproving familiar monotonicity, and therefore monotonicity.

(iv) Supportiveness.

The previous example is also a counter-example for supportiveness by only using $\mathcal{L}$ and its aggregation outcome $F(\mathcal{L})$ (see Fig. 9a, c). Note that $F(\mathcal{L})(a) = \text{undec}$ though no agent's opinion is undec on a.

$\square$

The following lemma, relating coherence to positive and negative support, will help us in proving results regarding the Endorsed Unanimity property for some functions.
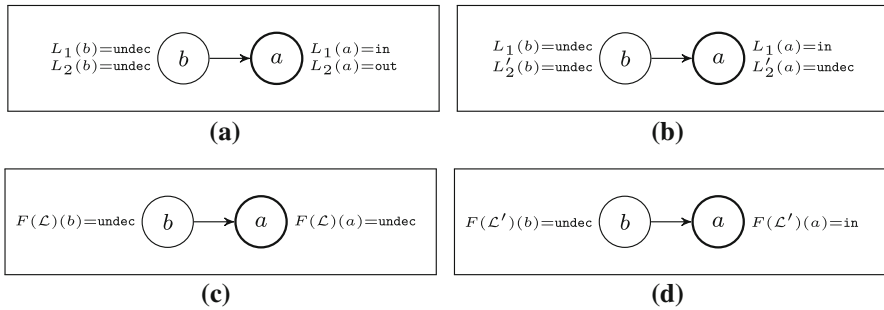
Fig. 9 Counter-example used in proposition 4(iii). **a** Labelling profile $\mathcal{L}$. **b** Labelling profile $\mathcal{L}'$. **c** Aggregated labelling $F(\mathcal{L})$. **d** Aggregated labelling $F(\mathcal{L}')$

Lemma 1 *Let L be a labelling giving full positive (resp. negative) support to an argument $a \in \mathcal{A}$. If L is coherent then $L(a) \neq$ out (resp. $L(a) \neq$ in). Furthermore, if L is at least 0-coherent then $L(a) =$ in (resp. $L(a) =$ out).*

Proof Let us prove this by contradiction. 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 ($m = |A(a) \cup D(a)|$). If L is coherent and $L(a) =$ out, then, by coherence, $Pro_L(a) \leq Con_L(a)$ but $Pro_L(a) = m$ and $Con_L(a) = 0$, and hence we obtain a contradiction, therefore $L(a) \neq$ out. Moreover, suppose that L is 0-coherent, then it cannot be that $L(a) =$ undec because that would mean, by definition of 0-coherence, that $Pro_L(a) = Con_L(a)$, which is not the case. The proof goes analogously for the case with full negative support. $\square$

In the following we analyse the satisfaction of the remaining properties (i.e., unanimity, endorsed unanimity, binary monotonicity, binary familiar monotonicity, independence and collective coherence) for each aggregation function.

A.2 Majority Function

Proposition 5 *The Majority function M:*
Satisfies the following properties:

  (i) *Unanimity;*
 (ii) *Endorsed Unanimity over $Coh_c(TODF)^n$ for $c \geq 0$.*
(iii) *Binary Monotonicity, and consequently, Binary Familiar Monotonicity.*
(iv) *Independence.*

Does not satisfy the following properties:

 (v) *Collective Coherence.*
(vi) *Endorsed Unanimity.*

Proof of Proposition 5 (i) Unanimity.

M trivially satisfies unanimity, since unanimity over an argument is the greatest majority that can be achieved.

Fig. 10 Counter-example used in proposition 5(v). **a** Agent's labelling $\mathcal{L}$. **b** Labelling profile $M(\mathcal{L})$

(ii) Endorsed Unanimity over $Coh_c(TODF)^n$ for $c \geq 0$.
Consider a TODF, and let $Ag = \{ag_1, \ldots, ag_n\}$ be the set of agents involved. Let $\mathcal{L} = (L_1, \ldots, L_n)$ be a labelling profile giving full positive support on argument a, i.e, for every $i \in Ag$, $Pro_{L_i}(a) = m$, with m being the number of descendants of a ($m = |A(a) \cup D(a)|$). By the Lemma 1, if every labelling is at least 0-coherent, we can ensure that for every i, $L_i(a) = \mathtt{in}$. Therefore, there is unanimity on the direct opinion of a, and hence $M(\mathcal{L})(a) = \mathtt{in}$. Analogously, we prove the case with full negative support.

(iii) Binary Monotonicity, and consequently, Binary Familiar Monotonicity.
Let a be an argument and $\mathcal{L}$ and $\mathcal{L}'$ two labelling profiles satisfying the required hypothesis of binary monotonicity setting; i.e, $\mathcal{L} = (L_1, \ldots, L_i, \ldots, L_{i+k}, \cdots, L_n)$, $\mathcal{L}' = (L_1, \ldots, L_i', \ldots, L_{i+k}', \ldots, L_n)$ such that $L_j(a) \neq L_j'(a) = l \in \{\mathtt{in}, \mathtt{out}\}$ for $j \in \{i, \ldots, i+k\}$. Let us assume $l = \mathtt{in}$ and $M(\mathcal{L})(a) = \mathtt{in}$, then we infer that $\mathtt{in}_{\mathcal{L}}(a) > \mathtt{out}_{\mathcal{L}}(a)$. Since $L_j(a) \neq L_j'(a) = \mathtt{in}$ for the agents $j \in \{i, \ldots, i+k\}$, this entails that $\mathtt{in}_{\mathcal{L}'}(a) > \mathtt{in}_{\mathcal{L}}(a) > \mathtt{out}_{\mathcal{L}}(a) > \mathtt{out}_{\mathcal{L}'}(a)$, and hence $M(\mathcal{L}')(a) = \mathtt{in}$. We can prove the case when $l = \mathtt{out}$ and $M(\mathcal{L})(a) = \mathtt{out}$ analogously.

(iv) Independence.
This property states that for any two profiles $\mathcal{L} = (L_1, \ldots, L_n)$ and $\mathcal{L}' = (L_1', \ldots, L_n')$ and an argument a; if for all agents $i \in \{1, \ldots, n\}$ $L_i(a) = L_i'(a)$ then $F(\mathcal{L})(a) = F(\mathcal{L}')(a)$. Note that if $L_i(a) = L_i'(a)$ for all agents $i \in \{1, \ldots, n\}$, then $\mathtt{in}_{\mathcal{L}}(a) = \mathtt{in}_{\mathcal{L}'}(a)$ and $\mathtt{out}_{\mathcal{L}}(a) = \mathtt{out}_{\mathcal{L}'}(a)$, which entails that $M(\mathcal{L})(a) = M(\mathcal{L}')(a)$.

(v) Collective Coherence.
The proof employs a counter-example with the following characteristics, let $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, \{a\} \rangle$ with argument b attacking target a, and with labelling profile $\mathcal{L} = (L)$, where $L(a) = \mathtt{in}$ and $L(b) = \mathtt{in}$, depicted in Fig. 10a. Computing the majority function over this profile we obtain the same labelling $M(\mathcal{L})(a) = \mathtt{in}$ and $M(\mathcal{L})(b) = \mathtt{in}$ (see Fig. 10b) which is not a coherent labelling.

(vi) Endorsed Unanimity.
The previous example provides as well a counter-example for endorsed unanimity. Note that the argument a receives full negative support though its final aggregated label is $\mathtt{in}$.

□

A.3 Opinion First Function

Proposition 6 *The Opinion First function OF:*
Satisfies the following properties:

- (i) *Unanimity;*
- (ii) *Endorsed Unanimity over $Coh(TODF)^n$;*
- (iii) *Binary Monotonicity, and consequently, Binary Familiar Monotonicity.*

Does not satisfy the following properties:

- (iv) *Collective Coherence;*
- (v) *Endorsed Unanimity;*
- (vi) *Independence.*

Proof of Proposition 6 (i) Unanimity.

OF trivially satisfies unanimity.

(ii) Endorsed Unanimity over $Coh(TODF)^n$.

Consider a TODF and let $Ag = \{ag_1, \dots, ag_n\}$ the agents involved. Let $\mathcal{L} = (L_1, \dots, L_n)$ be a labelling profile giving full positive support on argument a, i.e, for every $i \in Ag$, $Pro_{L_i}(a) = m$, with m being the number of descendants of a ($m = |A(a) \cup D(a)|$). By Lemma 1, if every labelling is coherent, then for every agent i, $L_i(a) \in \{\texttt{in}, \texttt{undec}\}$) and $\texttt{out}_{\mathcal{L}}(a) = 0$. We differentiate two cases:

(a) If at least one agent has labelled a with $\texttt{in}$. If so, $OF(\mathcal{L})(a) = \texttt{in}$, which satisfies endorsed unanimity, because $\texttt{in}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}}(a)$.

(b) All agents have labelled argument a as $\texttt{undec}$. 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. Since there is full positive support on a, then for each descendant b_a attacking a, the agents have labelled it with $\texttt{out}$; and (ii) for each descendant b_d defending a, the agents have labelled it with $\texttt{in}$. Therefore, the aggregated opinions for each attaching descendant will be $OF(\mathcal{L})(b_a) = \texttt{out}$ and $OF(\mathcal{L})(b_d) = \texttt{in}$ for each defending descendant. Therefore $Pro_{OF(\mathcal{L})}(a) = m$, and $Con_{OF(\mathcal{L})}(a) = 0$ and $OF(\mathcal{L})(a) = \texttt{in}$ since $Pro_{OF(\mathcal{L})}(a) > Con_{OF(\mathcal{L})}(a)$).

We can analogously prove the case of full negative support.

(iii) Binary Monotonicity, and consequently, Binary Familiar Monotonicity.

Let a be an argument and $\mathcal{L}$ and $\mathcal{L}'$ two labelling profiles satisfying the required hypothesis of binary monotonicity setting; i.e. $\mathcal{L} = (L_1, \dots, L_i, \dots, L_{i+k}, \cdots, L_n)$, $\mathcal{L}' = (L_1, \dots, L_i', \dots, L_{i+k}', \dots, L_n)$ such that $L_j(a) \neq L_j'(a) = l \in \{\texttt{in}, \texttt{out}\}$ for $j \in \{i, \dots, i+k\}$. Let us assume $l = \texttt{in}$ and $OF(\mathcal{L})(a) = \texttt{in}$, then 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 $j \in \{i, \dots, i+k\}$, it entails that $\texttt{in}_{\mathcal{L}'}(a) > \texttt{in}_{\mathcal{L}}(a) \geq \texttt{out}_{\mathcal{L}}(a) > \texttt{out}_{\mathcal{L}'}(a)$ and hence $OF(\mathcal{L}')(a) = \texttt{in}$. We can prove the case when $l = \texttt{out}$ analogously.

(iv) Collective Coherence.

The counter-example of Proposition 5(v) (displayed in Fig. 10a) also serves here
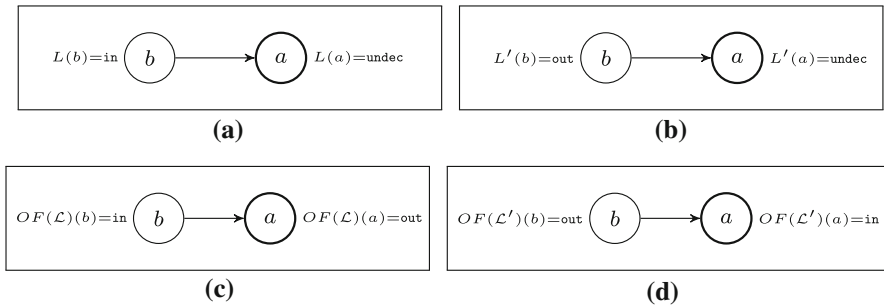
Fig. 11 Counter-example used in proposition 6(vi). **a** Labelling profile $\mathcal{L}$. **b** Labelling profile $\mathcal{L}'$. **c** Aggregated labelling by $OF(\mathcal{L})$. **d** Aggregated labelling $OF(\mathcal{L}')$

as a counter-example for OF; i.e., computing the opinion first function over that profile we obtain the following labelling $OF(\mathcal{L})(a) = \text{in}$ and $OF(\mathcal{L})(b) = \text{in}$ which is not a coherent labelling.

(v) Endorsed Unanimity.

The counter-example of Proposition 5(vi) (see Fig. 10a) also serves as a counter-example for OF; i.e., computing the opinion first function over that profile we obtain that $OF(\mathcal{L})(b) = \text{in}$ though the argument a receives full negative support.

(vi) Independence.

We prove this by counter-example (see Fig. 11). Consider a $TODF = \langle\{a, b\},$ $\{(b, a)\}, \emptyset, \{a\}\rangle$ with labelling profiles $\mathcal{L} = (L)$ and $\mathcal{L} = (L')$ such that $L(a) = \text{undec} = L'(a)$ and $L(b) = \text{in}$, $L'(b) = \text{out}$ (Fig. 11a, b). OF obtains the following labellings $OF(\mathcal{L})(a) = \text{out}$, $OF(\mathcal{L})(b) = \text{in}$ (Fig. 11c); and $OF(\mathcal{L}')(a) = \text{in}$, $OF(\mathcal{L}')(b) = \text{out}$ (Fig. 11d). Clearly, independence is not satisfied because $L(a) = L'(a)$ but $OF(\mathcal{L})(a) \neq OF(\mathcal{L}')(a)$.

□

A.4 Support First Function

Proposition 7 *The Support first function SF:*
Satisfies the following properties:

(i) *Collective Coherence;*
(ii) *Binary Familiar Monotonicity.*

Does not satisfy the following properties:

(iii) *Binary Monotonicity;*
(iv) *Unanimity;*
(v) *Endorsed Unanimity;*
(vi) *Independence.*

Proof (i) Collective Coherence.

Let a be an argument such that $SF(\mathcal{L})(a) = \text{in}$. By Definition 16, $Pro_{SF(\mathcal{L})}(a) >$

$Con_{SF(\mathcal{L})}(a)$ holds, or $Pro_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L})}(a)$ and $in_{\mathcal{L}}(a) > out_{\mathcal{L}}(a)$ holds. It thus 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 $SF(\mathcal{L})(a) = out$.

(ii) Binary Familiar Monotonicity.

Let $\mathcal{L}, \mathcal{L}'$ be two labelling profiles satisfying the required hypothesis of binary familiar monotonicity setting; i.e, let be two labelling profiles $\mathcal{L} = (L_1, \ldots, L_i, \cdots, L_{i+k}, \ldots, L_n)$, $\mathcal{L}' = (L_1, \ldots, L'_i, \ldots, L'_{i+k}, \ldots, L_n)$ such that $F(\mathcal{L})(a) = l \in \{in, out\}$, agents $ag_i, \ldots, ag_{i+k}$ only differing on their labellings of a (namely, for all b descendant of a, $L_j(b) = L'_j(b)$ for every $j \in \{i, \ldots, i+k\}$) and $L(a)_j \neq L'(a)_j = l$ for $j \in \{i, \ldots, i+k\}$. Let us assume $l = in$ and assume that $SF(\mathcal{L}) = in$. 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 $Con_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L}')}(a)$, and the aggregated indirect support of a does not change. We must consider two cases:

(a) If $Pro_{SF(\mathcal{L})}(a) > Con_{SF(\mathcal{L})}(a)$, then $Pro_{SF(\mathcal{L}')}(a) > Con_{SF(\mathcal{L}')}(a)$ and $SF(\mathcal{L}')(a) = in$.

(b) If $Pro_{SF(\mathcal{L})}(a) = Con_{SF(\mathcal{L})}(a)$ (and $Pro_{SF(\mathcal{L}')}(a) = Con_{SF(\mathcal{L}')}(a)$), then $in_{\mathcal{L}}(a) > out_{\mathcal{L}}(a)$. In this case we can conclude that $in_{\mathcal{L}'}(a) \geq in_{\mathcal{L}}(a)$ and $out_{\mathcal{L}}(a) \geq out_{\mathcal{L}'}(a)$, since $l = in$. Therefore, $SF(\mathcal{L}')(a) = in$.

Analogous reasoning can be applied when $l = out$ and $SF(\mathcal{L})(a) = out$, and hence the binary familiar monotonicity for SF holds.

(iii) Binary Monotonicity.

We will employ the counter-example depicted in Fig. 12. Our counter-example considers a $TODF = \langle \{a, b\}, \{(b, a)\}, \emptyset, \{a\} \rangle$ with labelling profiles $\mathcal{L} = (L)$, $L(a) = in = L(b)$ (Fig. 12a), and $\mathcal{L}' = (L')$, $L'(a) = out = L'(b)$ (Fig. 12b). The collective labellings obtained by SF for each profile are $SF(\mathcal{L})(a) = out$, $SF(\mathcal{L})(b) = in$ and $SF(\mathcal{L}')(a) = in$, $SF(\mathcal{L}')(b) = out$ (as shown in Fig. 12c, d).

This counter-example shows that due to the change of the direct opinion about argument a to out, the aggregated labelling for argument a changes to in (i.e., $SF(\mathcal{L}')(a) = in$).

(iv) Direct unanimity.

Using the counter-example of Fig. 12, both labelling profiles (i.e., $\mathcal{L}$ and $\mathcal{L}'$) are unanimous on the labels on a, but the aggregated label for a does not agree with them.

(v) Endorsed Unanimity.

The proof requires an example of an argument with two levels of descendants (see Fig. 13). 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 Fig. 13(a)). The aggregated labelling obtained by SF is $SF(\mathcal{L})(a) = in = SF(\mathcal{L})(c)$ and $SF(\mathcal{L})(b) = out$ (see Fig. 13b). This example shows that although a has full negative support (b is labelled in in $\mathcal{L}$), the aggregated label for a is accepted (labelled with in), which contradicts the support.
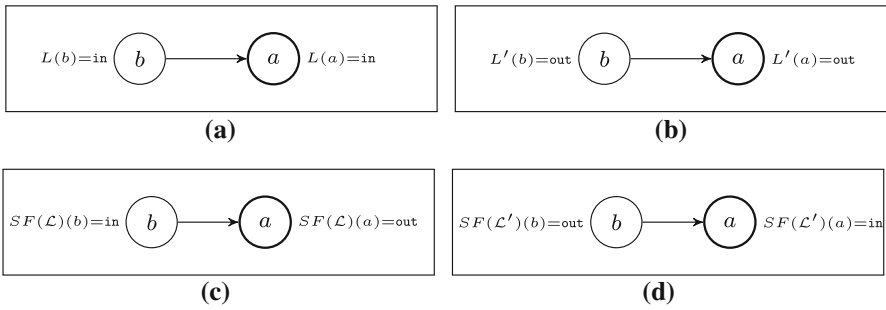
Fig. 12 Counter-example used in proposition 7(iii). **a** Labelling profile $\mathcal{L}$. **b** Labelling profile $\mathcal{L}'$. **c** Aggregated labelling $SF(\mathcal{L})$. **d** Aggregated labelling $SF(\mathcal{L}')$
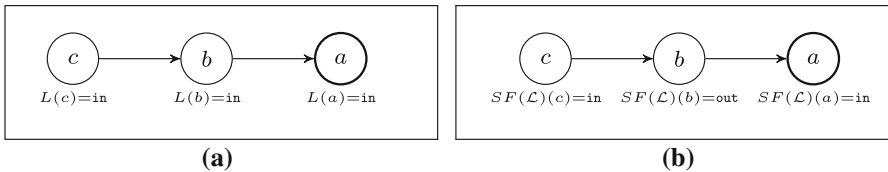

Fig. 13 Counter-example used in proposition 7(v). **a** Labelling Profile $\mathcal{L}$. **b** Aggregated Profile $SF(\mathcal{L})$

(vi) Independence.

The counter-example for proposition 6(vi) (see Fig. 11) also applies to SF. In this case, $L(a) = L'(a) = \texttt{undec}$, $SF(\mathcal{L})(a) = \texttt{out}$ and $SF(\mathcal{L}')(a) = \texttt{in}$.

□

A.5 Balanced Function

Proposition 8 *The Balanced function BF:*
Satisfies the following properties:

(i) *Collective Coherence;*
(ii) *Endorsed Unanimity over $Coh_0(TODF)^n$;*
(iii) *Binary Familiar Monotonicity.*

Does not satisfy the following properties:

(iv) *Binary Monotonicity;*
(v) *Unanimity;*
(vi) *Endorsed Unanimity;*
(vii) *Independence.*

Proof (i) Collective Coherence.

Let a be an argument such that $BF(\mathcal{L})(a) = \texttt{in}$. By Definition 17, $IO(\mathcal{L})(a) + DO(\mathcal{L})(a) > 0$. Thus, there are three cases: (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) $DO(\mathcal{L})(a) = 0$ and $IO(\mathcal{L})(a) = 1$. In any of these three cases, $IO(\mathcal{L})(a) \geq 0$, which implies that $Pro_{BF(\mathcal{L})}(a) \geq Con_{BF(\mathcal{L})}(a)$, and hence BF satisfies collective coherence. The case when $BF(\mathcal{L})(a) = \texttt{out}$ can also be proven analogously.

(ii) Endorsed Unanimity over $Coh_0(TODF)^n$.

Let's 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. By Lemma 1, $L_i(a) = \text{in}$ for every i and hence $DO(\mathcal{L})(a) = 1$. $Pro_{L_i}(a) = m$ implies that for every descendant b of a, $L_i(b) = \text{in}$ if $b \Vdash a$ and $L_i(b) = \text{out}$ if $b \mapsto a$. Therefore, for every descendant $b \in \{A(a) \cup D(a)\}$ its direct support will be $DO(\mathcal{L})(b) = 1$ or $DO(\mathcal{L})(b) = -1$ depending on whether b defends or attacks a, respectively. Moreover, this means that for b defending a $BF(\mathcal{L})(b) \in \{\text{in}, \text{undec}\}$ and for b attacking $BF(\mathcal{L})(b) \in \{\text{out}, \text{undec}\}$, 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 $BF(\mathcal{L})(a) = \text{in}$.

The case supposing full negative support for every agent can be proven analogously.

(iii) Binary Familiar Monotonicity.

Let $\mathcal{L}, \mathcal{L}'$ be two labelling profiles satisfying the required hypothesis of binary familiar monotonicity setting; i.e, let two labelling profiles $\mathcal{L} = (L_1, \ldots, L_i, \cdots, L_{i+k}, \cdots, L_n)$, $\mathcal{L}' = (L_1, \ldots, L_i', \ldots, L_{i+k}', \ldots, L_n)$ such that $F(\mathcal{L})(a) = l \in \{\text{in}, \text{out}\}$, agents $ag_i, \ldots, ag_{i+k}$ only differing on their labellings of a (namely, for all b descendant of a, $L_j(b) = L_j'(b)$ for every $j \in \{i, \ldots, i+k\}$) and $L(a)_j \neq L'(a)_j = l$ for $j \in \{i, \ldots, i+k\}$. Let us assume $l = \text{in}$ and $BF(\mathcal{L})(a) = \text{in}$ and we have that $DO(\mathcal{L})(a) \geq 0$. 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. $\text{in}_{\mathcal{L}}(a) \leq \text{in}_{\mathcal{L}'}(a)$ and $\text{out}_{\mathcal{L}}(a) \geq \text{out}_{\mathcal{L}'}(a)$ and we can deduce that $DO(\mathcal{L}')(a) \geq DO(\mathcal{L})(a) \geq 0$. From this, it follows that $DO(\mathcal{L}')(a) + IO(\mathcal{L})(a) \geq DO(\mathcal{L})(a) + IO(\mathcal{L})(a) \geq 1$, and hence $BF(\mathcal{L}')(a) = \text{in}$. We can analogously check the case where $BF(\mathcal{L})(a) = \text{out}$.

(iv) Binary Monotonicity.

Let consider the following example, depicted in Fig. 14, 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: $L(a) = \text{undec}$, $L(b) = \text{out}$ (see Fig. 14a) and $L'(a) = \text{in}$, $L'(b) = \text{in}$ (see Fig. 14b). The resulting labellings are $BF(\mathcal{L})(a) = \text{in}$, $BF(\mathcal{L})(b) = \text{out}$ (see Fig. 14c) and $BF(\mathcal{L}')(a) = \text{undec}$, $BF(\mathcal{L}')(b) = \text{in}$ (see Fig. 14d). As can be seen, $BF(\mathcal{L})(a) = \text{in}$ and by changing a's label to in on $\mathcal{L}'$ the aggregated labelling of a changes to undec due to the changes in the descendant's labels.

(v) Unanimity.

Fig. 15 represents a counter-example. Let a $TODF$ contain an 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 Fig. 15a):

(1) $L_1(a) = L_1(a_1) = L_1(a_2) = L_1(a_3) = \text{in}$ and $L_1(a_4) = L_1(a_5) = \text{out}$,

(2) $L_2(a) = L_2(a_1) = L_2(a_2) = L_2(a_4) = \text{in}$ and $L_2(a_3) = L_2(a_5) = \text{out}$, and

(3) $L_3(a) = L_3(a_1) = L_3(a_2) = L_3(a_5) = \text{in}$ and $L_3(a_3) = L_3(a_4) = \text{out}$.
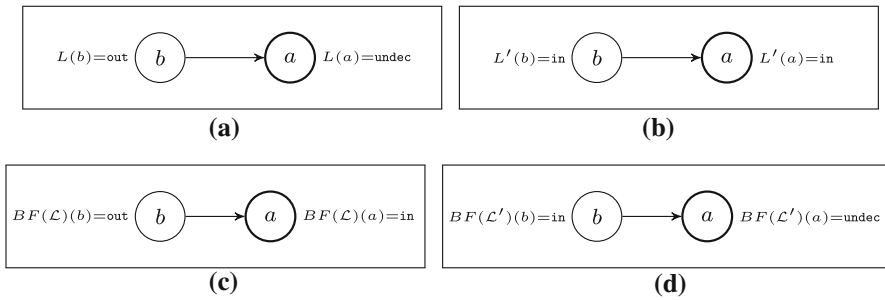
Fig. 14 Counter-example used in proposition 8(iv). **a** Labelling Profile $\mathcal{L}$. **b** Labelling Profile $\mathcal{L}'$. **c** Aggregated Profile $BF(\mathcal{L})$. **d** Aggregated Profile $BF(\mathcal{L}')$
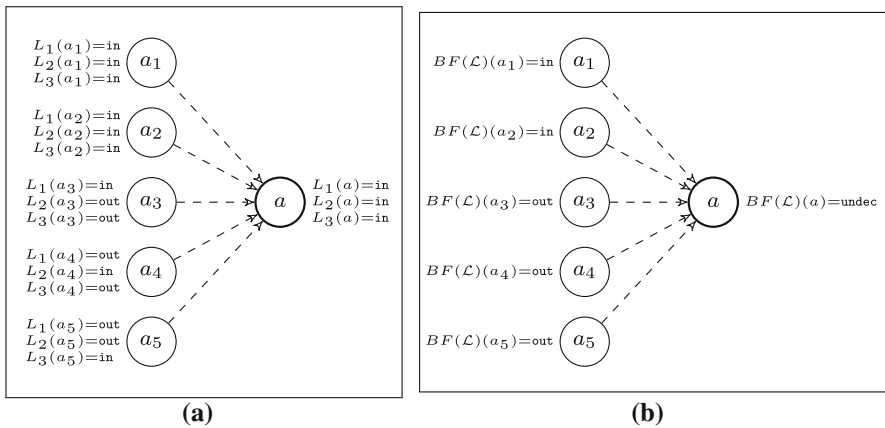

Fig. 15 Counter-example used in Proposition 8(v). On the left are the argument labellings, on the right is the result of the BF function
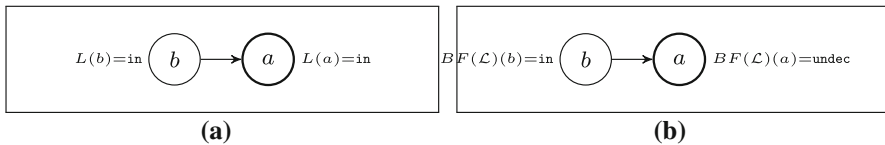

Fig. 16 Counter-example used in proposition 8(vi). **a** Labelling profile $\mathcal{L}$. **b** Aggregated Labelling $BF(\mathcal{L})$

Notice that the three agents agree on accepting the target, and hence there is unanimous opinion on a. Figure 15b 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 direct unanimity.

(vi) Endorsed Unanimity.

The result can be easily seen with the following counter-example, depicted in Fig. 16. Let be a $TODF = \langle\{a, b\}, \{(b, a)\}, \emptyset, \{a\}\rangle$ with the labelling profile $\mathcal{L}$

formed by only one single labelling, $L(a) = \text{in}$ and $L(b) = \text{in}$. The collective labelling obtained applying BF to $\mathcal{L}$ is the following: $BF(\mathcal{L})(a) = \text{undec}$ and $BF(\mathcal{L})(b) = \text{in}$. As we can see, a has full negative support but the collective label is undec instead of out.

(v) Independence.

The counter-example of proposition 6(vi) (see Fig. 11) applies to this function as well. In this case, $L(a) = L'(a) = \text{undec}$, $BF(\mathcal{L})(a) = \text{out}$ and $BF(\mathcal{L}')(a) = \text{in}$

□

References

Amgoud L, Maudet N, Parsons S (2000) Modelling dialogues using argumentation. In: Proceedings of the fourth international conference on multi-agent systems. IEEE Press, Boston, pp 31–38

Amgoud L, Cayrol C, Lagasquie-Schiex M-C, Livet P (2008) On bipolarity in argumentation frameworks. Int J Intell Syst 23(10):1062–1093

Appgree (2017). http://www.appgree.com/. 11 2017 last visited

Arrow KJ, Sen A (2002) Handbook of social choice and welfare, vol 1. Gulf Professional Publishing, Houston

Arrow KJ, Sen A, Suzumura K (2010) Handbook of social choice and welfare, vol 2. Elsevier, Amsterdam

Awad E, Booth R, Tohmé F, Rahwan I (2017) Judgement aggregation in multi-agent argumentation. J Log Comput 27(1):227

Aziz H, Brandt F, Elkind E, Skowron P (2017) Computational social choice: the first ten years and beyond. In: Computer science today, vol 10000

Baoqu (2017). http://beta.baoqu.org/. 11 2017 last visited

Baroni P, Giacomin M (2009) Semantics of abstract argumentation frameworks. In: Rahwan I, Simari GR (eds) Argumentation in artificial intelligence. Springer, Berlin, pp 25–44

Baroni P, Caminada M, Giacomin M (2011) An introduction to argumentation semantics. Knowl Eng Rev 26:365–410

Besnard P, Hunter A (2001) A logic-based theory of deductive arguments. Artif Intell 128:203–235

Better Reykjavík (2017). http://reykjavik.is/en/better-reykjavik-0. 11 2017 last visited

Caminada M (2006) On the issue of reinstatement in argumentation. Logics in artificial intelligence. Springer, Berlin, pp 111–123

Caminada MW, Gabbay DM (2009) A logical account of formal argumentation. Stud Log 93(2–3):109–145

Carr CS (2003) Using computer supported argument visualization to teach legal argumentation. In: Kirschner PA, Buckingham-Shum SJ, Carr CS (eds) Visualizing argumentation: software tools for collaborative and educational sense-making. Springer, London, pp 75–96

Cayrol C, Lagasquie-Schiex M-C (2005a) Gradual valuation for bipolar argumentation frameworks. In: Symbolic and quantitative approaches to reasoning with uncertainty. Springer, pp 366–377

Cayrol C, Lagasquie-Schiex M-C (2005b) On the acceptability of arguments in bipolar argumentation frameworks. In: Symbolic and quantitative approaches to reasoning with uncertainty. Springer, pp 378–389

Consider.it (2017). https://consider.it/. 11 2017 last visited

Coste-Marquis S, Devred C, Konieczny S, Lagasquie-Schiex M-C, Marquis P (2007) On the merging of Dung's argumentation systems. Artif Intell 171(10–15):730–753

Dawes SS (2008) The evolution and continuing challenges of e-governance. Publ Adm Rev 68(s1):S86–S102

Decide Madrid (2014). https://decide.madrid.es/. 11 2017 last visited

Decidim Barcelona (2017). https://www.decidim.barcelona/. 11 2017 last visited

Dietrich F (2007) A generalised model of judgment aggregation. Soc Choice Welf 28(4):529–565

Dietrich F, List C (2007) Strategy-proof judgment aggregation. Econ Philos 23(3):269–300

Dietrich F, Mongin P (2010) The premiss-based approach to judgment aggregation. J Econ Theory 145(2):562–582

Dokow E, Holzman R (2010) Aggregation of non-binary evaluations. Adv Appl Math 45(4):487–504

Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. Artif Intell 77:321–357

Dung PM, Kowalski RA, Toni F (2006) Dialectic proof procedures for assumption-based, admissible argumentation. Artif Intell 170(2):114–159

Endriss U, Moulin H (2016) Judgment aggregation. In: Brandt F, Conitzer V, Endriss U, Lang J, Procaccia AD (eds) Handbook of computational social choice. Cambridge University Press, New York, pp 399–426

Fox J, Barber D, Bardhan KD (1980) Alternatives to Bayes? A quantitative comparison with rule-based diagnostic inference. Methods Inf Med 19:210–215

Fox J, Krause P, Elvang-Gøransson M (1993) Argumentation as a general framework for uncertain reasoning. In: Proceedings of the ninth international conference on uncertainty in artificial intelligence. Morgan Kaufmann, San Matteo, pp 428–434

Gabbriellini S, Torroni P (2015) Microdebates: structuring debates without a structuring tool. AI Commun 29(1):31–51

Gaertner W (2009) A primer in social choice theory, Revised edn. Oxford University Press, Oxford

Ganzer-Ripoll J, López-Sánchez M, Rodriguez-Aguilar JA (2017a) A multi-agent argumentation framework to support collective reasoning. In: Aydoğan R, Baarslag T, Gerding E, Jonker CM, Julian V, Sanchez-Anguix V (eds) Conflict resolution in decision making. Springer International Publishing, Cham, pp 100–117

Ganzer-Ripoll J, Lopez-Sanchez M, Rodriguez-Aguilar JA, Parsons S (2017b) Collective decision making library. https://bitbucket.org/jariiia/argumentation-for-collective-decision-making. Accessed 18 Oct 2018

García AJ, Simari G (2004) Defeasible logic programming: an argumentative approach. Theory Pract Log Program 4(1):95–138

Jackson SK, Kuehn KM (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 J Glob Sustain Inf Soc 14(2):413–427

Kahn AB (1962) Topological sorting of large networks. Commun ACM 5(11):558–562

Klein M (2012) Enabling large-scale deliberation using attention-mediation metrics. Comput Support Coop Work 21(4–5):449–473

Klein M (2017) The deliberatorium: crowd-based innovation. http://cci.mit.edu/klein/deliberatorium.html. 11 2017 last visited

Klein M, Convertino G (2015) A roadmap for open innovation systems. J Soc Media Organ 2(1):1

Krause P, Ambler S, Elvang-Gørannson M, Fox J (1995) A logic of argumentation for reasoning under uncertainty. Comput Intell 11(1):113–131

Lang J, Slavkovik M, Vesic S (2016) Agenda separability in judgment aggregation. In: AAAI, pp 1016–1022

Leite J, Martins J (2011) Social abstract argumentation. In: Proceedings of the twenty-second international joint conference on artificial intelligence, pp 2287–2292

Lin F, Shoham Y (1989) Argument systems: a uniform basis for nonmonotonic reasoning. In: Proceedings of the 1st international conference on knowledge representation and reasoning. Morgan Kaufmann, San Mateo, pp 245–255

List C, Pettit P (2002) Aggregating sets of judgments: an impossibility result. Econ Philos 18(01):89–110

Loomio (2017). https://www.loomio.org/. 11 2017 last visited

McBurney P (2002) Rational interaction. Ph.D. thesis, Department of Computer Science, University of Liverpool

McBurney P, Parsons S (2002) Games that agents play: a formal framework for dialogues between autonomous agents. J Log Lang Inf 11(3):315–334

McBurney P, Parsons S (2009) Dialogue games for agent argumentation. In: Rahwan I, Simari G (eds) Argumentation in artificial intelligence. Springer, Dordrecht, pp 261–280

Modgil S, Caminada M (2009) Proof theories and algorithms for abstract argumentation frameworks. In: Rahwan I, Simari GR (eds) Argumentation in artificial intelligence. Springer, Dordrecht, pp 105–129

Modgil S, Prakken H (2013) A general account of argumentation with preferences. Artif Intell 195:361–397

Mongin P (2008) Factoring out the impossibility of logical aggregation. J Econ Theory 141(1):100–113

Parlement & Citoyens (2017). https://parlement-et-citoyens.fr/. 11 2017 last visited

Parsons S, Wooldridge M, Amgoud L (2003) Properties and complexity of formal inter-agent dialogues. J Log Comput 13(3):347–376

Petitions (2017). UK Government and Parliament. https://petition.parliament.uk/. 11 2017 last visited

Pigozzi G, Slavkovik M, van der Torre L (2008) Independence in judgment aggregation. In: Proceedings of the ninth international meeting of the society for social choice and welfare, SCW'08

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. Bologna, Italy, pp 85–94

Proietti C (2017) The dynamics of group polarization. In: The 6th international conference on logic, rationality and interaction, Hokkaido

Quoners (2017). http://quoners.es/. 11 2017 last visited

Rago A, Toni F (2017) Quantitative argumentation debates with votes for opinion polling. Springer, Cham, pp 369–385

Rahwan I, Simari GR (eds) (2009) Argumentation in artificial intelligence, vol 47. Springer, Dordrecht

Reed C, Rowe G (2004) Araucaria: software for argument analysis, diagramming and representation. Int J Artif Intell Tools 13(04):961–979

Sklar EI, Parsons S, Li Z, Salvit J, Perumal S, Wall H, Mangels J (2016) Evaluation of a trust-modulated argumentation-based interactive decision-making tool. J Auton Agents Multi-Agent Syst 30(1):136–173

Suthers D, Weiner A, Connelly J, 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. Washington, DC, pp 266–273

Sycara K (1990) Persuasive argumentation in negotiation. Theory Decis 28:203–242

Thagard P (2002) Coherence in thought and action. MIT Press, Cambridge

Van Gelder T (2003) Enhancing deliberation through computer supported argument visualization. In: Visualizing argumentation. Springer, pp 97–115

Verheij B (1995) Accrual of arguments in defeasible argumentation. In: Proceedings of the second Dutch/German workshop on nonmonotonic eeasoning. Utrecht, pp 217–224

Vreeswijk GA (1997) Abstract argumentation systems. Artif Intell 90(1):225–279

Walton DN, Krabbe ECW (1995) Commitment in dialogue: basic concepts of interpersonal reasoning. SUNY series in logic and language. State University of New York Press, Albany, NY, USA

Weerakkody V, Reddick CG (2012) Public sector transformation through e-government: experiences from Europe and North America. Routledge, New York

Affiliations

Jordi Ganzer-Ripoll[1] · Natalia Criado[1] · Maite Lopez-Sanchez[2] · Simon Parsons[1] · Juan A. Rodriguez-Aguilar[3]

Jordi Ganzer-Ripoll
jordi.ganzer_ripoll@kcl.ac.uk

Natalia Criado
natalia.criado@kcl.ac.uk

Maite Lopez-Sanchez
maite@maia.ub.es

Simon Parsons
simon.parsons@kcl.ac.uk

[1] Department of Informatics, King's College London, Strand, London WC2R 2LS, UK

[2] University of Barcelona, Barcelona, Spain

[3] Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain