

# Breaking the Cycle

## *Preference-Based Aggregation for Cyclic Argumentation Frameworks*

Michael A. MÜLLER<sup>1</sup>, Blaž ISTENIČ URH, Teodor-Ştefan ZOTESCU  
and Uille ENDRISS

*Institute for Logic, Language and Computation  
University of Amsterdam*

**Abstract.** We consider scenarios where a group of agents wish to simplify a given abstract argumentation framework—specifying a set of arguments and the attacks between them—by eliminating cycles in the attack-relation on the basis of their preferences over arguments. They do so by first aggregating their individual preferences into a collective preference order and then removing any attacks involved in a cycle that go against that order. Our analysis integrates insights from formal argumentation and social choice theory. We obtain sweeping impossibility results for essentially all standard methods of preference aggregation, showing that no Condorcet method and no positional scoring rule can uphold the fundamental principle expressing that views held by every single member of the group must be respected. But we also find that so-called representative-agent rules do offer this guarantee.

**Keywords.** abstract argumentation, social choice theory, preference aggregation, preference-based argumentation

### 1. Introduction

In a debate, in order to settle the issues at stake, the disputants should first determine which arguments ought to be considered and how those arguments relate to one another. This information can be represented by means of an abstract argumentation framework, where arguments are modelled as nodes in a directed graph, with edges corresponding to attacks between arguments [1]. Whether one argument *attacks* another is often uncontroversial and can be agreed upon by all individuals that consider the arguments. But whether one argument *defeats* another will sometimes depend on judgements about the relative strength of the arguments. We can model these judgements as preferences of the involved individuals [2,3,4,5].

In this paper we are interested in the process of aggregating the preferences of several agents before using the collective preference order thus obtained to determine which of the attacks result in defeat. We are specifically interested in using this process to eliminate cycles from the original argumentation framework, given that cycles in the attack-relation often make an unambiguous evaluation of which arguments to accept impossible [6]. We investigate the question of whether there are methods of aggregation that would adequately respect the views of the individual agents involved.

---

<sup>1</sup>Corresponding Author: Michael A. Müller, mam258@cantab.ac.uk.

We do so by adopting the methodology of social choice theory [7,8]. We assume that each individual agent provides us with a strict ranking of all the arguments involved in some cycle. Then we use a preference aggregation rule—which one might think of as a voting rule—to obtain a single such ranking, reflecting the collective preferences of the group as a whole. Natural candidates for aggregation rules include the well-known *Condorcet methods*, such as the Copeland and the Kemeny rule, and the equally well-known *positional scoring rules*, such as the Borda and the Plurality rule [9].

We can use such a ranking of arguments—be it an individual ranking or the collective ranking—to eliminate all attacks from lesser to more preferred arguments within a cycle. The resulting cycle-free graph will have an unambiguous semantics: accept all arguments that are either unattacked or that are defended by accepted arguments. To analyse whether a given preference aggregation rule results in an adequate process of collective cycle elimination, we make use of the *axiomatic method* of social choice theory [7,8]. Specifically, we formulate two very basic axioms, *Acceptance Unanimity* and *Rejection Unanimity*. They encode fundamental normative requirements one should expect to see satisfied by any reasonable method of preference aggregation (and thus by any reasonable method of collective cycle elimination). *Acceptance Unanimity* states that any argument accepted by all individual agents should also be accepted when we use the collective preference order to remove attacks that do not result in defeat. *Rejection Unanimity* encodes the corresponding property for arguments rejected by all individuals.

Surprisingly, we find that essentially all of the well-known preference aggregation rules widely used in practice and commonly studied in the literature violate both of our axioms. Concretely, we obtain impossibility results for all Condorcet methods and all positional scoring rules. On the bright side, we obtain possibility results in case the number of arguments involved in cycles is very small (at most three). More practically relevant, we find that all of the so-called *representative-agent rules* [10] satisfy our requirements.

*Related work.* The state of the art on the integration of formal argumentation and social choice theory, up to about 2016, is reviewed in depth by Bodanza et al. [11]. The problem that so far has received most attention is that of how to adequately aggregate multiple argumentation frameworks, each supplied by a different agent. Axiomatic studies of the normative properties of the aggregation rules one might use to this end include those of Tohmé et al. [12], Dunne et al. [13], Delobelle et al. [14], and Chen and Endriss [15]. Another direction concerns the aggregation of multiple extensions or labellings (i.e., sets of accepted arguments) for the same argumentation framework into a single collective extension. Examples for axiomatic studies of this kind include the works of Caminada and Pigozzi [16], Rahwan and Tohmé [17], Awad et al. [18], and Chen and Endriss [19]. Some of these also include impossibility results, and this line of work is the one most closely related to ours. In some sense, we focus on a restricted form of the problem of aggregating alternative extensions. Namely, the way in which we generate alternative extensions for each agent based on their own preference relation heavily constrains the range of extension-profiles we might encounter. The fact that we still obtain sweeping impossibility results, and that we do so on the basis of very weak axioms, is all the more surprising. Airiau et al. [20] analyse the variety of profiles one might encounter in a setting very similar to ours. Finally, we note that a particularly convincing motivation for the need to aggregate the views of multiple agents in the context of argumentation comes from online deliberation platforms. This point was first made by Leite and Martins [21], and Bernreiter et al. [22] further expand on this idea.

*Paper overview.* We develop our formal model of collective cycle elimination in Section 2, starting from relevant preliminaries in formal argumentation and social choice theory, leading up to the formulation of our two axioms. We then present both our impossibility and our possibility results in Section 3, before concluding with a brief discussion of possible alternative approaches to modelling collective cycle elimination.

## 2. Modelling Collective Cycle Elimination in Abstract Argumentation Frameworks

In this section, after recalling some relevant basics of formal argumentation theory, we introduce our model of collective cycle elimination for abstract argumentation frameworks. This includes a review of well-known preference aggregation rules and the formulation of basic normative requirements, known as *axioms* in social choice theory, that any reasonable mechanism for collective cycle elimination ought to satisfy.

### 2.1. Formal Argumentation Theory

We are going to build on the familiar model of abstract argumentation going back to the seminal work of Dung [1], abstracting away from the internal structure of arguments and instead focusing on the relationships that hold between them. An *argumentation framework* is a pair  $\langle \mathcal{A}, \rightarrow \rangle$ , where  $\mathcal{A}$  is a finite set of *arguments* and  $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$  is an *attack-relation* specifying for any two arguments whether the first attacks the second. We say that argument  $a$  ‘attacks’ argument  $b$  in case  $(a, b) \in \rightarrow$ , and we write  $a \rightarrow b$  for better readability. Argumentation frameworks can be visualised as directed graphs.

**Example 1.** The argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  with arguments  $\mathcal{A} = \{a, b, c, d\}$  and attack-relation  $\rightarrow = \{(a, b), (b, a), (a, c), (b, c), (c, d)\}$  is shown in Figure 1.

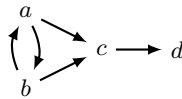


Figure 1. Example of an argumentation framework.

The argumentation framework of Figure 1 includes a cycle: arguments  $a$  and  $b$  attack each other. In general, for a given argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$ , a *cycle* is a set  $\{a_1, \dots, a_k\} \subseteq \mathcal{A}$  such that  $a_j \rightarrow a_{j+1}$  for all  $j < k$  as well as  $a_k \rightarrow a_1$ . We denote the union of all cycles in  $\langle \mathcal{A}, \rightarrow \rangle$  as  $Cyc(\mathcal{A}, \rightarrow)$ . Observing a cycle is not a problem in and of itself. Indeed, it is perfectly possible to develop rational arguments that are in conflict with one another. But if our goal is to define some notion of ‘winning’ argument(s) for a given argumentation framework, this type of conflict generates difficulties. For how do we decide which of  $a$  and  $b$  should ‘win’ if we have abstracted away from the internal structure of the arguments?

A number of different proposals have been made in the literature for how to extract the ‘winning’ arguments from a given argumentation framework [1,6]. One such proposal is to select the arguments that belong to its *grounded extension*.

**Definition 1** (Grounded extension). *The grounded extension of an argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  is the least set  $\Delta \subseteq \mathcal{A}$  that satisfies the following three properties:*

1. If  $a, b \in \Delta$ , then neither  $a \rightarrow b$  nor  $b \rightarrow a$ .
2. If  $a \in \Delta$  and  $b \rightarrow a$ , then there exists an argument  $c \in \Delta$  such that  $c \rightarrow b$ .
3. If for every  $b$  such that  $b \rightarrow a$  there exists a  $c \in \Delta$  such that  $c \rightarrow b$ , then  $a \in \Delta$ .

The first condition expresses that no two arguments in  $\Delta$  attack one another. If we think of  $c \in \Delta$  with  $c \rightarrow b$  as ‘defending’  $a \in \Delta$  with  $b \rightarrow a$ , then we can interpret the other two conditions as saying that  $\Delta$  coincides with the set of arguments defended by  $\Delta$ .

When an argument is part of the grounded extension, we say that it is *accepted* or *winning*. It turns out that the grounded extension of the framework of Figure 1 is the empty set.<sup>2</sup>

The grounded extension and other extensions, such as the *preferred extensions* or the *stable extensions*, are often referred to as *semantics* for argumentation frameworks [1,6]. Many of the challenges of defining an adequate semantics for argumentation frameworks can be traced back to the presence of cycles. Indeed, when there are no cycles, defining a natural notion of acceptance is straightforward: accept all unattacked arguments as well as all arguments defended by other accepted arguments. Most proposals for a semantics to be found in the literature, including the grounded extension, reduce to this simple definition for the special case of cycle-free argumentation frameworks. Already in his original paper on the topic, Dung showed that the different semantics he proposed all coincide in this case [1, Theorem 30].

## 2.2. Individual and Collective Cycle Elimination

When faced with a conflicting set of arguments, i.e., an argumentation framework with cycles, an individual might wish to eliminate some of the attacks based on her preferences over some of the arguments involved. There are several proposals in the literature for how to model such a process [2,3,4]. For instance, Bench-Capon [3] suggests that each argument speaks to a (moral or social) value, and individuals have preferences over such values. We note that Modgil [4] specifically considers preferences in the context of 2-cycles, i.e., cases where two arguments attack one another. One advantage of this setting is that it avoids the possible counter-intuitive consequence of potentially having to accept arguments that were in conflict before attacks were removed [5].

Here we employ a particularly simple strategy for eliminating cycles. For any given framework  $\langle \mathcal{A}, \rightarrow \rangle$ , we assume that every individual is equipped with a preference relation  $>$ , a strict partial order on  $\mathcal{A}$ , that (at the very least) strictly ranks all arguments in  $Cyc(\mathcal{A}, \rightarrow)$ , i.e., all those involved in a cycle. We then remove all attacks involved in a cycle that go against this ranking. Formally, given an argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  and a strict linear order  $>$  on  $Cyc(\mathcal{A}, \rightarrow)$ , we define a *defeat-relation*  $\Rightarrow$  on  $\mathcal{A}$  as follows:

$$a \Rightarrow b \text{ if and only if } a \rightarrow b \text{ and } a \text{ and } b \text{ are not part of the same cycle or } a \rightarrow b \text{ and } a > b$$

As we only use the part of  $>$  that ranks arguments in  $Cyc(\mathcal{A}, \rightarrow)$ , from here on we shall assume that any given  $>$  is simply a strict linear order on  $Cyc(\mathcal{A}, \rightarrow)$ . We use  $top(>)$  to refer to the top element of  $>$ . We sometimes write  $\rightarrow_{>}$  instead of  $\Rightarrow$ , to emphasise

<sup>2</sup>To see this, note that both  $a$  and  $b$  are undefended from each other’s attacks;  $c$  is defended by  $a$  from  $b$ ’s attack, and by  $b$  from  $a$ ’s attack, but  $a$  and  $b$  are not in the grounded extension; and although  $d$  is defended from the attack from  $c$  by  $a$  and  $b$ , again, these are not in the grounded extension.

that  $\succ$  was used to eliminate cycles. Observe that  $\langle \mathcal{A}, \Rightarrow \rangle$  is itself an argumentation framework—and, by construction, it has no cycles. At this point, an example is in order.

**Example 2.** Consider once more the argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  of Figure 1. For the preference relation  $\succ$  on  $\text{Cyc}(\mathcal{A}, \rightarrow) = \{a, b\}$  with  $a \succ b$ , we obtain  $\langle \mathcal{A}, \Rightarrow \rangle$  with  $\Rightarrow = \{(a, b), (a, c), (b, c), (c, d)\}$ , i.e., the attack from  $b$  to  $a$  is removed. While, as we previously saw, the grounded extension of  $\langle \mathcal{A}, \rightarrow \rangle$  is empty, that of  $\langle \mathcal{A}, \Rightarrow \rangle$  is  $\{a, d\}$ .

We want to choose the preference relation  $\succ$  used to eliminate cycles in a principled manner. Suppose the participants to the argument have settled on an argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$ . We can think of this as the debaters having settled on what arguments have been put forth and how they relate to each other. Then, we ask them to each submit a ranking over the arguments that occur in cycles before we aggregate these individual preferences into a single collective preference relation.

So let  $N = \{1, \dots, n\}$  be a set of *agents*. For the argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  under consideration, suppose each agent  $i \in N$  reports a strict linear order  $\succ_i$  on  $\text{Cyc}(\mathcal{A}, \rightarrow)$ , giving rise of a *profile*  $\succ = (\succ_1, \dots, \succ_n)$ . We shall use  $m$  to refer to the cardinality of  $\text{Cyc}(\mathcal{A}, \rightarrow)$ , so each of the preference relations involved is a strict ranking of the same  $m$  objects. A *preference aggregation rule* is a function  $F$  mapping any given profile of individual preferences to a single strict linear order  $\succ$  on  $\text{Cyc}(\mathcal{A}, \rightarrow)$ .<sup>3</sup> We are now ready to put all the pieces together and formulate our core definition.

**Definition 2** (Collective cycle elimination). *Given an argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$ , a profile  $\succ = (\succ_1, \dots, \succ_n)$  of individual preferences on  $\text{Cyc}(\mathcal{A}, \rightarrow)$ , and a preference aggregation rule  $F$  for  $m = |\text{Cyc}(\mathcal{A}, \rightarrow)|$  arguments, the result of collective cycle elimination is the cycle-free argumentation framework  $\langle \mathcal{A}, \Rightarrow \rangle$  we obtain by first using  $F$  to aggregate the profile into a single preference relation  $\succ = F(\succ_1, \dots, \succ_n)$  and then computing  $\Rightarrow$  as the defeat-relation for  $\langle \mathcal{A}, \rightarrow \rangle$  and  $\succ$ .*

In view of our notational conventions, note that we can write  $\rightarrow_{F(\succ)}$  instead of  $\Rightarrow$  if we need to explicitly refer to both the profile and the preference aggregation rule used.

### 2.3. Preference Aggregation Rules

We so far left open the question of how exactly preferences are to be aggregated. A rich variety of aggregation rules are discussed in the literature of social choice theory [7,8]. But almost all of the rules used in practice belong to one of two families, the *Condorcet methods* and the *positional scoring rules* [9]. Let us now define these families for our specific context of aggregating  $n$  strict rankings of  $m$  arguments into one such ranking.

For a given profile  $\succ = (\succ_1, \dots, \succ_n)$ , a *Condorcet winner* is an argument  $c \in \text{Cyc}(\mathcal{A}, \rightarrow)$  that beats any other argument in direct majority contests:

$$|\{i \in N \mid c \succ_i a\}| > |\{i \in N \mid a \succ_i c\}| \text{ for all arguments } a \in \text{Cyc}(\mathcal{A}, \rightarrow) \setminus \{c\}.$$

Then a *Condorcet method*  $F$  is any preference aggregation rule that ensures that, whenever the profile  $\succ$  to be aggregated has a Condorcet winner, that Condorcet winner ends

<sup>3</sup>The technical term commonly used in social choice theory for what we call ‘preference aggregation rule’ here is *social welfare function*, with  $\text{Cyc}(\mathcal{A}, \rightarrow)$  being the set of alternatives [7].

up as the top element of the collective preference relation  $F(>)$ . Well-known examples for rules that are Condorcet methods include the *Copeland rule* and the *Kemeny rule*.

In contrast, *positional scoring rules* are defined using *scoring vectors*. A *scoring vector* of length  $m$  is a vector  $\mathbf{s} = (s_1, \dots, s_m) \in \mathbb{R}^m$  of real numbers with  $s_1 \geq s_2 \geq \dots \geq s_m$  and  $s_1 > s_m$ . Any such scoring vector induces a positional scoring rule (PSR) as follows. For every agent  $i$  ranking a given argument  $a$  in position  $\ell$ , we award  $s_\ell$  points to  $a$ , and we finally rank the alternatives in terms of the total points awarded. Ties are broken lexicographically, either by means of some fixed order on agents or by means of some fixed order on arguments. The well-known *Borda rule* is the PSR with  $\mathbf{s} = (m - 1, m - 2, \dots, 0)$ . Other important examples include the *Plurality rule* with  $\mathbf{s} = (1, 0, \dots, 0)$ , which amounts to each agent giving one point to their most preferred argument only, and the *Veto rule* with  $\mathbf{s} = (0, \dots, 0, -1)$ , under which ever agent takes away one point from her least preferred argument.

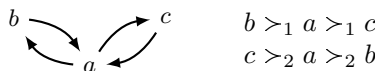
A useful fact regarding PSRs is that the aggregation rule being induced does not change under affine transformations of the scoring vector. In particular, for any given scoring vector  $\mathbf{s} = (s_1, \dots, s_m)$ , we can first subtract  $s_m$  from each score and then divide all of them by  $s_1 - s_m$ . Thus, w.l.o.g., we can think of any PSR for  $m$  arguments as being induced by a scoring vector  $\mathbf{s} = (s_1, \dots, s_m)$  with  $s_1 = 1$  and  $s_m = 0$ .

We are also going to consider a third family of preference aggregation rules, the so-called *representative-voter rules* [10], which are far less common in practice but still of theoretical interest. In our context, we shall refer to them as *representative-agent rules*. A representative-agent rule is any preference aggregation rule  $F$  that guarantees  $F(>) \in >$  for all profiles  $>$ . That is, these are rules that always return a ranking reported by one of the agents. Examples include the *modal ranking rule* [23], which returns one of the rankings reported most frequently, and the *average-voter rule* [10], which returns one of the rankings minimising the average Kendall-tau distance to the rankings in the profile. Importantly though, the representative-agent rules also include some very unappealing rules, notably the *dictatorships*, which always return the ranking of the same agent.

### 2.4. Normative Requirements for Collective Cycle Elimination

So which aggregation rule should we use? The standard approach in social choice theory to answer such questions is the *axiomatic method* [7,8]. The idea is to identify relevant normative requirements, so-called *axioms*, and to systematically investigate which rules satisfy which (combinations of) axioms. Here we shall propose two very weak such axioms that one would hope any reasonable mechanism for collective cycle elimination, and thus any preference aggregation rule defining such a mechanism, would satisfy.

**Example 3.** Consider the scenario sketched below, with an argumentation framework for three arguments and two agents with opposite preferences regarding these arguments.



Here, the two agents differ in how they eliminate cycles, but they both accept  $b$  and  $c$ , and they both reject  $a$ . It is desirable that the same arguments are collectively accepted.

We capture this intuition using the following two axioms.

**Axiom 1** (Acceptance Unanimity). *A preference aggregation rule  $F$  for  $m$  arguments and  $n$  agents respects the axiom of Acceptance Unanimity if, for every argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  with  $|\text{Cyc}(\mathcal{A}, \rightarrow)| = m$  and every profile  $\succ = (\succ_1, \dots, \succ_n)$ , it is the case that any argument that belongs to the grounded extension of  $\langle \mathcal{A}, \rightarrow_{\succ_i} \rangle$  for every agent  $i \in N$  also belongs to the grounded extension of  $\langle \mathcal{A}, \rightarrow_{F(\succ)} \rangle$ .*

**Axiom 2** (Rejection Unanimity). *A preference aggregation rule  $F$  for  $m$  arguments and  $n$  agents respects the axiom of Rejection Unanimity if, for every argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  with  $|\text{Cyc}(\mathcal{A}, \rightarrow)| = m$  and every profile  $\succ = (\succ_1, \dots, \succ_n)$ , it is the case that any argument not belonging to the grounded extension of  $\langle \mathcal{A}, \rightarrow_{\succ_i} \rangle$  for any agent  $i \in N$  also does not belong to the grounded extension of  $\langle \mathcal{A}, \rightarrow_{F(\succ)} \rangle$ .*

Thus, Acceptance Unanimity expresses the fundamental idea that, if argument  $a$  were to be accepted by every single agent once they have eliminated cycles according to their individual preferences,  $a$  should also be accepted if we aggregate the individual preferences before eliminating cycles. Rejection Unanimity similarly stipulates that unilaterally rejected arguments should also be collectively rejected.

Unanimity is one of the most common and most basic axioms considered in social choice theory; it is closely related to the notion of *Pareto Efficiency* [7]. Unanimity with respect to something *not* being the case, as for our axiom of Rejection Unanimity, has also been called *groundedness* in the literature [15]. In the context of argumentation theory, Dunne et al. [13] have studied a related notion under the name of (attack) *closure*.

### 3. Impossibility and Possibility Results

Intuitively speaking, our Unanimity axioms are both very desirable and very undemanding, so one would hope and expect them to be satisfied by almost any reasonable aggregation rule. The technical results we present in this section show that, unfortunately, this intuition is not correct: it turns out to be impossible to satisfy our axioms for most rules of practical interest. We prove sweeping impossibility results for Condorcet methods and PSRs in case there are at least 4 arguments that are involved in cycles. But we also obtain possibility results for argumentation frameworks with fewer cycle arguments. Finally, we present a general possibility result showing that all representative-agent rules can guarantee both Unanimity axioms for all argumentation frameworks.

#### 3.1. Groundwork

The following technical lemma spells out sufficient conditions for an aggregation rule  $F$  to fail both Acceptance Unanimity and Rejection Unanimity. Observe that for any profile  $\succ$  with  $F(\succ) \notin \succ$ , it must be the case that for every agent  $i \in N$  there exists a *distinguishing pair* of arguments  $(x_i, y_i)$  such that  $x_i \succ_i y_i$  but  $F(\succ)$  ranks  $y_i$  over  $x_i$ .

**Lemma 1.** *A preference aggregation rule  $F$  for  $m \geq 4$  arguments and  $n \geq 3$  agents fails both Acceptance Unanimity and Rejection Unanimity if there exists a profile  $\succ = (\succ_1, \dots, \succ_n)$  for which at least one of the following two conditions holds:*

- (i)  $F(\succ) \notin \succ$  and for all distinct agents  $i, j \in N$  there exist distinguishing pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  such that either  $\{x_i, y_i\} \cap \{x_j, y_j\} = \emptyset$  or  $y_i = y_j$ .



(ii)  $top(F(\succ)) \neq top(\succ_i)$  for all agents  $i \in N$ .

*Proof.* First, observe that (ii) implies (i). Indeed, for any profile  $\succ$  satisfying (ii), by choosing distinguishing pairs  $(x_i, y_i)$  with  $x_i = top(\succ_i)$  and  $y_i = top(F(\succ))$ , we find (i) to be satisfied as well. So we are left with proving that (i) is a sufficient condition.

So consider an arbitrary profile  $\succ$  for  $m \geq 4$  arguments that satisfies condition (i), as well as an aggregation rule  $F$  for such profiles. We are going to construct an argumentation framework  $\langle \mathcal{A}, \rightarrow \rangle$  such that  $F$  fails both axioms when applied to  $\langle \mathcal{A}, \rightarrow \rangle$  and  $\succ$ . Let  $\mathcal{A} = \{x_i, y_i \mid i \in N\} \cup \{b, c\} \cup \{a_1, \dots, a_k\}$  and let  $\rightarrow = \{(x_i, y_i), (y_i, x_i), (x_i, b) \mid i \in N\} \cup \{(b, c)\} \cup \{(a_\ell, a_{\ell+1}) \mid \ell < k\} \cup \{(a_k, a_1)\}$ , where  $k$  is chosen such that there are exactly  $m$  arguments in cycles (see Figure 2). Intuitively, we let each distinguishing pair be involved in a mutual attack and connect them suitably to arguments  $b$  and  $c$ . Arguments  $a_1, \dots, a_k$  only make sure enough arguments occur in cycles. Note that this construction is in general possible for  $n \geq 3$  only if  $m \geq 4$ .

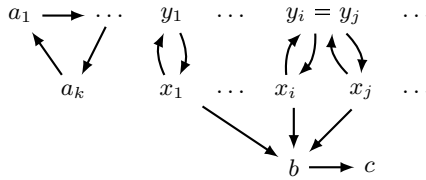


Figure 2. Illustration of the argumentation framework used for Lemma 1.

Since each  $x_i$  is only (mutually) attacked by  $y_i$ , and each agent  $i$  ranks her own  $x_i$  above  $y_i$ , the former is unattacked in  $\langle \mathcal{A}, \rightarrow_{\succ_i} \rangle$ . Thus,  $c$  belongs to the grounded extension for every agent, and  $b$  does not belong to that of any of them. But for the aggregate it is the other way round: it ranks every  $y_i$  above the corresponding  $x_i$ , and thus  $b$  will be in the grounded extension and  $c$  will not. So  $c$  is a witness for the violation of Acceptance Unanimity, and  $b$  is a witness for the violation of Rejection Unanimity.  $\square$

### 3.2. Impossibility Result for Condorcet Methods

Using Lemma 1, we can easily prove an impossibility result for Condorcet methods.

**Proposition 1.** *Every Condorcet method for  $m \geq 4$  arguments and  $n \geq 3$  agents violates both Acceptance Unanimity and Rejection Unanimity (unless  $n = m = 4$ ).*

*Proof.* Let  $F$  be an arbitrary Condorcet method for  $n \geq 3$  agents and  $m \geq 4$  arguments, for which  $n = m = 4$  is not the case. Consider the following profile  $\succ$ :

$$\begin{aligned} a_2 >_i a_1 >_i a_3 >_i a_4 >_i \dots >_i a_m & \text{ for every } i \in N \text{ with } i \equiv 0 \pmod 3 \\ a_3 >_i a_1 >_i a_2 >_i a_4 >_i \dots >_i a_m & \text{ for every } i \in N \text{ with } i \equiv 1 \pmod 3 \\ a_4 >_i a_1 >_i a_2 >_i a_3 >_i \dots >_i a_m & \text{ for every } i \in N \text{ with } i \equiv 2 \pmod 3 \end{aligned}$$

Then  $a_1$  is the Condorcet winner,<sup>4</sup> so  $a_1 = top(F(\succ))$ , while for each agent  $i$ , we have that  $a_1 \neq top(\succ_i)$ . Thus, by Lemma 1, condition (ii),  $F$  violates both axioms.  $\square$

Interestingly, not only does this proof technique fail to apply in case  $n = m = 4$ , but in Section 3.5 we shall see that for these parameters we in fact obtain a possibility result.

<sup>4</sup>Importantly, this holds for any  $n \geq 3$  and  $m \geq 4$ , as long as  $n = m = 4$  is not the case.



### 3.3. Impossibility Result for Positional Scoring Rules

Next, we establish a similar impossibility result for PSRs.

**Proposition 2.** *Every PSR for  $m \geq 4$  arguments violates both Acceptance Unanimity and Rejection Unanimity for some number  $n$  of agents.*

*Proof.* We first cover the case of  $m = 4$ . So consider an arbitrary PSR for  $m = 4$  arguments. W.l.o.g., this rule is induced by a scoring vector  $(1, s_2, s_3, 0)$  with  $1 \geq s_2 \geq s_3 \geq 0$ . We distinguish three cases:

- (1)  $0 \leq s_2 \leq \frac{1}{2}$ , but without the combination  $s_2 = \frac{1}{2}$  and  $s_3 = 0$
- (2)  $s_2 = \frac{1}{2}$  and  $s_3 = 0$
- (3)  $\frac{1}{2} < s_2 \leq 1$

For case (1), consider the following profile:

$$\begin{array}{l|l} 2 \text{ agents:} & a > b > c > d \\ 3 \text{ agents:} & b > a > d > c \\ 1 \text{ agent:} & c > a > b > d \end{array}$$

Let  $S_a$  be the total score of  $a$ , and so forth. The aggregate ranking is  $b > a > c > d$ , given that  $S_b = 3 + 2s_2 + s_3 > S_a = 2 + 4s_2 > S_c = 1 + 2s_3 > S_d = 3s_3$ . Now condition (i) of Lemma 1 applies, with distinguishing pairs  $(a, b)$  and  $(d, c)$ , so both axioms indeed are violated.

For case (2), we instead take the following profile:

$$\begin{array}{l|l} 1 \text{ agent:} & c > b > a > d \\ 1 \text{ agent:} & a > b > d > c \\ 1 \text{ agent:} & d > b > c > a \end{array}$$

We get  $S_b = \frac{3}{2} > S_a = S_c = S_d = 1$ , and condition (ii) of Lemma 1 applies.

For case (3), consider the following family of profiles:

$$\begin{array}{l|l} k \text{ agents:} & c > b > a > d \\ (k + 1) \text{ agents:} & a > b > d > c \\ (k + 2) \text{ agents:} & d > b > c > a \end{array}$$

We get the following scores:

$$\begin{array}{ll} S_a = k + 1 + ks_3 & S_c = k + (k + 2)s_3 \\ S_b = (3k + 3)s_2 & S_d = k + 2 + (k + 1)s_3 \end{array}$$

So  $S_d > S_a$  and  $S_d \geq S_c$ . Next, we show that there exists a  $k$  for which furthermore  $S_b > S_d$ , i.e., for which the aggregate will rank  $b$  at the top (unlike any of the agents). As  $s_2 \geq s_3$ , we have  $S_d \leq k + 2 + (k + 1)s_2$ . But  $k + 2 + (k + 1)s_2 < (3k + 3)s_2$  holds exactly when  $\frac{k+2}{2k+2} < s_2$ . As  $\frac{1}{2} < s_2$ , we can find a sufficiently large  $k$  that satisfies  $\frac{k+2}{2k+2} < s_2$  and thus ensures  $S_b > S_d$ . So condition (ii) of Lemma 1 applies also here.

It remains to be shown that the impossibility thus established for  $m = 4$  extends to all  $m \geq 4$ . But this is immediate: If we consider argumentation frameworks with  $m > 4$  cycle arguments of which  $m - 4$  arguments are isolated arguments that each attack themselves, we can carry through the exact same construction as above.  $\square$

It is worth noting that there are many values of  $n$  for which all PSRs violate the axioms. The construction used in the proof for case (1) works for any number of agents that is divisible by 6. That for case (2) only requires divisibility by 3. Finally, since case (3) is a limit argument, the PSRs will violate the axioms for any sufficiently large  $n$ .

### 3.4. Possibility Results for Argumentation Frameworks with Few Cycle Arguments

Propositions 1 and 2 are powerful impossibility results that rule out most aggregation rules in the general case. But it turns out that we can establish possibility results for specific parameters, when the number of arguments involved in cycles is very small. Specifically, for  $m = 2$  cycle arguments, all reasonable aggregation rules will work.

**Proposition 3.** *Every Condorcet method and every PSR for  $m = 2$  arguments satisfies both Acceptance Unanimity and Rejection Unanimity.*

*Proof.* For  $m = 2$ , there are only 2 arguments to compare, so all Condorcet methods and all PSRs reduce to the *simple majority rule*, under which the aggregate ranking is the ranking chosen most often [9]. This also implies that the aggregate ranking is chosen by at least one agent. Thus, if all agents accept (or reject) some argument, then the aggregate will also accept (or reject) that argument.  $\square$

For argumentation frameworks with at most  $m = 3$  arguments involved in cycles, we are able to identify at least one aggregation rule that satisfies our axioms, namely the Plurality rule (with ties broken according to the preferences of some fixed agent).

**Proposition 4.** *The Plurality rule with agent-based tie-breaking for  $m = 3$  arguments satisfies both Acceptance Unanimity and Rejection Unanimity.*

*Proof (sketch).* The central step in the proof is to show that, for  $m = 3$ , the Plurality rule with agent-based tie-breaking will always return a ranking that accepts the same cycle arguments as at least one agent. Note that, for this rule  $F$ , there is always an agent  $i$  such that  $top(>_i) = top(F(>))$  and, for  $(a, b) \in F(>)$ , there is always an agent  $j$  such that  $a >_j b$  (either by agent-based tie-breaking or because  $a = top(>_j)$ ). A careful case distinction now shows that, depending on the structure of the argumentation framework, the aggregate ranking will agree either with  $i$  or  $j$  on the cycle arguments. For instance, if  $a = top(F(>))$  and arguments  $b$  and  $c$  occur together in a cycle while  $a$  is self-attacking, then only the ranking between  $b$  and  $c$  matters and agent  $j$  agrees with the aggregate on the cycle arguments. In contrast, if all three arguments occur in a cycle together, then  $i$  would be our agent of choice.  $\square$

### 3.5. Possibility Result for Representative-Agent Rules

We end on a positive note and show that there is a large family of aggregation rules that satisfies both our axioms for any number of agents and any number of arguments.

**Proposition 5.** *Every representative-agent rule satisfies both Acceptance Unanimity and Rejection Unanimity.*

*Proof.* By definition, any representative-agent rule copies the ranking of one of the agents, so any property satisfied by all agents must also be satisfied by the aggregate.  $\square$

Whether this positive result is a suitable way around our impossibilities is up for debate. While the case for certain representative-agent rules has been made in the literature [10,23], they clearly are not as widely accepted as the Condorcet methods or the PSRs. Further, this family also includes obviously unattractive rules such as the dictatorships. But since the aggregation rule knows nothing about the underlying argumentation framework, it might be a good choice to follow the submitted rankings. For instance, arguments might share premises or rely on similar reasoning patterns, which might make certain rankings more reasonable than others. Representative-agent rules guarantee that we do not end up disregarding these hidden constraints.

It is worth pointing out that Proposition 3 is essentially a corollary to Proposition 5. Finally, recall that the impossibility result for Condorcet methods (Proposition 1) does not apply when  $n = m = 4$ . Indeed, for  $n = m = 4$ , there exists a Condorcet method that satisfies both axioms. What is special about the case of  $n = m = 4$  is that an argument can be a Condorcet winner only if it is ranked first by at least one agent. So the rule that selects the ranking where the Condorcet winner, if it exists, is the top argument and that otherwise selects the ranking of agent 1 is both a Condorcet method and a representative-agent rule—so Proposition 5 applies.

#### 4. Conclusion

We introduced a formal model to analyse methods for eliminating cycles from an abstract argumentation framework on the basis of the preferences held by the members of a group. We introduced two basic normative requirements encoding the following principle: unanimously held views by all members of the group should be reflected in the outcome returned by any such method. Our findings show that it is essentially impossible to uphold this principle by means of standard methods of preference aggregation, although the family of representative-agent rules offers a promising way out.

While it is not uncommon to uncover impossibilities in the context of social choice, we note that, at the technical level, our impossibilities are not immediately related to any of the best-known results in the field [7,8]. These typically exploit a conflict between the axiom of *independence* (or another axiom that implies it, such as *strategyproofness*) and the axiom of *nondictatorship*. This is a pattern we also often find in work at the interface of formal argumentation with social choice theory [12,15,17].

For future work, it will be important to explore alternative design choices for collective cycle elimination. For instance, asking everyone to report *separate rankings* for each cycle would be another natural choice (but care would need to be taken to ensure consistency in case of overlapping cycles). And while we opted for eliminating *all attacks* in a cycle going against our preferences, another natural option would be to only eliminate *one attack* per cycle. In general, we could also consider using preferences in other ways than eliminating attacks [5]. Further, it seems that our findings do not depend fully on the context of cycle elimination, as Lauren et al. [24] show that there are similar results even outside of abstract argumentation, namely in assumption-based argumentation.

*Acknowledgements.* We acknowledge the support of the Dutch Research Council (NWO) through project 639.023.811 (VICI “Collective Information”). The work of the second author furthermore was partially supported by the Austrian Science Fund (FWF) through project LoDEx I 6372-N (Grant-DOI 10.55776/I6372).

## References

- [1] Dung PM. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and  $n$ -Person Games. *Artificial Intelligence*. 1995;77(2):321-57.
- [2] Amgoud L, Cayrol C. On the Acceptability of Arguments in Preference-Based Argumentation. In: *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI)*; 1998. p. 1-7.
- [3] Bench-Capon TJM. Persuasion in Practical Argument Using Value-Based Argumentation Frameworks. *Journal of Logic and Computation*. 2003;13(3):429-48.
- [4] Modgil S. Reasoning about Preferences in Argumentation Frameworks. *Artificial Intelligence*. 2009;173(9):901-34.
- [5] Amgoud L, Vesic S. A New Approach for Preference-Based Argumentation Frameworks. *Annals of Mathematics and Artificial Intelligence*. 2011;63(2):149-83.
- [6] Baroni P, Caminada M, Giacomin M. Abstract Argumentation Frameworks and their Semantics. In: *Handbook of Formal Argumentation*. College Publications; 2018. p. 159-236.
- [7] Arrow KJ, Sen A, Suzumura K, editors. *Handbook of Social Choice and Welfare*. vol. 1. North-Holland; 2002.
- [8] Brandt F, Conitzer V, Endriss U, Lang J, Procaccia AD, editors. *Handbook of Computational Social Choice*. Cambridge University Press; 2016.
- [9] Zwicker WS. Introduction to the Theory of Voting. In: *Handbook of Computational Social Choice*. Cambridge University Press; 2016. p. 23-56.
- [10] Endriss U, Grandi U. Binary Aggregation by Selection of the Most Representative Voter. In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*; 2014. p. 668-74.
- [11] Bodanza G, Tohmé FA, Auday M. Collective Argumentation: A Survey of Aggregation Issues around Argumentation Frameworks. *Argument & Computation*. 2017;8(1):1-34.
- [12] Tohmé FA, Bodanza GA, Simari GR. Aggregation of Attack Relations: A Social-choice Theoretical Analysis of Defeasibility Criteria. In: *Proceedings of the 5th International Symposium on Foundations of Information and Knowledge Systems (FoIKS)*. Springer; 2008. p. 8-23.
- [13] Dunne PE, Marquis P, Wooldridge M. Argument Aggregation: Basic Axioms and Complexity Results. In: *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA)*. IOS Press; 2012. p. 129-40.
- [14] Delobelle J, Konieczny S, Vesic S. On the Aggregation of Argumentation Frameworks: Operators and Postulates. *Journal of Logic and Computation*. 2018;1671-99.
- [15] Chen W, Endriss U. Preservation of Semantic Properties in Collective Argumentation: The Case of Aggregating Abstract Argumentation Frameworks. *Artificial Intelligence*. 2019;269:27-48.
- [16] Caminada M, Pigozzi G. On Judgment Aggregation in Abstract Argumentation. *Autonomous Agents and Multiagent Systems*. 2011;22(1):64-102.
- [17] Rahwan I, Tohmé FA. Collective Argument Evaluation as Judgement Aggregation. In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*; 2010. p. 417-24.
- [18] Awad E, Booth R, Tohmé F, Rahwan I. Judgement Aggregation in Multi-agent Argumentation. *Journal of Logic and Computation*. 2017;27(1):227-59.
- [19] Chen W, Endriss U. Aggregating Alternative Extensions of Abstract Argumentation Frameworks: Preservation Results for Quota Rules. In: *Proceedings of the 7th International Conference on Computational Models of Argument (COMMA)*; 2018. p. 425-36.
- [20] Airiau S, Bonzon E, Endriss U, Maudet N, Rossit J. Rationalisation of Profiles of Abstract Argumentation Frameworks: Characterisation and Complexity. *Journal of Artificial Intelligence Research*. 2017;60:149-77.
- [21] Leite J, Martins J. Social Abstract Argumentation. In: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*; 2011. p. 2287-92.
- [22] Bernreiter M, Maly J, Nardi O, Woltran S. Combining Voting and Abstract Argumentation to Understand Online Discussions. In: *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*; 2024. In press.
- [23] Caragiannis I, Procaccia AD, Shah N. Modal Ranking: A Uniquely Robust Voting Rule. In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*; 2014. p. 616-22.
- [24] Lauren S, Belardinelli F, Toni F. Aggregating Bipolar Opinions. In: *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*; 2021. p. 746-54.