**Noname manuscript No.**
(will be inserted by the editor)

# On Judgment Aggregation in Abstract Argumentation

**Martin Caminada · Gabriella Pigozzi**

Received: date / Accepted: date

**Abstract** Judgment aggregation is a field in which individuals are required to vote for or against a certain decision (the *conclusion*) while providing reasons for their choice. The reasons and the conclusion are logically connected propositions. The problem is how a collective judgment on logically interconnected propositions can be defined from individual judgments on the same propositions. It turns out that, despite the fact that the individuals are logically consistent, the aggregation of their judgments may lead to an inconsistent group outcome, where the reasons do not support the conclusion. However, in this paper we claim that collective irrationality should not be the only worry of judgment aggregation. For example, judgment aggregation would not reject a consistent combination of reasons and conclusion that no member voted for. In our view this may not be a desirable solution. This motivates our research about when a social outcome is 'compatible' with the individuals' judgments. The key notion that we want to capture is that any individual member has to be able to defend the collective decision. This is guaranteed when the group outcome is compatible with its members views. Judgment aggregation problems are usually studied using classical propositional logic. However, for our analysis we use an argumentation approach to judgment aggregation problems. Indeed the question of how individual evaluations can be combined into a collective one can also be addressed in abstract argumentation. We

Martin Caminada
Individual and Collective Reasoning
Computer Science and Communication
University of Luxembourg
6, Rue Richard Coudenhove Kalergi, L-1359 – LUXEMBOURG
Tel.: +352-466644–5485
Fax: +352 46 66 44 5621
E-mail: martin.caminada@uni.lu

Gabriella Pigozzi
Individual and Collective Reasoning
Computer Science and Communication
University of Luxembourg
6, Rue Richard Coudenhove Kalergi, L-1359 – LUXEMBOURG
Tel.: +352-466644–5442
Fax: +352 46 66 44 5621
E-mail: gabriella.pigozzi@uni.lu

introduce three aggregation operators that satisfy the condition above, and we offer two definitions of compatibility. Not only does our proposal satisfy a good number of standard judgment aggregation postulates, but it also avoids the problem of individual members of a group having to become committed to a group judgment that is in conflict with their own individual positions.

**Keywords** Judgment aggregation · Discursive dilemma · Argumentation · Group decision making

## 1 Introduction

Groups are often required to make decisions and to justify them. For example, courts have to provide reasons for declaring a defendant liable or innocent, governments need to allege raises in income taxes, hiring committees must be able to explain why they decided to hire a certain candidate. Groups decisions are the result of the aggregation of the individual judgments on the issues at stake.

Judgment aggregation is a recent discipline in which group members have to vote for or against a certain decision (the *conclusion*) and also provide reasons for their choice. Reasons, conclusion and the logical connections between them are given in the decision problem. Judgment aggregation investigates how individual consistent judgments on these logically interconnected propositions shall be mapped into an equally consistent social judgment on the same propositions [23, 22, 24, 26]. The difficulty lies in the fact that a natural aggregation procedure, like the majority rule, turns out to generate possibly inconsistent collective outcomes, where the reasons do not support the selected conclusion. Even worse, it can be shown that *any* judgment aggregation procedure that satisfies some desirable properties is condemned to produce sometimes irrational outcomes.

Whereas the literature on judgment aggregation is concerned with the unpleasant occurrence of irrational collective outcomes, our interest is not only to guarantee a consistent group outcome, but also that such outcome is 'compatible' with the individual judgments. Group inconsistency is not the only undesirable outcome. It may happen, for example, that majority rule selects as social outcome a consistent combination of reasons and conclusion that actually no member voted for. Such situation may be not a desirable collective outcome as it may conflict with some of its members' judgments. This motivates our research question: when is a group outcome 'compatible' with its members' judgments? We are interested in group decision making in which any group member is able to defend the group decision without having to argue against his own private beliefs. This is what we call a 'compatible' group decision.

As an example of a judgment aggregation problem and of the desirability of our property, suppose that three European leaders, Jan Peter, Yves and Jean-Claude, meet to decide whether to proceed to a bailout of Hortis Bank by nationalizing it. They agree that the bailout of Hortis Bank should be done if and only if the following are judged to be both true: (i) the survival of Hortis Bank is critical for Europe's financial sector, and (ii) Hortis Bank can be legally nationalized under current European law. Suppose that Jan Peter, Yves and Jean-Claude make their judgments according to Table 1.

It can then happen that, despite Jan Peter, Yves and Jean-Claude being logically consistent individuals, they may have to face a situation in which two of them believe

| | $p$ =Critical survival? | $q$=Legal nationalization? | $r$=Bailout? |
|---|---|---|---|
| Jan Peter | Yes | No | No |
| Yves | No | Yes | No |
| Jean-Claude | Yes | Yes | Yes |
| Majority | Yes | Yes | No |

**Table 1** Discursive dilemma. The Hortis Bank should be nationalized only if it is the case that its survival is critical for Europe's financial sector and that it can be legally nationalized under current European law.

that the Hortis Bank should not be nationalized. However, they will not be able to justify their decision because two of them are also convinced that the survival of Hortis Bank is critical for Europe's financial sector and (another) majority deems that Hortis Bank can be legally nationalized under current European law. So a majority believes that both (i) and (ii) are true *and* that the Hortis Bank should not be nationalized, which is a violation of the adopted decision rule. What should they decide? Should the Hortis Bank not be nationalized despite the fact that there is a majority believing that there are reasons for taking that action, or should they proceed to bailout Hortis Bank on the basis of the favorable judgment of only one person? This is an instance of the so-called *discursive dilemma* that troubles judgment aggregation. Now, suppose that proposition-wise majority voting in the example above gives a logically consistent solution (for this it is enough, for example, to switch Jan Peter's judgment on the critical survival issue from yes to no). However, Jean-Claude is the only one favoring the bailout of Hortis Bank. Should Jean-Claude revise his belief and oppose the bailout? Is the fact that Jean-Claude deems Jan Peter and Yves equally intelligent, informed and impartial as himself enough for him to change his opinion? We claim that it should not, as Jean-Claude must be able to defend the decision of the group he belongs to.

A set of arguments and a defeat relation among them is called *argumentation framework*. Given an argumentation framework, argumentation theory identifies and characterizes the sets of arguments (*extensions*) that can reasonably survive the conflicts expressed in the argumentation framework, and therefore can collectively be accepted. In general, there are several possible extensions for a set of arguments and a defeat relation on them [13]. Formal argumentation can be seen as an abstract generalized way of nonmonotonic reasoning, and several nonmonotonic formalisms including Nute's Defeasible Logic [18], Simari's DeLP [16], logic programming and Default Logic [13] have been reformulated in the form of formal argumentation.

In this paper we are interested in argumentation in a multi-agent setting [38, 7]. Given an argumentation framework, different individuals may provide different evaluations regarding what should be accepted and rejected. How can individual evaluations be mapped into a collective one? Similarly as in judgment aggregation, where the acceptance or rejection of a proposition may yield to the acceptance or rejection of another one, in argumentation the acceptance of one argument may force to reject another one [34]. The aggregation of individual evaluations of a given argumentation framework raises the same problems as the aggregation of individual judgments. We will see that argument-by-argument majority voting may result in an unacceptable extension, as the proposition-wise majority voting may output an inconsistent collective judgment set. Judgment aggregation can be then addressed as the problem of combining different individual evaluations of the situation represented by an argumentation framework. The reason for using abstract argumentation is twofold: on the one hand, the existence of different argumentation semantics allows us to be flexible when defining which social outcomes are permissible. On the other hand, it allows us to bring judgment aggregation from classical logic to nonmonotonic reasoning.

Any mechanism dealing with the aggregation of different positions needs to handle conflicts. We define and investigate three operators, sceptical, credulous and super credulous, and investigate their properties. We will show that, by iterating the aggregation process, not only we ensure a collective consistent decision, but that this is also unique. The operators introduced here are suitable aggregation operators for group decisions where the participation of the group does not lead any individual to endorse a position against his beliefs. Hence, they reflect a 'compatible' collective outcome with the individual positions.

The paper is structured as follows: in Section 2 we present the problem of judgment aggregation. Section 3 is devoted to outline the abstract argumentation framework, and in Section 4 we express concepts like complete, grounded, preferred, stable or semi-stable semantics in terms of argument labellings. In Section 5 we show how judgment aggregation concepts can be applied to formal argumentation. The sceptical, credulous and super-credulous outcomes are introduced in Sections 6, 7, 8 respectively. We then prove some properties of these operators in Section 9. Section 10 on related work and the conclusions end the paper.

## 2 Judgment Aggregation

The problem of judgment aggregation was discussed by Kornhauser and Sager [22, 23]. In their example, a court has to make a decision on whether a person is liable of breaching a contract. The judges have to reach a verdict following the legal doctrine. However, the problem of aggregating individual judgments on logically connected propositions is not confined to the domain in which the legal doctrine dictates that certain judgments are to be made by reference to certain laws.

The judgment aggregation framework includes a set of issues (and their negations) on which the judgments have to be made. This is called the *agenda*. The propositions in the agenda are sentences in classical propositional logic, though more expressive logics like modal, predicate, conditional and deontic can also be used [8]. Additional constraints may be added. These are not part of the agenda, but restrict the sets of admissible judgments. For example, if $p, q, r, \ldots$ are proposition variables, the agenda of the bailout of the bank example contains the propositions $\{p, \neg p, q, \neg q, r, \neg r\}$ with the additional requirement that $(p \wedge q) \leftrightarrow r$. An example of an agenda without additional constraints is $\{p, \neg p, p \rightarrow q, \neg(p \rightarrow q), q, \neg q\}$. It is assumed that all the individuals express logically consistent judgments on the propositions of the agenda. How the individual judgments on the propositions in the agenda should be aggregated into a social consistent judgment on the same propositions is the question that judgment aggregation aims to answer.

Let us consider the bailout of Hortis Bank example again and the judgments as in Table 1. Jan Peter, Yves and Jean-Claude express a consistent opinion, i.e. they say yes to the bailout if and only if they say yes to both $p$ and $q$. However, proposition-wise majority voting (consisting in the separate aggregation of the votes for each proposition $p$, $q$ and $r$ via majority rule) results in a majority for $p$ and $q$ and yet a majority for $\neg r$. This is an inconsistent collective result, in the sense that $\{p, q, \neg r, (p \wedge q) \leftrightarrow r\}$ is inconsistent in propositional logic. The problem lies in the fact that majority voting can lead a group of rational agents to endorse an irrational collective judgment. The literature on judgment aggregation refers to such situation as the *discursive dilemma*.

Majority voting fails to guarantee a consistent outcome whenever the agenda has a minimally inconsistent subset of three or more propositions [31,10]. In the bailout of the Hortis bank example, a minimally inconsistent set with three propositions is $\{p, q, \neg r\}$. Furthermore, the problem of aggregating individual judgments is not restricted to majority voting, but it applies to all aggregation procedures satisfying some desirable conditions. We now introduce the formal framework of judgment aggregation and the properties imposed on the aggregation rule.

A set of agents $N = \{1, \ldots, n\}$ makes judgments on logically interconnected propositions. $L$ is a language with atomic propositions $p, q, r, \ldots$, including the complex formulas $\neg p, (p \wedge q), (p \vee q), (p \rightarrow q), (p \leftrightarrow q)$. The agenda is denoted by $\Phi \subseteq L$. $\Phi$ is the set of propositions on which the agents have to express a judgment and it does not contain tautologies or contradictions. The agenda is assumed to be finite and is closed under negation: if $p \in \Phi$, then $\neg p \in \Phi$.[1] Each doubly negated proposition $\neg\neg p$ is identified with the non negated proposition $p$. A subset $J \subseteq \Phi$ is called (individual or collective) *judgment set* and it is the set of propositions believed by an individual or the group. A judgment set is *consistent* if it is a consistent set in $L$, and is *complete* if, for any $p \in \Phi$, $p \in J$ or $\neg p \in J$ but not both. An $n$-tuple $(J_1, \ldots, J_n)$ of agent judgment sets is called a *profile*. Finally, a *judgment aggregation rule* $F$ assigns a collective judgment set $J$ to each profile $(J_1, \ldots, J_n)$ of agent judgment sets.

Judgment aggregation problems are a generalization of the preference aggregation issues in the tradition of social choice theory [1,24,9]. Following an axiomatic approach in the Arrowian style of social choice theory [1], List and Pettit [24,25] showed that, given an agenda with at least two atomic propositions and at least one suitable composite proposition (and their negations), there exists no judgment aggregation rule $F$ that satisfies universal domain, collective rationality, anonymity and systematicity.

**Universal domain**: The domain of $F$ is the set of all profiles of consistent and complete judgment sets.

**Collective rationality**: Only complete, consistent and deductively closed collective judgments are permissible as outputs.

**Anonymity**: For any profiles $(J_1, \ldots, J_n)$, $(J'_1, \ldots, J'_n)$ in the domain that are permutations of each other, $F(J_1, \ldots, J_n) = F(J'_1, \ldots, J'_n)$. Intuitively, this means that all agents have equal weight.

**Systematicity**: For any $p, q \in \Phi$ and any profiles $(J_1, \ldots, J_n)$, $(J'_1, \ldots, J'_n)$ in the domain, if $\forall j \in N, p \in J_j \leftrightarrow q \in J'_j$, then $p \in F(J_1, \ldots, J_n) \leftrightarrow q \in F(J'_1, \ldots J'_n)$. This condition ensures that the collective judgment on each proposition depends only on the agent judgments on that proposition, and that the aggregation rule is the same across all propositions.

Systematicity is a strong condition and can be weakened to independence:

**Independence**: For any $p \in \Phi$ and any profiles $(J_1, \ldots, J_n)$, $(J'_1, \ldots, J'_n)$ in the domain, if $\forall j \in N, p \in J_j \leftrightarrow p \in J'_j$, then $p \in F(J_1, \ldots, J_n) \leftrightarrow p \in F(J'_1, \ldots J'_n)$. In other words, independence is systematicity without the neutrality condition, requiring that all propositions are equally treated.

[1] To increase readability, only the positive issues of the agenda are listed in the tables. It is assumed that, for any issue in the agenda, an individual deems that issue to be true if and only if he deems its negation to be false.

For many agendas, the independence condition is sufficient for an impossibility result to occur [25]. A stronger impossibility theorem was proved by weakening the anonymity condition to non-dictatorship [32]:

**Non-dictatorship**: There exists no $i \in N$ such that, for any profile $(J_1, \ldots, J_n)$ in the domain, $F(J_1, \ldots, J_n) = J_i$.

One more result from the judgment aggregation literature should be mentioned. It is the theorem that restate precisely Arrow's famous impossibility theorem for judgment aggregation: if an aggregation rule satisfies universal domain, collective rationality, independence and unanimity, then it is dictatorial [9,11]. Unanimity corresponds to Arrow's weak Pareto principle and requires that, if all the members in the group adopt the same position on a certain issue, that position will be adopted at the collective level as well.

Inconsistent collective outcomes as in the example above rest upon the issue-by-issue aggregation of propositions that are logically connected. Just like Arrow's condition of "independence of irrelevant alternatives" plays a central role in his famous impossibility results, independence remains a controversial condition in judgment aggregation. Our operators for aggregating individual evaluations of an argumentation framework also use an argument-by-argument counting procedure. In this way we can ensure the group to be responsive to the individuals' evaluations. However, in order to avoid incoherent collective outcomes, the process is iterated until a reasonable (according to the semantics in place) extension is obtained.

Judgment aggregation is traditionally studied in classical propositional logic, where agents assign 0/1 values to logically connected propositions.[2] However, the question of how individual evaluations can be combined into a collective one can also be addressed in abstract argumentation: given an argumentation framework and individual evaluations of what to accept and what to reject, what is the resulting group evaluation? In the next sections we define judgment aggregation problems in abstract argumentation. After introducing the main notions of the abstract argumentation formalism, we show how judgment aggregation concepts can be applied to formal argumentation.

## 3 Argumentation Preliminaries

An *argumentation framework* [13] consists of a set of arguments and a defeat relation on these arguments. In order to simplify the discussion, we only consider finite argumentation frameworks.

**Definition 1** Let $U$ be the universe of all possible arguments. An *argumentation framework* is a pair $(Ar, \mathit{def})$ where $Ar$ is a finite subset of $U$ and $\mathit{def} \subseteq Ar \times Ar$.

We say that an argument $A$ *defeats* an argument $B$ iff $(A, B) \in \mathit{def}$.

An argumentation framework can be represented as a directed graph in which the arguments are represented as nodes and the defeat relations are represented as arrows. For instance, argumentation framework $(Ar, \mathit{def})$ where $Ar = \{A, B, C, D, E\}$ and $\mathit{def} = \{(A, B), (B, A), (B, C), (C, D), (D, E), (E, C)\}$ is represented in Figure 1.

[2] For representations of judgment aggregation problems in more expressive logics than standard propositional logic, see [8].

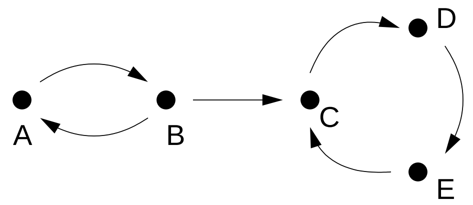


**Fig. 1** An argumentation framework represented as a directed graph.

The shorthand notation $A^+$ and $A^-$ stands for, respectively, the set of arguments defeated by $A$ and the set of arguments that defeat $A$. Likewise, if $\mathcal{A}rgs$ is a set of arguments, then we write $\mathcal{A}rgs^+$ for the set of arguments that is defeated by at least one argument in $\mathcal{A}rgs$, and $\mathcal{A}rgs^-$ for the set of arguments that defeat at least one argument in $\mathcal{A}rgs$. In the definition below, $F(\mathcal{A}rgs)$ stands for the set of arguments that are acceptable in the sense of [13].

**Definition 2** Let $(Ar, \mathit{def})$ be an argumentation framework, $A \in Ar$ and $\mathcal{A}rgs \subseteq Ar$.
We define $A^+$ as $\{B \mid A \; \mathit{def} \; B\}$ and $\mathcal{A}rgs^+$ as $\{B \mid A \; \mathit{def} \; B$ for some $A \in \mathcal{A}rgs\}$.
We define $A^-$ as $\{B \mid B \; \mathit{def} \; A\}$ and $\mathcal{A}rgs^-$ as $\{B \mid B \; \mathit{def} \; A$ for some $A \in \mathcal{A}rgs\}$.
$\mathcal{A}rgs$ is *conflict-free* iff $\mathcal{A}rgs \cap \mathcal{A}rgs^+ = \emptyset$.
$\mathcal{A}rgs$ *defends* an argument $A$ iff $A^- \subseteq \mathcal{A}rgs^+$.
We define the function $F : 2^{Ar} \rightarrow 2^{Ar}$ as $F(\mathcal{A}rgs) = \{A \mid A$ is defended by $\mathcal{A}rgs\}$.

In the definition below, definitions of grounded, preferred and stable semantics are described in terms of complete semantics, which has the advantage of making the proofs in the remainder of this paper more straightforward. These descriptions are not literally the same as the ones provided by Dung [13], but as was first stated in [2], these are in fact equivalent to Dung's original versions of grounded, preferred and stable semantics.

**Definition 3** Let $(Ar, \mathit{def})$ be an argumentation framework and let $\mathcal{A}rgs \subseteq Ar$ be a conflict-free set of arguments.

- $\mathcal{A}rgs$ is *admissible* iff $\mathcal{A}rgs \subseteq F(\mathcal{A}rgs)$.
- $\mathcal{A}rgs$ is a *complete* extension iff $\mathcal{A}rgs = F(\mathcal{A}rgs)$.
- $\mathcal{A}rgs$ is a *grounded* extension iff $\mathcal{A}rgs$ is the minimal (w.r.t. set-inclusion) complete extension.
- $\mathcal{A}rgs$ is a *preferred* extension iff $\mathcal{A}rgs$ is a maximal (w.r.t. set-inclusion) complete extension.
- $\mathcal{A}rgs$ is a *stable* extension iff $\mathcal{A}rgs$ is a complete extension that defeats every argument in $Ar \backslash \mathcal{A}rgs$.
- $\mathcal{A}rgs$ is a *semi-stable* extension iff $\mathcal{A}rgs$ is a complete extension where $\mathcal{A}rgs \cup \mathcal{A}rgs^+$ is maximal (w.r.t. set-inclusion).

As an example, in the argumentation framework of Figure 1 $\{B, D\}$ is a stable extension, $\{A\}$ is a preferred extension which is not stable or semi-stable, $\emptyset$ is the grounded extension, and $\{B\}$ is an admissible set which is not a complete extension.

It is known that for every argumentation framework, there exists at least one admissible set (the empty set), exactly one grounded extension, one or more complete extensions, one or more preferred extensions and zero or more stable extensions. Moreover, when the set of arguments in the argumentation framework is finite, as is assumed in the current paper, there also exist one or more semi-stable extensions.

An overview of how the various extensions are related to each other is provided in Figure 2. The fact that every stable extension is also a semi-stable extension, and that every semi-stable extension is also a preferred extension was first stated in [2]. All other relations shown in Figure 2 have originally been stated in [13].

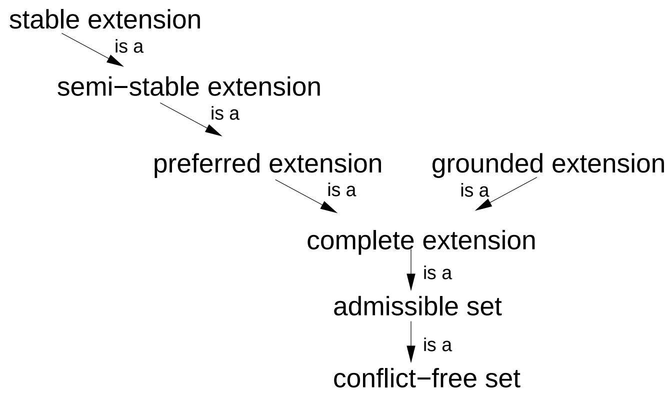


**Fig. 2** An overview of argumentation semantics (extension based).

In essence, an argumentation semantics can be seen as a function that, given an argumentation framework, yields zero or more sets of arguments which can be collectively accepted.

**Definition 4** Let $U$ be the universe of arguments, and let $\mathcal{AF}$ be the set of all possible argumentation frameworks using this universe. An *extension based semantics* is a function $\mathcal{S} : \mathcal{AF} \rightarrow 2^{2^U}$.

For instance, if we have an argumentation framework $AF = (Ar, \mathit{def})$ then:

- $\mathcal{S}_{stable}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is a stable extension of $AF\}$
- $\mathcal{S}_{semi-stable}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is a semi-stable extension of $AF\}$
- $\mathcal{S}_{preferred}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is a preferred extension of $AF\}$
- $\mathcal{S}_{complete}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is a complete extension of $AF\}$
- $\mathcal{S}_{grounded}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is the grounded extension of $AF\}$
- $\mathcal{S}_{admissible}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is an admissible set of $AF\}$
- $\mathcal{S}_{conflict-free}(AF) = \{\mathcal{A}rgs \mid \mathcal{A}rgs$ is a conflict-free set of $AF\}$

As illustrated in Figure 2, it holds for any argumentation framework $AF = (Ar, \mathit{def})$ that $\mathcal{S}_{stable}(AF) \subseteq \mathcal{S}_{semi-stable}(AF) \subseteq \mathcal{S}_{preferred}(AF) \subseteq \mathcal{S}_{complete}(AF)$ as well as that $\mathcal{S}_{grounded}(AF) \subseteq \mathcal{S}_{complete}(AF) \subseteq \mathcal{S}_{admissible} \subseteq \mathcal{S}_{conflict-free}$.

## 4 Argument Labellings

The concepts of admissibility, as well as that of complete, grounded, preferred, stable or semi-stable semantics were originally stated in terms of sets of arguments. It is equally well possible, however, to express these concepts using argument labellings. This approach was pioneered by Pollock [35] and Jakobovits and Vermeir [20], and has more recently been extended by Caminada [2,3], Vreeswijk [41] and Verheij [39]. The idea of a labelling is to associate with each argument exactly one label, which can either be `in`, `out` or `undec`. The label `in` indicates that the argument is explicitly accepted, the label `out` indicates that the argument is explicitly rejected, and the label `undec` indicates that the status of the argument is undecided, meaning that one abstains from an explicit judgment whether the argument is `in` or `out`.

**Definition 5** Let $(Ar, \mathit{def})$ be an argumentation framework. A *labelling* is a total function $\mathcal{L} : Ar \longrightarrow \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$.

We write $\mathtt{in}(\mathcal{L})$ for $\{A \mid \mathcal{L}(A) = \mathtt{in}\}$, $\mathtt{out}(\mathcal{L})$ for $\{A \mid \mathcal{L}(A) = \mathtt{out}\}$ and $\mathtt{undec}(\mathcal{L})$ for $\{A \mid \mathcal{L}(A) = \mathtt{undec}\}$. Sometimes, we write a labelling $\mathcal{L}$ as a triple $(\mathcal{A}rgs_1, \mathcal{A}rgs_2, \mathcal{A}rgs_3)$ where $\mathcal{A}rgs_1 = \mathtt{in}(\mathcal{L})$, $\mathcal{A}rgs_2 = \mathtt{out}(\mathcal{L})$ and $\mathcal{A}rgs_3 = \mathtt{undec}(\mathcal{L})$.

We distinguish three special kinds of labellings. The *all-in labelling* is a labelling that labels every argument `in`. The *all-out labelling* is a labelling that labels every argument `out`. The *all-undec labelling* is a labelling that labels every argument `undec`.

**Definition 6** Let $\mathcal{L}$ be a labelling of argumentation framework $(Ar, \mathit{def})$. We say that $\mathcal{L}$ is conflict-free iff for each $A, B \in Ar$, if $\mathcal{L}(A) = \mathtt{in}$ and $B$ defeats $A$ then $\mathcal{L}(B) \neq \mathtt{in}$.

**Definition 7** Let $\mathcal{L}$ be a labelling of argumentation framework $(Ar, \mathit{def})$ and let $A \in Ar$. We say that:

1. $A$ is *illegally* `in` iff $A$ is labelled `in` but not all its defeaters are labelled `out`
2. $A$ is *illegally* `out` iff $A$ is labelled `out` but it does not have a defeater that is labelled `in`
3. $A$ is *illegally* `undec` iff $A$ is labelled `undec` but either all its defeaters are labelled `out` or it has a defeater that is labelled `in`.

We say that an argument is *legally* `in` iff it is labelled `in` and is not illegally `in`. We say that an argument is *legally* `out` iff it is labelled `out` and is not illegally `out`. We say that an argument is *legally* `undec` iff it is labelled undec and is not illegally `undec`.

**Definition 8** An *admissible labelling* is a labelling without arguments that are illegally `in` and without arguments that are illegally `out`.

**Definition 9** A *complete labelling* is a labelling without arguments that are illegally `in`, without arguments that are illegally `out` and without arguments that are illegally `undec`.

From Definition 6, 8 and 9 it immediately follows that each complete labelling is also an admissible labelling, and each admissible labelling is also a conflict-free labelling.

**Lemma 1 (Lemma 1 [3])** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be complete labellings of argumentation framework $AF = (Ar, \mathit{def})$. It holds that $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2)$ iff $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2)$.*

It is interesting to notice that an admissible labelling corresponds with the notion of an admissible set.

**Theorem 1 (Theorem 1, [3])** *Let $(Ar, \mathit{def})$ be an argumentation framework and $\mathcal{A}rgs \subseteq Ar$. $\mathcal{A}rgs$ is an admissible set iff there exists an admissible labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$.*

The notion of a complete labelling corresponds to the notion of a complete extension.

**Theorem 2 (Theorem 2, [3])** *Let $(Ar, def)$ be an argumentation framework and $\mathcal{A}rgs \subseteq Ar$. $\mathcal{A}rgs$ is a complete extension iff there exists a complete labelling $\mathcal{L}$ with* $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$.

As an aside, in [2], a slightly different definition of a (complete) labelling is given, but equivalence can be shown.

**Theorem 3** *Let $\mathcal{L}$ be a labelling of argumentation framework $(Ar, def)$. It holds that $\mathcal{L}$ is a complete labelling iff for each $A \in Ar$:*

1. *$A$ is labelled* `in` *by $\mathcal{L}$ iff every defeater of $A$ is labelled* `out` *by $\mathcal{L}$, and*
2. *$A$ is labelled* `out` *by $\mathcal{L}$ iff $A$ has a defeater that is labelled* `in` *by $\mathcal{L}$.*

*Proof*
"$\Longrightarrow$": Let $\mathcal{L}$ be a complete labelling. From the fact that $\mathcal{L}$ does not have any argument illegally `in` or illegally `out`, it follows that if $A$ is labelled `in` then every defeater of $A$ is labelled `out` (i). We now prove that if every defeater of $A$ is labelled `out` then $A$ is labelled `in`. Let $A$ be an argument of which all defeaters are labelled `out`. Then $A$ cannot be labelled `out` (because then $A$ would be illegally `out` and $\mathcal{L}$ would not be a complete labelling). Also, $A$ cannot be labelled `undec` (because then $A$ would be illegally `undec` and $\mathcal{L}$ would not be a complete labelling). From the fact that $A$ cannot be labelled `out` or `undec` it follows that $A$ is labelled `in`. So if every defeater of $A$ is labelled `out` then $A$ is labelled `in` (ii).
From the fact that $\mathcal{L}$ does not have any argument illegally `out` it follows that if $A$ is labelled `out`, then $A$ has a defeater that is labelled `in` (iii). We now prove that if $A$ has a defeater that is labelled `in` then $A$ is labelled `out`. Let $A$ be an argument that has a defeater that is labelled `in`. Then $A$ cannot be labelled `in` (because then $A$ would be illegally `in` and $\mathcal{L}$ would not be a complete labelling). Also, $A$ cannot be labelled `undec` (because then $A$ would be illegally `undec` and $\mathcal{L}$ would not be a complete labelling). From the fact that $A$ cannot be labelled `in` or `undec` it follows that $A$ is labelled `out`. So if $A$ has a defeater that is labelled `in` then $A$ is labelled `out` (iv).
From (i) and (ii) it follows that (1). From (iii) and (iv) it follows that (2).
"$\Longleftarrow$": Let $\mathcal{L}$ be a labelling of which every argument $A$ satisfies (1) and (2). From (1) it immediately follows that if $A$ is labelled `in` then every defeater of $A$ is labelled `out`, so $\mathcal{L}$ does not have any argument illegally `in`. From (2) it immediately follows that if $A$ is labelled `out` then $A$ has a defeater that is labelled `in`, so $\mathcal{L}$ does not have any argument illegally `out`. We now prove that $\mathcal{L}$ also does not have any argument illegally `undec`. Let $A$ be an argument that is illegally `undec`. We distinguish two cases.

- All defeaters of $A$ are labelled `out`. But then from (1) it follows that $A$ is labelled `in`. Contradiction.
- $A$ has a defeater that is labelled `in`. But then from (2) it follows that $A$ is labelled `out`. Contradiction.

Both cases end in a contradiction, so $A$ cannot be illegally `undec`.

**Definition 10** Let $\mathcal{L}$ be a complete labelling.

- We say that $\mathcal{L}$ is a *grounded labelling* iff $\mathtt{in}(\mathcal{L})$ is minimal (w.r.t. set inclusion) among all complete labellings.

– We say that $\mathcal{L}$ is a *preferred labelling* iff $\mathtt{in}(\mathcal{L})$ is maximal (w.r.t. set inclusion) among all complete labellings.
– We say that $\mathcal{L}$ is a *stable labelling* iff $\mathtt{undec}(\mathcal{L}) = \emptyset$.
– We say that $\mathcal{L}$ is a *semi-stable labelling* iff $\mathtt{undec}(\mathcal{L})$ is minimal (w.r.t. set inclusion) among all complete labellings.

The notions of a grounded, preferred, stable and semi-stable labelling correspond to the notions of a grounded, preferred, stable and semi-stable extension, respectively.

**Theorem 4** *Let $(Ar, def)$ be an argumentation framework and $\mathcal{A}rgs \subseteq Ar$.*

– *$\mathcal{A}rgs$ is the grounded extension iff there exists a grounded labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$*
– *$\mathcal{A}rgs$ is a preferred extension iff there exists a preferred labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$.*
– *$\mathcal{A}rgs$ is a stable extension iff there exists a stable labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$.*
– *$\mathcal{A}rgs$ is a semi-stable extension iff there exists a semi-stable labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$.*

*Proof* Using the results of Theorem 2 this then follows in a straightforward way from Definition 3 and Definition 10.

An overview of how the various labellings are related to each other is provided in Figure 3.

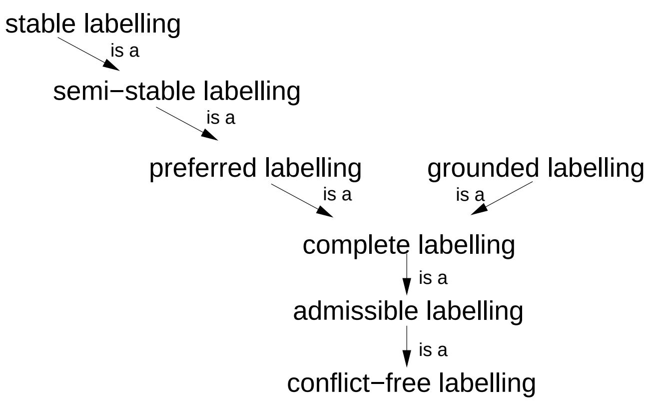


**Fig. 3** An overview of argumentation semantics (labelling based).

In essence, a labelling based semantics can be seen as a function that, given an argumentation framework, yields zero or more labellings, each of which can be seen as a reasonable position that one can take in the presence of the argumentation framework.

**Definition 11** Let $\mathcal{AF}$ be the set of all possible argumentation frameworks using a universe $U$. Let $\mathcal{L}abellings$ be $\{\mathcal{L} \mid$ there exists an argumentation framework $AF \in \mathcal{AF}$ such that $\mathcal{L}$ is a labelling of $AF\}$. A *labelling based semantics* is a function $\mathcal{T} : \mathcal{AF} \to 2^{\mathcal{L}abellings}$.

For instance, if we have an argumentation framework $AF = (Ar, def)$ then:

- $\mathcal{T}_{stable}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is a stable labelling of $AF\}$
- $\mathcal{T}_{semi-stable}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is a semi-stable labelling of $AF\}$
- $\mathcal{T}_{preferred}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is a preferred labelling of $AF\}$
- $\mathcal{T}_{complete}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is a complete labelling of $AF\}$
- $\mathcal{T}_{grounded}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is the grounded labelling of $AF\}$
- $\mathcal{T}_{admissible}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is an admissible labelling of $AF\}$
- $\mathcal{T}_{conflict-free}(AF) = \{\mathcal{L} \mid \mathcal{L}$ is a conflict-free labelling of $AF\}$

As illustrated in Figure 3, it holds that for any argumentation framework $AF = (Ar, def)$ that $\mathcal{T}_{stable}(AF) \subseteq \mathcal{T}_{semi-stable}(AF) \subseteq \mathcal{T}_{preferred}(AF) \subseteq \mathcal{T}_{complete}(AF)$ as well as that $\mathcal{T}_{grounded}(AF) \subseteq \mathcal{T}_{complete}(AF) \subseteq \mathcal{T}_{admissible}(AF) \subseteq \mathcal{T}_{conflict-free}(AF)$.

In [40], [5] and [28] it is shown how an admissible labelling corresponds to a position that can be defended in a rational discussion. Consider again the argumentation framework of Figure 1.

Proponent: "I accept argument D."

Opponent: "If you accept D, then you must reject D's defeater C. Based on which grounds?"

Proponent: "I reject C because I accept C's defeater B."

Opponent: "If you accept B, then you must reject B's defeater A. Based on which grounds?"

Proponent: "I reject A because I accept A's defeater B."

The rules of the discussion, formally described in [5], allow for the opponent to question the position of the proponent. If the proponent is able to answer these questions in a coherent way, the result will be an admissible labelling (arguments introduced by the proponent are labelled `in`, arguments introduced by the opponent are labelled `out`, and all unmentioned arguments are labelled `undec`).

In essence, an admissible labelling can be seen as a position in which one is able to provide reasons for each argument that one accepts and reasons for each argument that one rejects. If one is then also able to provide reasons for each argument one abstains from, the position will also be a complete labelling. In the remaining part of this paper, we mainly focus on admissible and complete labellings.

## 5 Applying Judgment Aggregation to Argumentation

Judgment aggregation studies the question of how various opinions of a group of agents can be put together to form an overall outcome of the entire group. The idea is that if each individual position is "reasonable", then the overall outcome should also be "reasonable". When judgment aggregation is done using classical logic, "reasonable" usually means that the set of formulas has to be consistent and closed under classical entailment. In formal argumentation, however, the wide variety of argumentation semantics means there are different ways of defining what is "reasonable".

Before introducing the operators for the aggregation of individual labellings of a given argumentation framework, we want to show how the classical discursive dilemma of judgment aggregation can be represented as an argumentation framework. We recall that in the discursive dilemma all individuals have to either accept or reject every issue in the agenda. Using argumentation, the individuals have the possibility to abstain on some arguments (this is represented by the `undec` labelling). Judgment aggregation with abstention has been explored in [17, 12], as we discuss in Section 10. Here, in order to capture the discursive dilemma, $\mathtt{undec}(\mathcal{L}) = \emptyset$.

Figure 4 shows the directed graph for the argumentation framework $AF = (Ar, def)$ where $Ar = \{p, q, r, \bar{p}, \bar{q}, \bar{r}\}$ (corresponding to the agenda of the Hortis Bank example), and $def = \{(p, \bar{p}), (\bar{p}, p), (q, \bar{q}), (\bar{q}, q), (\bar{p}, r), (\bar{q}, r), (r, \bar{r})\}$.

The labellings corresponding to the three individuals taking part in the bailout decision are indexed with the initials of the corresponding individual and are: $\mathcal{L}_{JP} = (\{p, \bar{q}, \bar{r}\}, \{q, \bar{p}, r\}, \emptyset)$, $\mathcal{L}_Y = (\{q, \bar{p}, \bar{r}\}, \{p, \bar{q}, r\}, \emptyset)$ and $\mathcal{L}_{JC} = (\{p, q, r\}, \{\bar{p}, \bar{q}, \bar{r}\}, \emptyset)$.

All three labellings $\mathcal{L}_{JP}$, $\mathcal{L}_Y$ and $\mathcal{L}_{JC}$ are admissible. However, if we apply majority voting argument-by-argument, we obtain an inadmissible collective labelling $\mathcal{L}_{Coll} = (\{p, q, \bar{r}\}, \{\bar{p}, \bar{q}, r\}, \emptyset)$, corresponding to the logically inconsistent collective judgment set in the discursive dilemma example. This shows that judgment aggregation problems can be mapped into an argumentation framework, though we do not claim here that this is possible for *all* judgment aggregation problems.

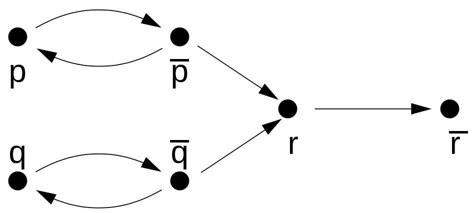


**Fig. 4** The discursive dilemma as an argumentation framework.

Given a set of individuals $N = \{1, \ldots, n\}$, we now need to define a general labellings aggregation operator $F_{AF}$ that assigns a collective labelling $\mathcal{L}_{Coll}$ to each profile $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$ of individual labellings. A note about notation is in order: while in judgment aggregation profiles are n-tuples of individual judgment sets, the aggregation operator $F_{AF}$ is defined on sets of individual labellings. The reason for using n-tuples is that several individuals may submit the same judgment sets. However, the motivation for our operators is to avoid situations in which any group member is forced to commit herself to a position that goes against his opinions. Therefore, cardinality considerations do not play a role in our approach, so we can take a profile to be a set of individual labellings.

**Definition 12** Let $\mathcal{L}abellings$ be the set of all possible labellings of argumentation framework $AF = (Ar, def)$. A general labellings aggregation operator is a function $F_{AF} : 2^{\mathcal{L}abellings} - \{\emptyset\} \rightarrow \mathcal{L}abellings$ such that $F_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}) = \mathcal{L}_{Coll}$.

We can now state the corresponding conditions of universal domain, collective rationality, anonymity, and independence for $F_{AF}$.

**Universal domain**: The domain of $F_{AF}$ is the set of all profiles of individual labellings belonging to semantics $\mathcal{T}_{admissible}$, $\mathcal{T}_{conflict-free}$ or $\mathcal{T}_{complete}$.

**Collective rationality**: $F_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is a labelling belonging to semantics $\mathcal{T}_{admissible}$, $\mathcal{T}_{conflict-free}$ or $\mathcal{T}_{complete}$.

**Anonymity**: Anonymity requires that the labelling submitted by an individual is indistinguishable from the labelling submitted by another individual. In our framework profiles are sets of individual labellings. Since in sets the correspondence between an individual and his submitted labelling is not defined, anonymity trivially holds.

**Independence**: For any $A \in Ar$ and any profiles $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$, $\{\mathcal{L}'_1, \ldots, \mathcal{L}'_n\}$ in the domain, if for some label $l(A) \in \{\texttt{in}, \texttt{out}, \texttt{undec}\}$, we have that $\forall j \in N, \mathcal{L}_j(A) = l(A) \leftrightarrow \mathcal{L}'_j(A) = l(A)$, then $F_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ and $F_{AF}(\{\mathcal{L}'_1, \ldots, \mathcal{L}'_n\})$ must assign the same label $l$ to $A$.

Universal domain and collective rationality capture respectively the types of input and output that an aggregation function can accept and must return. The corresponding conditions in judgment aggregation fix that permissible inputs are consistent and complete judgment sets, and the the aggregation function should return complete, consistent and deductively closed judgment sets (the only exception being incomplete judgment sets as in [17,12]). Argumentation is a flexible framework in which we can define aggregation operators for individual labellings and study their behavior under different semantics. This is reflected in the universal domain and collective rationality conditions above, where the inputs can be admissible, conflict-free or complete labellings (hence specifying to which semantics they belong).

The anonymity condition requires that all individuals count the same in the collective decision. Finally, the independence condition states that, the collective label of any argument $A$ should depend exclusively on the labellings that the individuals in a certain profile assign to $A$ and not on other arguments. Thus, if the individuals of two profiles assign the same labels to an argument $A$, then the collective labellings of $A$ assigned to the two profiles must also coincide.

Our general aim is to be flexible on the issue of argumentation semantics. Nevertheless, some semantics would be too restrictive for the aim of judgment aggregation. As an example, consider argumentation framework $AF = (Ar, \mathit{def})$ with $Ar = \{A, B\}$ and $\mathit{def} = \{(A, B), (B, A)\}$. If one, for instance, would apply preferred semantics, then there would be two preferred labellings: $\mathcal{L}_1 = (\{A\}, \{B\}, \emptyset)$ and $\mathcal{L}_2 = (\{B\}, \{A\}, \emptyset)$. However, aggregating these labellings to any preferred labelling would pose serious difficulties, because one would have to make an arbitrary choice which of these two preferred labellings would be selected as outcome labelling. Similar problems exist when one applies stable or semi-stable semantics.

Applying a unique status semantics (such as grounded [13], ideal [14] or eager [4]) would make judgment aggregation trivial, since it would follow that for each argumentation framework, only one reasonable position is possible. And if disagreement is not possible in the first place, then there is also effectively no need for any form of judgment aggregation. However, a unique status semantics can also be seen as over-restrictive, which is why in the current work we focus on multiple status semantics.

In the remaining part of this paper, we will focus on conflict-free labellings, admissible labellings and complete labellings. To see why this approach makes sense, consider again the example mentioned above. Labellings $\mathcal{L}_1$ and $\mathcal{L}_2$ are both complete, admissible and conflict-free labellings. It would be perfectly reasonable to define the aggregated labelling $\mathcal{L}_{Coll}$ as $(\emptyset, \emptyset, \{A, B\})$. That is, if agent 1 claims that $A$ should be accepted and $B$ should be rejected ($\mathcal{L}_1$) and agent 2 claims that $A$ should be rejected and $B$ should be accepted ($\mathcal{L}_2$) then it would be reasonable for the group as a whole to abstain from any explicit opinion on $A$ or $B$. Also, it holds that $\mathcal{L}_{Coll}$ is a complete, admissible and conflict-free labelling.

Since each stable, semi-stable, preferred, or grounded labelling is also a complete (and therefore admissible and conflict-free) labelling, we do not overly restrict ourselves by requiring every input labelling to be complete (or admissible or conflict-free). The general approach will be to make minimal assumptions about the input labellings when studying the properties of the proposed aggregation functions. Then, in some cases we will observe that once one starts to have additional requirements on the input-labellings, the aggregation function will satisfy more advanced properties as well.

## 6 The Sceptical Outcome

Conceptually, the idea of the sceptical aggregation is the following. All participants gather in a meeting which is aimed at constructing the sceptical outcome. The chair of the meeting then asks for each argument the opinion of the participants. If all participants unanimously think the argument should be accepted, then the argument is initially accepted. If all participants unanimously think the argument should be rejected, then the argument is initially rejected. In all other cases, the group as a whole does not have an explicit opinion about the argument. After all arguments have been treated this way, the meeting goes to the second phase. The chairman then reviews whether each accept or reject can still be justified from thus derived group outcome. An argument that is accepted without every defeater being rejected can no longer be accepted, since the position of the group as a whole does not provide sufficient justification for this. An argument that is rejected without a defeater that is accepted can no longer be rejected, since the position of the group as a whole does not provide sufficient justification for this. In each of these two cases, the group has to abstain from having an explicit opinion about the argument. This is an iterative process, since once one abstains from having an explicit opinion about a particular argument, it can cause explicit positions (accepts or rejects) of other arguments to be no longer justified. Thus, one has to go on until the group no longer has explicit opinions that are not justified. After this second phase is over, the result will be a position that is "smaller or equal" (less or equally committed) to each individual position of the participants. That is, each argument that is accepted by the group is also accepted by each individual participant, and each argument that is rejected by the group is also rejected by each individual participant. Furthermore, the group outcome is also self-justifying (each rejected argument has a defeater that is accepted, and each accepted argument has all its defeaters rejected). Within these constraints, the sceptical outcome is the most committed (biggest) position possible.

**Definition 13** Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two labellings of argumentation framework $AF = (Ar, def)$. We say that $\mathcal{L}_1$ is less or equally committed as $\mathcal{L}_2$ ($\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$) iff $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2)$ and $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2)$.

It can be observed that $\sqsubseteq$ is a partial order on labellings. That is:

1. $\mathcal{L} \sqsubseteq \mathcal{L}$ (reflexive)
2. if $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ and $\mathcal{L}_2 \sqsubseteq \mathcal{L}_3$ then $\mathcal{L}_1 \sqsubseteq \mathcal{L}_3$ (transitive)
3. if $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ and $\mathcal{L}_2 \sqsubseteq \mathcal{L}_1$ then $\mathcal{L}_1 = \mathcal{L}_2$ (anti-symetric)

Since "$\sqsubseteq$"is a partial order, we will sometimes talk about labelling $\mathcal{L}_2$ being "bigger or equal" to labelling $\mathcal{L}_1$, when $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$. Similarly, for certain sets of labellings we can also refer to a maximal or minimal, sometimes even to the biggest or smallest labelling, all with respect to the partial order defined by "$\sqsubseteq$".

Before continuing the formalization of the above described procedure, we first need to introduce some additional properties of labellings with respect to the order imposed by "$\sqsubseteq$".

**Theorem 5** *Let $\mathcal{L}$ be a labelling of argumentation framework $AF = (Ar, \mathit{def})$. The set of admissible labellings that are smaller or equal to $\mathcal{L}$ has a (unique) biggest element.*

*Proof* The all-undec labelling is admissible and smaller or equal to $\mathcal{L}$. Hence, there exists at least one admissible labelling that is smaller or equal to $\mathcal{L}$. Since we only consider finite argumentation frameworks, it follows that there exists only a finite number of admissible labellings smaller or equal to $\mathcal{L}$. It then directly follows that among the admissible labellings smaller or equal to $\mathcal{L}$, there exists at least one maximal one. We now prove that this maximal labelling is also unique. Let $\mathcal{L}_{ma_1}$ be a maximal labelling satisfying that $\mathcal{L}_{ma_1}$ is admissible and $\mathcal{L}_{ma_1} \sqsubseteq \mathcal{L}$. Also, let $\mathcal{L}_{ma_2}$ be a maximal labelling satisfying that $\mathcal{L}_{ma_2}$ is admissible and $\mathcal{L}_{ma_2} \sqsubseteq \mathcal{L}$. Now let $\mathcal{L}_{ma_3}$ be $\{(A, \mathtt{in}) \mid \mathcal{L}_{ma_1}(A) = \mathtt{in} \vee \mathcal{L}_{ma_2}(A) = \mathtt{in}\} \cup \{(A, \mathtt{out}) \mid \mathcal{L}_{ma_1}(A) = \mathtt{out} \vee \mathcal{L}_{ma_2}(A) = \mathtt{out}\} \cup \{(A, \mathtt{undec}) \mid \mathcal{L}_{ma_1}(A) = \mathtt{undec} \wedge \mathcal{L}_{ma_2}(A) = \mathtt{undec}\}$. It is relatively straightforward to observe that $\mathcal{L}_{ma_3}$ assigns *at least* one label to each argument. The fact that $\mathcal{L}_{ma_3}$ also assigns *at most* one label to each argument can be seen as follows. Suppose $\mathcal{L}_{ma_3}$ would assign more than one label to argument $A$. Then $A$ would be labelled both in and out. This means that either $\mathcal{L}_{ma_1}(A) = \mathtt{in}$ and $\mathcal{L}_{ma_2} = \mathtt{out}$, or that $\mathcal{L}_{ma_1}(A) = \mathtt{out}$ and $\mathcal{L}_{ma_2}(A) = \mathtt{in}$. From the fact that $\mathcal{L}_{ma_1}(A) \sqsubseteq \mathcal{L}$ and $\mathcal{L}_{ma_2}(A) \sqsubseteq \mathcal{L}$ it would then follow that $\mathcal{L}$ assigns both in and out to $A$, which is impossible. Therefore, $\mathcal{L}_{ma_3}$ assigns at most one label to each argument. This, together with the earlier observed fact that $\mathcal{L}_{ma_3}$ assigns at least one label to each argument, implies that $\mathcal{L}_{ma_3}$ assigns exactly one label to each argument. Hence, $\mathcal{L}_{ma_3}$ is a well-defined labelling of $AF$.
Also, it is relatively straightforward to observe that $\mathcal{L}_{ma_1} \sqsubseteq \mathcal{L}_{ma_3}$ and $\mathcal{L}_{ma_2} \sqsubseteq \mathcal{L}_{ma_3}$. We now prove that $\mathcal{L}_{ma_3}$ is an admissible labelling. If $\mathcal{L}_{ma_3}(A) = \mathtt{in}$ then we can see as follows that $\mathcal{L}_{ma_3}(B) = \mathtt{out}$ for every defeater $B$ of $A$. Let $B$ be an arbitrary defeater of $A$. From the fact that $\mathcal{L}_{ma_3}(A) = \mathtt{in}$ it follows that $\mathcal{L}_{ma_1}(A) = \mathtt{in}$ or $\mathcal{L}_{ma_2}(A) = \mathtt{in}$. Assume without loss of generality that $\mathcal{L}_{ma_1}(A) = \mathtt{in}$ (the case of $\mathcal{L}_{ma_2}(A) = \mathtt{in}$ goes similar). Then, from the fact that $\mathcal{L}_{ma_1}$ is an admissible labelling, it follows that $A$'s defeater $B$ is labelled out by $\mathcal{L}_{ma_1}$. This then implies that $B$ is also labelled out by $\mathcal{L}_{ma_3}$.
If $\mathcal{L}_{ma_3}(A) = \mathtt{out}$ then we can see as follows that $\mathcal{L}_{ma_3}(B) = \mathtt{in}$ for at least one defeater $B$ of $A$. From the fact that $\mathcal{L}_{ma_3}(A) = \mathtt{out}$ it follows that $\mathcal{L}_{ma_1}(A) = \mathtt{out}$ or $\mathcal{L}_{ma_2}(A) = \mathtt{out}$. Assume without loss of generality that $\mathcal{L}_{ma_1}(A) = \mathtt{out}$ (the case of $\mathcal{L}_{ma_2}(A) = \mathtt{out}$ goes similar). Then, from the fact that $\mathcal{L}_{ma_1}$ is an admissible labelling, it follows that $A$ has a defeater $B$ that is labelled in by $\mathcal{L}_{ma_1}$. This implies that $B$ is also labelled in by $\mathcal{L}_{ma_3}$.
From the fact that each argument that is labelled in by $\mathcal{L}_{ma_3}$ has all its defeaters labelled out by $\mathcal{L}_{ma_3}$, and that each argument that is labelled out by $\mathcal{L}_{ma_3}$ has at least one defeater that is labelled in by $\mathcal{L}_{ma_3}$, it follows that $\mathcal{L}_{ma_3}$ is an admissible labelling. The fact that $\mathcal{L}_{ma_3}$ is a labelling satisfying both admissibility and being less or equal to $\mathcal{L}$, together with the fact that $\mathcal{L}_{ma_1}$ is a *maximal* labelling satisfying both admissibility and being less or equal to $\mathcal{L}$ means that $\mathcal{L}_{ma_3}$ cannot be bigger than $\mathcal{L}_{ma_1}$. The fact that $\mathcal{L}_{ma_1} \sqsubseteq \mathcal{L}_{ma_3}$ then implies that $\mathcal{L}_{ma_1} = \mathcal{L}_{ma_3}$. The fact that $\mathcal{L}_{ma_3}$ is a labelling satisfying both admissibility and being less or equal to $\mathcal{L}$, together with the fact that $\mathcal{L}_{ma_2}$ is a *maximal* labelling satisfying both admissibility and being less or equal to $\mathcal{L}$ means that $\mathcal{L}_{ma_3}$ cannot be bigger than $\mathcal{L}_{ma_2}$. The fact that $\mathcal{L}_{ma_2} \sqsubseteq \mathcal{L}_{ma_3}$ then implies that $\mathcal{L}_{ma_2} = \mathcal{L}_{ma_3}$. From $\mathcal{L}_{ma_1} = \mathcal{L}_{ma_3}$ and that $\mathcal{L}_{ma_2} = \mathcal{L}_{ma_3}$ then implies that $\mathcal{L}_{ma_1} = \mathcal{L}_{ma_2}$. This means that the set of admissible

labellings that are smaller or equal to $\mathcal{L}$ has exactly one maximal element, which is then automatically also the biggest element.

It should be mentioned that Theorem 5 is one of the points where the labellings approach to argumentation differs from the traditional sets and extensions approach. In general it is *not* the case that for each set of arguments $\mathcal{A}rgs$, the set of all admissible subsets of $\mathcal{A}rgs$ has a biggest element. The possible existence of more than one preferred extension is a clear counterexample of this. The labellings approach to argumentation therefore satisfies a property that is not satisfied by the sets and extensions approach; a property that will be quite useful for current purposes. For instance, it allows for the well-definedness of the *down-admissible* labelling.

**Definition 14** Let $\mathcal{L}$ be a labelling of argumentation framework $AF = (Ar, def)$. The *down-admissible* labelling of $\mathcal{L}$ is the biggest element of the set of all admissible labellings that are smaller or equal to $\mathcal{L}$.

The contraction function is meant to relabel an argument from `in` or `out` to `undec`.

**Definition 15** Let $\mathcal{L}abellings$ be the set of all possible labellings of argumentation framework $AF = (Ar, def)$. The contraction function is a function $c_{AF} : \mathcal{L}abellings \times Ar \rightarrow \mathcal{L}abellings$ such that $c_{AF}(\mathcal{L}, A) = (\mathcal{L} - \{(A, \mathtt{in}), (A, \mathtt{out})\}) \cup \{(A, \mathtt{undec})\}$.

The idea of a contraction sequence is to keep applying contraction steps until the result is an admissible labelling.

**Definition 16** Let $\mathcal{L}$ be a labelling of argumentation framework $AF = (Ar, def)$. A *contraction sequence from* $\mathcal{L}$ is a list of labellings $[\mathcal{L}_1, \ldots, \mathcal{L}_m]$ such that:

1. $\mathcal{L}_1 = \mathcal{L}$,
2. for each $j \in \{1, \ldots, m\}$: $\mathcal{L}_{j+1} = c_{AF}(\mathcal{L}_j, A)$, where $A$ is an argument that is illegally `in` or illegally `out` in $\mathcal{L}_j$, and
3. $\mathcal{L}_m$ is a labelling without any illegal `in` or illegal `out`.

It should be mentioned that for every labelling $\mathcal{L}$, there exists at least one contraction sequence for it. This is because we consider only finite argumentation frameworks, from which it follows that the contraction sequence will terminate after a finite number of steps.

The idea of Lemma 2 and 3 below is that if an argument is illegally `in` (or illegally `out`) then it cannot be made legally `in` (or legally `out`) just by relabelling other arguments to `undec`.

**Lemma 2** *Let $\mathcal{L}$ be a labelling where $A$ is illegally* `in`*. It holds that in each labelling $\mathcal{L}' \sqsubseteq \mathcal{L}$: if $A$ is labelled* `in` *by $\mathcal{L}'$, then $A$ is illegally* `in` *in $\mathcal{L}'$.*

*Proof* The fact that $A$ is illegally `in` in $\mathcal{L}$ implies that $\mathcal{L}(A)$ has at least one defeater $B$ with $\mathcal{L}(B) \neq \mathtt{out}$. We distinguish two cases:

1. $\mathcal{L}(B) = \mathtt{in}$. The fact that $\mathcal{L}' \sqsubseteq \mathcal{L}$ implies that $\mathcal{L}'(B)$ is either `in` or `undec`. In both cases, $A$ has a defeater that is not labelled `out` by $\mathcal{L}'$, so $A$ is illegally `in` in $\mathcal{L}'$.
2. $\mathcal{L}(B) = \mathtt{undec}$. The fact that $\mathcal{L}' \sqsubseteq \mathcal{L}$ implies that $\mathcal{L}'(B) = \mathtt{undec}$, so $A$ has a defeater that is not labelled `out` by $\mathcal{L}'$, so $A$ is illegally `in` in $\mathcal{L}'$.

**Lemma 3** *Let $\mathcal{L}$ be a labelling where $A$ is illegally* `out`*. It holds that in each labelling $\mathcal{L}' \sqsubseteq \mathcal{L}$: if $A$ is labelled* `out` *by $\mathcal{L}'$, then $A$ is illegally* `out` *in $\mathcal{L}'$.*

*Proof* The fact that $A$ is illegally `out` in $\mathcal{L}$ implies that $\mathcal{L}(A)$ does not have a defeater that is labelled `in` by $\mathcal{L}$. That is, each defeater of $A$ is either labelled `out` or `undec` by $\mathcal{L}$. Let $B$ be an arbitrary defeater of $A$. We distinguish two cases:

1. $\mathcal{L}(B) = \texttt{out}$. The fact that $\mathcal{L}' \sqsubseteq \mathcal{L}$ implies that $\mathcal{L}'(B)$ is either `out` or `undec`. In both cases, $\mathcal{L}'(B) \neq \texttt{in}$.
2. $\mathcal{L}(B) = \texttt{undec}$. The fact that $\mathcal{L}' \sqsubseteq \mathcal{L}$ implies that $\mathcal{L}'(B) = \texttt{undec}$, so $\mathcal{L}'(B) \neq \texttt{in}$.

In both cases, we have that $\mathcal{L}'(B) \neq \texttt{in}$. Since this holds for any arbitrary defeater $B$ of $A$, it follows that $A$ does not have any defeater that is labelled `in` by $\mathcal{L}'$, so $A$ is illegally `out` in $\mathcal{L}'$.

The next thing to prove is that a contraction sequence calculates the down-admisssible labelling.

**Theorem 6** *Let $\mathcal{L}$ be a labelling of argumentation framework $AF = (Ar, def)$, let $\mathcal{L}_{da}$ be the down-admissible labelling of $\mathcal{L}$ and let $[\mathcal{L}_1, \ldots, \mathcal{L}_m]$ be a contraction sequence from $\mathcal{L}$. It holds that $\mathcal{L}_m = \mathcal{L}_{da}$.*

*Proof* We first prove that for each $\mathcal{L}_j$ ($j \in \{1, \ldots, m\}$) in the contraction sequence it holds that $\mathcal{L}_j \sqsubseteq \mathcal{L}$. We prove this by induction over $j$.

basis: From the fact that "$\sqsubseteq$" is reflexive, it follows that $\mathcal{L} \sqsubseteq \mathcal{L}$. From the fact that $\mathcal{L}_1 = \mathcal{L}$ it then follows that $\mathcal{L}_1 \sqsubseteq \mathcal{L}$.

step: Suppose that $\mathcal{L}_j \sqsubseteq \mathcal{L}$. From the fact that $\mathcal{L}_{j+1} \sqsubseteq \mathcal{L}_j$ it then follows that $\mathcal{L}_{j+1} \sqsubseteq \mathcal{L}$, since "$\sqsubseteq$" is transitive.

The next thing to be proved is that for each $\mathcal{L}_j$ ($j \in \{1, \ldots, m\}$) in the contraction sequence, it holds that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$. We prove this by induction over $j$.

basis: From Definition 14 it immediately follows that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}$. From the fact that $\mathcal{L}_1 = \mathcal{L}$ it then follows that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_1$.

step: Suppose that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$ with $j \leq m - 1$. We now prove that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_{j+1}$. Let $A$ be the argument such that $\mathcal{L}_{j+1} = c_{AF}(\mathcal{L}_j, A)$. We distinguish two possibilities.

1. $A$ is illegally `in` in $\mathcal{L}_j$. From the fact that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$ (induction hypothesis) it follows that (Lemma 2) if $A$ is labelled `in` by $\mathcal{L}_{da}$, $A$ would be illegally `in` in $\mathcal{L}_{da}$. Since $\mathcal{L}_{da}$ is an admissible labelling, it does not have an argument that is illegally `in`. Therefore, $A$ cannot be labelled `in` by $\mathcal{L}_{da}$. $A$ also cannot be labelled `out` by $\mathcal{L}_{da}$ (otherwise it would not hold that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$) so $A$ has to be labelled `undec` by $\mathcal{L}_{da}$.
2. $A$ is illegally `out` in $\mathcal{L}_j$. From the fact that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$ (induction hypothesis) it follows that (Lemma 3) if $A$ is labelled `out` by $\mathcal{L}_{da}$, $A$ would be illegally `out` in $\mathcal{L}_{da}$. Since $\mathcal{L}_{da}$ is an admissible labelling, it does not have an argument that is illegally `out`. Therefore, $A$ cannot be labelled `out` by $\mathcal{L}_{da}$. $A$ also cannot be labelled `in` by $\mathcal{L}_{da}$ (otherwise it would not hold that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$) so $A$ has to be labelled `undec` by $\mathcal{L}_{da}$.

Using the thus derived fact that $A$ is labelled `undec` by $\mathcal{L}_{da}$, we now continue with the prove that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_{j+1}$

- Let $B$ be an argument that is labelled `in` by $\mathcal{L}_{da}$. Then $B \neq A$. From the fact that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$ it follows that $\mathcal{L}_j(B) = \texttt{in}$. From the fact that $\mathcal{L}_{j+1}(B) = \mathcal{L}_j(B)$ for each $B \neq A$ it follows that $\mathcal{L}_{j+1}(B) = \texttt{in}$.

– Let $B$ be an argument that is labelled out by $\mathcal{L}_{da}$. Then $B \neq A$. From the fact that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$ it follows that $\mathcal{L}_j(B) = \texttt{out}$. From the fact that $\mathcal{L}_{j+1}(B) = \mathcal{L}_j(B)$ for each $B \neq A$ it follows that $\mathcal{L}_{j+1}(B) = \texttt{out}$.

From the thus derived fact that each argument labelled in by $\mathcal{L}_{da}$ is also labelled in by $\mathcal{L}_{j+1}$, and that each argument labelled out by $\mathcal{L}_{da}$ is also labelled out by $\mathcal{L}_{j+1}$ it follows that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_{j+1}$.

We have now proved that for each $\mathcal{L}_j$ ($j \in \{1, \ldots, m\}$) it holds that $\mathcal{L}_j \sqsubseteq \mathcal{L}$ and it holds that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$. In particular, this also holds for $\mathcal{L}_m$. Furthermore, $\mathcal{L}_m$ does not contain an illegal in or illegal out, and is therefore an admissible labelling. Hence, $\mathcal{L}_m$ is an admissible labelling such that $\mathcal{L}_m \sqsubseteq \mathcal{L}$. From the fact that $\mathcal{L}_{da}$ is the unique biggest admissible labelling such that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}$, it then directly follows that $\mathcal{L}_m \sqsubseteq \mathcal{L}_{da}$. However, it also holds that $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_m$ (since for each $j \in \{1, \ldots, m\}$: $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_j$). From the fact that $\mathcal{L}_m \sqsubseteq \mathcal{L}_{da}$ and $\mathcal{L}_{da} \sqsubseteq \mathcal{L}_m$ it then directly follows that $\mathcal{L}_m = \mathcal{L}_{da}$, since "$\sqsubseteq$" is a partial order satisfying anti-symmetry.

Now that the preliminary theory has been treated, we can return to the main question of how to formalize the procedure of the meeting described in the beginning of the section. The first phase of the meeting, which yields the initial outcome ($\mathcal{L}_{sio}$), can be described as follows.

**Definition 17** Let $\mathcal{L}abellings$ be the set of all possible labellings of argumentation framework $AF = (Ar, def)$. The sceptical initial aggregation operator is a function $sio_{AF} : 2^{\mathcal{L}abellings} - \{\emptyset\} \rightarrow \mathcal{L}abellings$ such that $sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}) =$
$\{(A, \texttt{in}) \mid \forall i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \texttt{in}\} \cup$
$\{(A, \texttt{out}) \mid \forall i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \texttt{out}\} \cup$
$\{(A, \texttt{undec}) \mid \exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) \neq \texttt{in} \wedge \exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) \neq \texttt{out}\}$.

We first observe that $sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is well-defined in the sense that it assigns exactly one label to each argument. We will sometimes write $\mathcal{L}_1 \sqcap \mathcal{L}_2$ as an abbreviation for $sio_{AF}(\{\mathcal{L}_1, \mathcal{L}_2\})$.

**Lemma 4** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ ($n \geq 1$) be admissible labellings of argumentation framework $AF = (Ar, def)$ and $\mathcal{L}_{sio} = sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ be the associated sceptical initial labelling. It holds that $\mathcal{L}_{sio} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$).*

*Proof* This follows directly from Definition 17.

**Proposition 1** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ be conflict-free labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{sio}$ be $sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{sio}$ is also a conflict-free labelling.*

It is interesting to notice that if the semantics merely requires conflict-freeness, then $sio_{AF}$ satisfies *all* the standard judgment aggregation conditions. This means that $sio_{AF}$ satisfies universal domain, collective rationality, anonymity and independence.

However, if the semantics requires the stronger condition of admissibility, then $sio_{AF}$ no longer satisfies coherence. As an example, consider the argumentation framework $AF = (Ar, def)$ with $Ar = \{A, B, C, D\}$ and $def = \{(A, B), (B, A), (A, C), (B, C), (C, D)\}$. This argumentation framework is shown in Figure 5. Both $\mathcal{L}_1 = (\{A, D\},$

$\{B, C\}, \emptyset)$ and $\mathcal{L}_2 = (\{B, D\}, \{A, C\}, \emptyset)$ are admissible labellings, but $sio_{AF}(\{\mathcal{L}_1, \mathcal{L}_2\}) = (\{D\}, \{C\}, \{A, B\})$, which is not admissible. We will now define a new aggregation operator ($so_{AF}$) which does yield admissible labellings. In exchange, we will have to give up the property of unanimity. We will come back to this point at the end of this section.

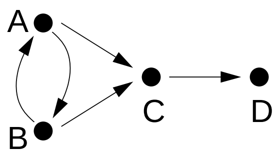


**Fig. 5** The sceptical aggregation operator does not satisfy unanimity.

The result of the second phase of the meeting can then be characterized as follows. Notice that this result can be calculated using a contraction sequence.

**Definition 18** Let $\mathcal{L}abellings$ be the set of all labellings of argumentation framework $AF = (Ar, def)$. The *sceptical aggregation operator* is a function $so_{AF} : 2^{\mathcal{L}abellings} - \{\emptyset\} \rightarrow \mathcal{L}abellings$ such that $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is the down-admissible labelling of $sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

One positive feature of the thus described sceptical aggregation is that if an argument is accepted (or rejected) by the group outcome, it is also accepted (or rejected) by each individual member of the group. Hence, if a member needs to explain (perhaps in public) why his group accepts or rejects a particular argument, he will be able to do so without having to go against his own private opinions. This is stated by the following theorem.

**Theorem 7** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ ($n \geq 1$) be labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{so}$ be $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{so}$ is the biggest admissible labelling such that for every $i \in \{1, \ldots, n\}$: $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_i$.*

*Proof* We first prove that for each $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Let $\mathcal{L}_{sio} = sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. From Definition 18 it follows that $\mathcal{L}_{so}$ is the down-admissible labelling of $\mathcal{L}_{sio}$. Therefore, $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{sio}$. Lemma 4 states that $\mathcal{L}_{sio} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Therefore, $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$).
Now that we know that $\mathcal{L}_{so}$ is an admissible labelling such that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$), the next thing to show is that $\mathcal{L}_{so}$ is also the *biggest* admissible labelling such that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Let $\mathcal{L}'$ be an admissible labelling with $\mathcal{L}' \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). From the fact that $\mathcal{L}' \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$) it follows that $\mathcal{L}' \sqsubseteq \mathcal{L}_{sio}$. This can be seen as follows.

- Let $A \in \mathtt{in}(\mathcal{L}')$. Then from $\mathcal{L}' \sqsubseteq \mathcal{L}_i$ it follows that $A \in \mathtt{in}(\mathcal{L}_i)$ (for each $i \in \{1, \ldots, n\}$). From Definition 17 it then follows that $A \in \mathtt{in}(\mathcal{L}_{sio})$.
- Let $A \in \mathtt{out}(\mathcal{L}')$. Then from $\mathcal{L}' \sqsubseteq \mathcal{L}_i$ it follows that $A \in \mathtt{out}(\mathcal{L}_i)$ (for each $i \in \{1, \ldots, n\}$). From Definition 17 it then follows that $A \in \mathtt{out}(\mathcal{L}_{sio})$.

From the fact that $\mathcal{L}'$ is an admissible labelling that is smaller or equal to $\mathcal{L}_{sio}$, and the fact that $\mathcal{L}_{so}$ is the *biggest* admissible labelling that is smaller or equal to $\mathcal{L}_{sio}$, it then follows that $\mathcal{L}' \sqsubseteq \mathcal{L}_{so}$.

It is interesting that we can also use $so_{AF}$ if the rationality requirement is changed from "admissible" to "complete". That is, if the input labellings are complete, then the aggregated labelling will automatically also be complete.

**Theorem 8** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ $(n \geq 1)$ be complete labellings of argumentation framework $AF = (Ar, def)$. It holds that the sceptical outcome labelling $\mathcal{L}_{so} = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is a complete labelling.*

*Proof* From the fact that $\mathcal{L}$ is an admissible labelling, it immediatelly follows that every argument labelled `in` is legally `in`, and that every argument labelled `out` is legally `out`. So the only thing still to be proved for $\mathcal{L}_{so}$ to be a complete labelling is that every argument labelled `undec` is legally `undec`. Let $A$ be an arbitrary argument such that $\mathcal{L}(A) =$ `undec`. Suppose $A$ would be illegally `undec`. We distinguish two possibilities.

1. $A$ has a defeater that is labelled `in` by $\mathcal{L}_{so}$. But then $\mathcal{L}_{so}$ would not be maximal because it would be possible to relabel $A$ to `out` (since $A$ is also labelled `out` by every $\mathcal{L}_i$ $(i \in \{1, \ldots, n\})$).
2. $A$ has all its defeaters labelled `out` by $\mathcal{L}_{so}$. But then $\mathcal{L}_{so}$ would not be maximal because it would be possible to relabel $A$ to `in` (since $A$ is also labelled `in` by every $\mathcal{L}_i$ $(i \in \{1, \ldots, n\})$).

In both cases, the result is a contradiction. Therefore, $A$ cannot be illegally `undec` in $\mathcal{L}_{so}$.

Before we move to the credulous outcome, we want to return to the observation that the aggregation operator $sio_{AF}$ preserves unanimity at the expense of losing coherence when coherence is interpreted as admissibility. If collective rationality requires the group labelling to be admissible, $sio_{AF}$ cannot guarantee to output an admissible labelling even under admissible inputs. We have then defined the operator $so_{AF}$ that guarantees admissible outputs but does not preserve unanimity.

The argumentation system of Figure 5 is an example of what in the literature is known as *floating conclusions* [27], i.e. statements that are supported in each extension but by different arguments. In default logic, when a theory has multiple extensions, the sceptical approach states that a conclusion should be endorsed only if it is contained in every extension. Horty [19] considers the sceptical policy applied to multiple argument extensions and questions it. The reason for not accepting floating conclusions is that they are precarious:

> The point is not that floating conclusions might be wrong; any conclusion drawn through defeasible reasoning might be wrong. The point is that a statement supported only as floating conclusion seems to be less secure than the same statement when it is uniformly supported by a common argument. ([19], p.65)

The instance shown in Figure 5 and the discussion above about floating conclusions reminds the *Paretian dilemma*, a variation of the discursive dilemma presented by Nehring in [30]. The Pareto criterion says that, if every individual prefers one collective outcome over another, that outcome should be socially selected. In [30], a three-judges court has:

> to decide whether a defendant has to pay damages to the plaintiff. Legal doctrine requires that damages are due if and only if the following three premises are established: 1) the defendant had a duty to take care, 2) the defendant behaved negligently, 3) his negligence caused damage to the plaintiff. [p.1]

| | $p =$ Duty | $q =$ Negligence | $r =$ Causation | $s =$ Damages |
|---|---|---|---|---|
| Judge 1 | 1 | 1 | 0 | 0 |
| Judge 2 | 0 | 1 | 1 | 0 |
| Judge 3 | 1 | 0 | 1 | 0 |
| Majority | 1 | 1 | 1 | 0 |

**Table 2** Paretian dilemma. $p =$ Duty, $q =$ Negligence, $r =$ Causation, $s = (p \wedge q \wedge r) =$ Damages.

Suppose that the three judges vote as in Table 2.

The table above shows that, when aggregating issue-by-issue, the reason-based collective judgment (the judgments on $p$, $q$ and $r$) can conflict with the Pareto criterion. A majority agrees on each $p$, $q$ and $r$, supporting the decision that the defendant has to pay damages to the plaintiff, contrary to the unanimous judgment that damages should be denied. Like $sio_{AF}$, proposition-wise majority voting preserves unanimity but cannot ensure a consistent social outcome.

Nehring proves that, all well-behaved (i.e. anonymous or non-dictatorial) aggregation rules are prone to the Paretian dilemma, tantamount to saying that no reason-based group decision can be guaranteed. How negative is this result? In [30] it is argued that when the reasons are epistemically independent "all relevant information about the outcome decision is contained in the agents' premise judgments. [...] Indeed, under epistemic independence of premises it is easy to understand how a group aggregation rule can *rightly* override a unanimous outcome judgment" [*ibid.* p.36]. Furthermore, the normative force of the Pareto criterion depends on the type of social decision. The Pareto criterion should be ensured when the individuals have a *shared self-interest* in the final outcome, whereas it can be relaxed when they *share responsibility* for the decision. Judicial decisions are clear instances of shared responsibility situations, while other group decisions may be self-interest driven. Nehring's analysis concludes that the Pareto criterion and reason-based group decisions are two principles that may come into conflict. However, such conflict does not mean that one of these two principles is ill-founded.

## 7 The Credulous Outcome

The idea of the credulous outcome is that it is not too bad for a group opinion to accept or reject arguments that are not accepted or rejected by each individual member, as long as each individual member's private opinions are not directly against the group outcome. That is, if the group outcome is that an argument should be accepted, then each member either believes that this argument has to be accepted or simply has no opinion about it. Similarly, if the group outcome is that the argument has to be rejected, then each member either believes that this argument should be rejected or has no opinion about it.

The first thing to define is what it means for two labellings to be *compatible*.

**Definition 19** Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two labellings of argumentation framework $(Ar, def)$. We say that $\mathcal{L}_1$ is *compatible* with $\mathcal{L}_2$, denoted as $\mathcal{L}_1 \approx \mathcal{L}_2$, iff $\mathtt{in}(\mathcal{L}_1) \cap \mathtt{out}(\mathcal{L}_2) = \emptyset$ and $\mathtt{out}(\mathcal{L}_1) \cap \mathtt{in}(\mathcal{L}_2) = \emptyset$.

It can be shown that the compatibility relation "$\approx$" satisfies reflexity ($\mathcal{L} \approx \mathcal{L}$) and symmetry (if $\mathcal{L}_1 \approx \mathcal{L}_2$ then $\mathcal{L}_2 \approx \mathcal{L}_1$) but not transitivity (from $\mathcal{L}_1 \approx \mathcal{L}_2$ and $\mathcal{L}_2 \approx \mathcal{L}_3$

it does not follow that $\mathcal{L}_1 \approx \mathcal{L}_3$). As an example of why $\approx$ does not satisfy transitivity, consider the argumentation framework of Figure 5, with $\mathcal{L}_1 = (\{A, D\}, \{B, C\}, \emptyset)$, $\mathcal{L}_2 = (\emptyset, \emptyset, \{A, B, C, D\})$ and $\mathcal{L}_3 = (\{B, D\}, \{A, C\}, \emptyset)$. Here, $\mathcal{L}_1 \approx \mathcal{L}_2$ and $\mathcal{L}_2 \approx \mathcal{L}_3$ but $\mathcal{L}_1 \not\approx \mathcal{L}_3$.

It is also possible to give a different equivalent description of "$\approx$".

**Proposition 2** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be labellings of argumentation framework $(Ar, def)$. It holds that $\mathcal{L}_1 \approx \mathcal{L}_2$ iff $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$ and $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.*

For some of the proofs, we will use the results of Proposition 2 without explicitly referring to it. That is, to prove that $\mathcal{L}_1 \approx \mathcal{L}_2$ we simply prove that $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$ and $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.

It holds that "$\sqsubseteq$" is a stronger condition than "$\approx$". That is, if $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ then $\mathcal{L}_1 \approx \mathcal{L}_2$.

Conceptually, the process of constructing the credulous outcome is done as follows. All participants gather in a meeting. The chair of the meeting then asks for each argument the opinion of the participants. If at least one participant accepts the argument, and nobody explicitly rejects it, then the argument is initially accepted. If at least one participant rejects the argument and nobody explicitly accepts it, then the argument is initially rejected. After all arguments have been treated this way, the meeting goes to the second phase. The chairman then reviews whether each accept or reject can still be justified from thus derived group outcome. Each argument that is accepted or rejected without a justification can no longer be accepted or rejected, so the group has to abstain from having an explicit opinion about it. This is an iterative process, since once one abstains from having an explicit opinion about a particular argument, it can cause explicit positions (accepts or rejects) of other arguments to be no longer justified. Thus, one has to go on until the group no longer has explicit opinions that are not justified.

**Definition 20** Let $\mathcal{L}abellings$ be the set of all possible labellings of argumentation framework $AF = (Ar, def)$. The *credulous initial aggregation operator* is a function $cio_{AF} : 2^{\mathcal{L}abellings} - \{\emptyset\} \rightarrow \mathcal{L}abellings$ such that $cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}) =$
$\{(A, \mathtt{in}) \mid \exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{in} \wedge \neg\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{out}\} \cup$
$\{(A, \mathtt{out}) \mid \exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{out} \wedge \neg\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{in}\} \cup$
$\{(A, \mathtt{undec}) \mid \forall i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{undec} \vee (\exists i \in \{1, \ldots, n\} : \mathcal{L}_i = \mathtt{in} \wedge \exists i \in \{1, \ldots, n\} : \mathcal{L}_i = \mathtt{out})\}$.

We will sometimes write $\mathcal{L}_1 \sqcup \mathcal{L}_2$ as an abbreviation for $cio_{AF}(\{\mathcal{L}_1, \mathcal{L}_2\})$.

**Theorem 9** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ be labellings of argumentation framework $AF = (Ar, def)$ and let $\mathcal{L}_{cio} = cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ be the credulous initial labelling. It holds that $\forall i \in \{1, \ldots, n\} : \mathcal{L}_{cio} \approx \mathcal{L}_i$.*

*Proof* This follows directly from Definition 19 and Definition 20.

**Definition 21** Let $\mathcal{A}dm\mathcal{L}abellings$ be the set of all admissible labellings of argumentation framework $AF = (Ar, def)$. The credulous aggregation operator is a function $co_{AF} : 2^{\mathcal{A}dm\mathcal{L}abellings} - \{\emptyset\} \rightarrow \mathcal{A}dm\mathcal{L}abellings$ such that $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is the down-admissible labelling of $cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

It holds that the credulous outcome labelling is compatible with each of the participants' individual labelling. For the proof, we first need the following lemma.

**Lemma 5** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ be labellings of argumentation framework $(Ar, def)$. It holds that if $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ and $\mathcal{L}_2 \approx \mathcal{L}_3$ then $\mathcal{L}_1 \approx \mathcal{L}_3$.*

*Proof* We need to prove that $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_3) \cup \mathtt{undec}(\mathcal{L}_3)$ and that $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_3) \cup \mathtt{undec}(\mathcal{L}_3)$.

- Let $A \in \mathtt{in}(\mathcal{L}_1)$. Then, from the fact that $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ it follows that $A \in \mathtt{in}(\mathcal{L}_2)$. From the fact that $\mathcal{L}_2 \approx \mathcal{L}_3$ it then follows that $A \in \mathtt{in}(\mathcal{L}_3) \cup \mathtt{undec}(\mathcal{L}_3)$.
- Let $A \in \mathtt{out}(\mathcal{L}_1)$. Then, from the fact that $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ it follows that $A \in \mathtt{out}(\mathcal{L}_2)$. From the fact that $\mathcal{L}_2 \approx \mathcal{L}_3$ it then follows that $A \in \mathtt{out}(\mathcal{L}_3) \cup \mathtt{undec}(\mathcal{L}_3)$.

**Theorem 10** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ ($n \geq 1$) be admissible labellings of argumentation framework $AF = (Ar, def)$, let $\mathcal{L}_{cio} = cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ be the associated credulous initial labelling, and let $\mathcal{L}_{co} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ be the associated credulous outcome labelling. It holds that $\mathcal{L}_{co} \approx \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$).*

*Proof* From Definition 21 and Definition 14, it immediately follows that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_{cio}$. Also, from Definition 19 and 20, it immediately follows that $\mathcal{L}_{cio} \approx \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Lemma 5 then allows us to derive that $\mathcal{L}_{co} \approx \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$).

Although for the sceptical aggregation, it holds that the outcome is a complete labelling when the individual input labellings are complete (Theorem 8), this does not hold for the credulous aggregation. That is, even if the input labellings $\mathcal{L}_1, \ldots, \mathcal{L}_n$ are complete, the credulous outcome labelling $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ does not need to be complete. Take the example of $(Ar, def)$ where $Ar = \{A, B, C, D, E, F\}$ and $def = \{(A, B), (B, A), (C, D), (D, C), (B, E), (D, E), (E, F)\}$. This argumentation framework is shown in Figure 6. Let $\mathcal{L}_1 = (\{A\}, \{B\}, \{C, D, E, F\})$ and $\mathcal{L}_2 = (\{C\}, \{D\}, \{A, B, E, F\})$. Both $\mathcal{L}_1$ and $\mathcal{L}_2$ are complete, but the credulous outcome labelling $\mathcal{L}_{co}$ is $(\{A, C\}, \{B, D\}, \{E, F\})$ which is not complete.

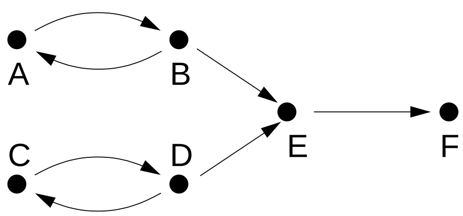


**Fig. 6** The credulous aggregation operator may not yield a complete labelling.

## 8 The Super Credulous Outcome

It is even possible to define a reasonable outcome that is even more credulous than the credulous outcome; we call this new outcome *super credulous*. The idea is to start with the credulous outcome and then "expand" it, that is, to make it bigger by relabelling illegal `undec`s to `in`s and `out`s. Conceptually, what happens is the following. After the meeting has determined the credulous outcome, the chairman holds a third phase, in which he tries to see whether the opinions can be extended. For this, he takes a look at each argument where the meeting does not yet have an opinion about (that is accepted neither rejected collectively). If there are sufficient grounds to accept it (that is, if all its defeaters are already rejected) then the group will accept it. If there are sufficient grounds to reject it (that is, if it has a defeater that is accepted) then it will be rejected. This goes on until each argument that can be accepted is accepted and each argument that can be rejected is rejected. The result will be the super credulous outcome.

**Definition 22** Let $\mathcal{L}abellings$ be the set of all possible labellings of argumentation framework $AF = (Ar, def)$. The *expansion function* $e_{AF} : \mathcal{L}abellings \times Ar \rightarrow \mathcal{L}abellings$ is defined as follows.

- $e_{AF}(\mathcal{L}, A) = (\mathcal{L} - \{(A, \texttt{undec})\}) \cup \{(A, \texttt{in})\}$ if $A$ is illegally `undec` in $\mathcal{L}$ and all its defeaters are `out`,
- $e_{AF}(\mathcal{L}, A) = (\mathcal{L} - \{(A, \texttt{undec})\}) \cup \{(A, \texttt{out})\}$ if $A$ is illegally `undec` in $\mathcal{L}$ and it has a defeater that is `in`, and
- $e_{AF}(\mathcal{L}, A) = \mathcal{L}$ in all other cases.

**Definition 23** Let $\mathcal{L}_a$ be an admissible labelling of argumentation framework $AF = (Ar, def)$. An *expansion sequence* from $\mathcal{L}_a$ is a list of labellings $[\mathcal{L}_1, \ldots, \mathcal{L}_m]$ $(m \geq 1)$ such that:

1. $\mathcal{L}_1 = \mathcal{L}_a$,
2. for each $j \in \{1, \ldots, m-1\}$: $\mathcal{L}_{j+1} = e_{AF}(\mathcal{L}_j, A)$, where $A$ is an argument that is illegally `undec` in $\mathcal{L}_j$, and
3. $\mathcal{L}_m$ is a labelling that does not have any illegal `undec`.

It should be mentioned that for any admissible labelling $\mathcal{L}_a$ of $AF$, there exists at least one super credulous expansion sequence. This is because there are only a finite number of arguments to do an expansion step on.

**Lemma 6** *Let $\mathcal{L}_a$ be an admissible labelling of argumentation framework $AF = (Ar, def)$, and let $[\mathcal{L}_1, \ldots, \mathcal{L}_m]$ be an expansion sequence from $\mathcal{L}_a$. It holds that $\mathcal{L}_m$ is a complete labelling with $\mathcal{L}_a \sqsubseteq \mathcal{L}_m$.*

*Proof* We first prove, by induction, that every $\mathcal{L}_j$ $(j \in \{1, \ldots, m\})$ is bigger or equal to $\mathcal{L}_a$.

basis: From the fact that "$\sqsubseteq$" is reflexive, it follows that $\mathcal{L}_a \sqsubseteq \mathcal{L}_a$. From the fact that $\mathcal{L}_1 = \mathcal{L}_a$ it then follows that $\mathcal{L}_a \sqsubseteq \mathcal{L}_1$.
step: Suppose that $\mathcal{L}_a \sqsubseteq \mathcal{L}_j$ $(j \in \{1, \ldots, m-1\})$. From the fact that $\mathcal{L}_j \sqsubseteq \mathcal{L}_{j+1}$ it follows that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{j+1}$ since "$\sqsubseteq$" is transitive.

From the thus proved fact that every $\mathcal{L}_a \sqsubseteq \mathcal{L}_j$ $(j \in \{1, \ldots, m\})$, it then directly follows that $\mathcal{L}_a \sqsubseteq \mathcal{L}_m$.
We now prove, by induction, that every $\mathcal{L}_j$ $(j \in \{1, \ldots, m\})$ is an admissible labelling.

basis: $\mathcal{L}_1 = \mathcal{L}_a$ and $\mathcal{L}_a$ is an admissible labelling.
step: Suppose $\mathcal{L}_j$ is an admissible labelling. Let $A$ be the argument on which the expansion step to $\mathcal{L}_{j+1}$ took place. That is, $\mathcal{L}_{j+1} = e_{AF}(\mathcal{L}_j, A)$. We now prove that $\mathcal{L}_{j+1}$ is an admissible labelling. Let $B$ be an argument labelled `in` by $\mathcal{L}_{j+1}$. We distinguish two cases.

1. $B \neq A$. Then $B$ is labelled `in` by $\mathcal{L}_j$. Since $\mathcal{L}_j$ is an admissible labelling (induction hypothesis), it follows that all defeaters of $B$ are labelled `out` by $\mathcal{L}_j$.
2. $B = A$. Then, from the fact that $B$ qualified for an expension step resulting in $B$ being labelled `in`, it follows (Definition 22) that all defeaters of $B$ are labelled `out` by $\mathcal{L}_j$.

In both cases, it holds that all defeaters of $B$ are labelled `out` by $\mathcal{L}_j$. Since $\texttt{out}(\mathcal{L}_j) \subseteq \texttt{out}(\mathcal{L}_{j+1})$ it follows that all defeaters of $B$ are also labelled `out` by $\mathcal{L}_{j+1}$. Hence, $B$ is legally `in` in $\mathcal{L}_{j+1}$.
Now that we have proven that each argument labelled `in` by $\mathcal{L}_{j+1}$ is legally `in`, the next thing to prove is that each argument labelled `out` by $\mathcal{L}_{j+1}$ is legally `out`. Let $B$ be an argument labelled `out` by $\mathcal{L}_{j+1}$. We distinguish two cases.

1. $B \neq A$. Then $B$ is labelled `out` by $\mathcal{L}_j$. Since $\mathcal{L}_j$ is an admissible labelling (induction hypothesis), it follows that $B$ has a defeater (say $C$) that is labelled `in` by $\mathcal{L}_j$.
2. $B = A$. Then, from the fact that $B$ qualified for an expension step resulting in $B$ being labelled `out`, it follows (Definition 22) that $B$ has a defeater (say $C$) that is labelled `in` by $\mathcal{L}_j$.

In both cases, it holds that $B$ has a defeater $C$ that is labelled `in` by $\mathcal{L}_j$. Since $\texttt{in}(\mathcal{L}_j) \subseteq \texttt{in}(\mathcal{L}_{j+1})$ it follows that $B$'s defeater $C$ is also labelled `in` by $\mathcal{L}_{j+1}$. Hence, $B$ is legally `out` in $\mathcal{L}_{j+1}$.
From the thus derived facts that each argument labelled `in` by $\mathcal{L}_{j+1}$ is legally `in`, and that each argument labelled `out` by $\mathcal{L}_{j+1}$ is legally `out` it follows that $\mathcal{L}_{j+1}$ is an admissible labelling.

From the thus proved fact that for each $j \in \{1, \ldots, m\}$ it holds that $\mathcal{L}_j$ is an admissible labelling, it directly follows that $\mathcal{L}_m$ is an admissible labelling. This, and the fact that $\mathcal{L}_m$ does not have any illegal `undec`, implies that $\mathcal{L}_m$ is a complete labelling.

**Theorem 11** *Let $\mathcal{L}_a$ be an admissible labelling of argumentation framework $AF = (Ar, def)$. The set of complete labellings that are bigger or equal to $\mathcal{L}_a$ has a (unique) smallest element.*

*Proof* Lemma 6 implies that there exists at least one complete labelling (namely $\mathcal{L}_m$) that is bigger or equal to $\mathcal{L}_a$. This, together with the fact that there is only a finite number of complete labellings bigger or equal to $\mathcal{L}_a$ implies that the set of complete labellings that are bigger or equal to $\mathcal{L}_a$ has at least one minimal element. In order to prove that this minimal element is also the smallest element, we have to prove that no other minimal element exists.
Let $\mathcal{L}_{mc_1}$ and $\mathcal{L}_{mc_2}$ be minimal complete labellings such that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_1}$ and $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_2}$. Now consider the sceptical outcome labelling associated with $\mathcal{L}_{mc_1}$ and $\mathcal{L}_{mc_2}$. That is, let $\mathcal{L}_{mc_3} = so_{AF}(\{\mathcal{L}_{mc_1}, \mathcal{L}_{mc_2}\})$. From the fact that $\mathcal{L}_{mc_1}$ and $\mathcal{L}_{mc_2}$ are complete labellings, it follows from Theorem 8 that $\mathcal{L}_{mc_3}$ is also a complete labelling. From Definition 18 it directly follows that $\mathcal{L}_{mc_3} \sqsubseteq \mathcal{L}_{mc_1}$ and $\mathcal{L}_{mc_3} \sqsubseteq \mathcal{L}_{mc_2}$. Let $\mathcal{L}_{si}$ be $\mathcal{L}_{mc_1} \sqcap \mathcal{L}_{mc_2}$. Then from the fact that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_1}$ and $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_2}$ it follows that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{si}$. From the fact that $\mathcal{L}_{mc_3}$ is the (unique) biggest admissible labelling such that $\mathcal{L}_{mc_3} \sqsubseteq \mathcal{L}_{si}$ it then follows that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_3}$. So $\mathcal{L}_{mc_3}$ is a complete labelling such that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_3}$. From the fact that $\mathcal{L}_{mc_1}$ is a minimal complete labelling such that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_1}$ and the earlier observed fact that $\mathcal{L}_{mc_3} \sqsubseteq \mathcal{L}_{mc_1}$ it then follows that $\mathcal{L}_{mc_3} = \mathcal{L}_{mc_1}$. Similarly, from the fact that $\mathcal{L}_{mc_2}$ is a minimal complete labelling such that $\mathcal{L}_a \sqsubseteq \mathcal{L}_{mc_2}$ and the earlier observed fact that $\mathcal{L}_{mc_3} \sqsubseteq \mathcal{L}_{mc_2}$ it then

follows that $\mathcal{L}_{mc_3} = \mathcal{L}_{mc_2}$. From $\mathcal{L}_{mc_3} = \mathcal{L}_{mc_1}$ and $\mathcal{L}_{mc_3} = \mathcal{L}_{mc_2}$ it then follows that $\mathcal{L}_{mc_1} = \mathcal{L}_{mc_2}$.

Using Theorem 11, we can then define the *up-complete labelling* of an admissible labelling.

**Definition 24** Let $\mathcal{L}_a$ be an admissible labelling of argumentation framework $AF = (Ar, def)$. The *up-complete* labelling of $\mathcal{L}_a$ is the smallest element of the set of all complete labellings that are bigger or equal to $\mathcal{L}_a$.

**Definition 25** Let $\mathcal{L}abellings$ be the set of all labellings of argumentation framework $AF = (Ar, def)$. The *super credulous aggregation operator* is a function $sco_{AF} : 2^{\mathcal{L}abellings} - \{\emptyset\} \rightarrow \mathcal{L}abellings$ such that $sco(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is the up-complete labelling of $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

We now prove that the super credulous expansion sequence is a way of calculating the super credulous outcome.

**Theorem 12** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ $(n \geq 1)$ be admissible labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{co} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ be the associated credulous outcome and $\mathcal{L}_{sco} = sco_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ be the associated super credulous outcome. Let $[\mathcal{L}'_1, \ldots, \mathcal{L}'_m]$ be an associated super credulous expansion sequence. It holds that $\mathcal{L}'_m = \mathcal{L}_{sco}$.*

*Proof* From Lemma 6 it follows that $\mathcal{L}'_m$ is a complete labelling that is bigger or equal to $\mathcal{L}_{co}$. From the fact that $\mathcal{L}_{sco}$ is the smallest complete labelling that is bigger or equal to $\mathcal{L}'_m$ it then follows that $\mathcal{L}_{sco} \sqsubseteq \mathcal{L}'_m$.
We now prove, by induction, that for every $j \in \{1, \ldots, m\}$: $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$.

basis: From the fact that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_{sco}$ and that $\mathcal{L}'_1 = \mathcal{L}_{co}$ it follows that $\mathcal{L}'_1 \sqsubseteq \mathcal{L}_{sco}$.

step: Suppose $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$. Let $A$ be the argument on which the expansion step is done. That is, $A$ is the argument such that $e_{AF}(\mathcal{L}'_j, A) = \mathcal{L}'_{j+1}$. We now have to prove that $\mathcal{L}'_{j+1} \sqsubseteq \mathcal{L}_{sco}$. That is, we have to prove that $\texttt{in}(\mathcal{L}'_{j+1}) \subseteq \texttt{in}(\mathcal{L}_{sco})$ and that $\texttt{out}(\mathcal{L}'_{j+1}) \subseteq \texttt{out}(\mathcal{L}_{sco})$.

- Let $B \in \texttt{in}(\mathcal{L}'_{j+1})$. If $B \neq A$ then it follows that $\mathcal{L}'_{j+1}(B) = \mathcal{L}_j(B)$, so from the fact that $\mathcal{L}'_{j+1}(B) = \texttt{in}$ it follows that $\mathcal{L}'_j(B) = \texttt{in}$, and from $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$ (the induction hypothesis) it then follows that $\mathcal{L}_{sco}(B) = \texttt{in}$, so $B \in \texttt{in}(\mathcal{L}_{sco})$.
  If $B = A$ (so the expansion step was done on $B$) then it follows that each defeater $C$ of $B$ is labelled $\texttt{out}$ by $\mathcal{L}'_j$. From $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$ (the induction hypothesis) it then follows that each defeater $C$ of $B$ is also labelled $\texttt{out}$ by $\mathcal{L}_{sco}$. From the fact that $\mathcal{L}_{sco}$ is a complete labelling, it then follows that $B$ is labelled $\texttt{in}$. So $B \in \texttt{in}(\mathcal{L}_{sco})$.
  In both cases (both $A \neq B$ and $A = B$) we have that $B \in \texttt{in}(\mathcal{L}_{sco})$, so we have that $\texttt{in}(\mathcal{L}'_{j+1}) \subseteq \texttt{in}(\mathcal{L}_{sco})$.
- Let $B \in \texttt{out}(\mathcal{L}'_{j+1})$. If $B \neq A$ then it follows that $\mathcal{L}'_{j+1}(B) = \mathcal{L}_j(B)$, so from the fact that $\mathcal{L}'_{j+1}(B) = \texttt{out}$ it follows that $\mathcal{L}'_j(B) = \texttt{out}$, and from $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$ (the induction hypothesis) it then follows that $\mathcal{L}_{sco}(B) = \texttt{out}$, so $B \in \texttt{out}(\mathcal{L}_{sco})$.
  If $B = A$ (so the expansion step was done on $B$) then it follows that $B$ has a defeater $C$ that is labelled $\texttt{in}$ by $\mathcal{L}'_j$. From $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$ (the induction hypothesis) it then follows that $C$ is also labelled $\texttt{in}$ by $\mathcal{L}_{sco}$. From the fact that $\mathcal{L}_{sco}$ is a complete labelling, it then follows that $B$ is labelled $\texttt{out}$. So $B \in \texttt{out}(\mathcal{L}_{sco})$.
  In both cases (both $A \neq B$ and $A = B$) we have that $B \in \texttt{out}(\mathcal{L}_{sco})$, so we have that $\texttt{out}(\mathcal{L}'_{j+1}) \subseteq \texttt{out}(\mathcal{L}_{sco})$.

From the thus derived facts that $\mathtt{in}(\mathcal{L}'_{j+1}) \subseteq \mathtt{in}(\mathcal{L}_{sco})$, and that $\mathtt{out}(\mathcal{L}'_{j+1}) \subseteq \mathtt{out}(\mathcal{L}_{sco})$ it follows that $\mathcal{L}_{j+1} \sqsubseteq \mathcal{L}_{sco}$.

From the thus obtained fact that for each $j \in \{1, \dots, m\}$: $\mathcal{L}'_j \sqsubseteq \mathcal{L}_{sco}$ it directly follows that $\mathcal{L}'_m \sqsubseteq \mathcal{L}_{sco}$. This, together with the earlier observed fact that $\mathcal{L}_{sco} \sqsubseteq \mathcal{L}'_m$ implies that $\mathcal{L}'_m = \mathcal{L}_{sco}$.

Now that we have proved that the procedure of the super credulous expansion sequence indeed produces the super credulous outcome, we are now getting ready to prove that this super credulous outcome is also compatible with each individual input labelling. To do this, we first need the following lemma.

**Lemma 7** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be admissible labellings of argumentation framework $(Ar, def)$, such that $\mathcal{L}_1 \approx \mathcal{L}_2$. Let $A$ be an argument that is illegally* `undec` *in $\mathcal{L}_1$. It holds that $\mathcal{L}'_1 = e_{AF}(\mathcal{L}_1, A)$ is again an admissible labelling with $\mathcal{L}'_1 \approx \mathcal{L}_2$.*

*Proof* The fact that $\mathcal{L}'_1$ is an admissible labelling follows from the proof of Lemma 6 (induction step of the second induction proof). We now prove that $\mathcal{L}'_1 \approx \mathcal{L}_2$. For this, we need to prove that (Proposition 2) $\mathtt{in}(\mathcal{L}'_1) \subseteq \mathtt{in}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$ and that $\mathtt{out}(\mathcal{L}'_1) \subseteq \mathtt{out}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.

- Let $B \in \mathtt{in}(\mathcal{L}'_1)$. If $B \neq A$ then $\mathcal{L}'{}_1(B) = \mathcal{L}_1(B)$, so $\mathcal{L}_1(A) = \mathtt{in}$. From the fact that $\mathcal{L}_1 \approx \mathcal{L}_2$ it follows that $B \in \mathtt{in}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.
  If $B = A$ then each defeater $C$ of $B$ is labelled `out` by $\mathcal{L}_1$. From the fact that $\mathcal{L}_1 \approx \mathcal{L}_2$ it follows that each defeater $C$ of $B$ is labelled either `out` or `undec` by $\mathcal{L}_2$. This then means that $B$ is labelled either `in` or `undec` by $\mathcal{L}_2$ (it cannot be labelled `out` because then it would be illegally `out`). That is, $B \in \mathtt{in}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.
- Let $B \in \mathtt{out}(\mathcal{L}'_1)$. If $B \neq A$ then $\mathcal{L}'_1(B) = \mathcal{L}_1(B)$, so $\mathcal{L}_1(A) = \mathtt{out}$. From the fact that $\mathcal{L}_1 \approx \mathcal{L}_2$ it follows that $B \in \mathtt{out}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.
  If $B = A$ then $B$ has a defeater $C$ that is labelled `in` by $\mathcal{L}_1$. From the fact that $\mathcal{L}_1 \approx \mathcal{L}_2$ it follows that $C$ is labelled either `in` or `undec` by $\mathcal{L}_2$. This then means that $B$ is labelled either `out` or `undec` by $\mathcal{L}_2$ (it cannot be labelled `in` because then it would be illegally `in`). That is, $B \in \mathtt{out}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$.

From the thus observed facts that $\mathtt{in}(\mathcal{L}'_1) \subseteq \mathtt{in}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$ and that $\mathtt{out}(\mathcal{L}'_1) \subseteq \mathtt{out}(\mathcal{L}_2) \cup \mathtt{undec}(\mathcal{L}_2)$, it follows that $\mathcal{L}'_1 \approx \mathcal{L}_2$.

Using Lemma 7, it is not too difficult to show that $\mathcal{L}_{sco}$ is compatible with each $\mathcal{L}_i$, assuming that each $\mathcal{L}_i$ is admissible.

**Theorem 13** *Let $\mathcal{L}_1, \dots, \mathcal{L}_n$ $(1 \leq n)$ be admissible labellings of argumentation framework $AF = (Ar, def)$, and let $\mathcal{L}_{sco} = sco_{AF}(\{\mathcal{L}_1, \dots, \mathcal{L}_n\})$ be the associated super credulous outcome. It holds that for every $i \in \{1, \dots, n\}$: $\mathcal{L}_{sco} \approx \mathcal{L}_i$.*

*Proof* Let $\mathcal{L}_i$ be an arbitrary element of $\{\mathcal{L}_1, \dots, \mathcal{L}_n\}$. Let $[\mathcal{L}'_1, \dots, \mathcal{L}'_m]$ be a super credulous expansion sequence from $\mathcal{L}_{co} = co_{AF}(\{\mathcal{L}_1, \dots, \mathcal{L}_n\})$. We now prove, by induction, that for each $j \in \{1, \dots, m\}$ it holds that $\mathcal{L}_i \approx \mathcal{L}'_j$.

basis: From the fact that $\mathcal{L}'_1 = \mathcal{L}_{co}$ and that $\mathcal{L}_{co} \approx \mathcal{L}_i$ (Theorem 10) it follows that $\mathcal{L}'_1 \approx \mathcal{L}_i$. Furthermore, it should be mentioned that $\mathcal{L}'_1$ is an admissible labelling

step: Suppose that for some $j \in \{1, \dots, m-1\}$, $\mathcal{L}'_j$ is an admissible labelling with $\mathcal{L}'_j \approx \mathcal{L}_i$. Then Lemma 7 tells us that $\mathcal{L}'_{j+1}$ is an admissible labelling with $\mathcal{L}'_{j+1} \approx \mathcal{L}_i$.

From the thus obtained fact that for each $j \in \{1, \dots, m\}$: $\mathcal{L}'_j \approx \mathcal{L}_i$, it directly follows that $\mathcal{L}'_m \approx \mathcal{L}_i$. And from the fact that $\mathcal{L}'_m = \mathcal{L}_{sco}$ (Theorem 12) it then directly follows that $\mathcal{L}_{sco} \approx \mathcal{L}_i$. The fact that the compatibility relation is symmetric then directly implies that $\mathcal{L}_1 \approx \mathcal{L}_{sco}$.

## 9 Some Properties

In this section we examine some properties of the theory that has been developed until now. In particular, we examine how the sceptical, credulous and super credulous aggregation operators relate to each other.

### 9.1 Relating the sceptical, credulous and super credulous outcome

An interesting property is that the credulous outcome labelling is bigger or equal to the sceptical outcome labelling.

**Theorem 14** *Let $\mathcal{L}_1, \dots, \mathcal{L}_n$ be admissible labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{so} = so_{AF}(\{\mathcal{L}_1, \dots, \mathcal{L}_n\})$ be the sceptical outcome labelling and let $\mathcal{L}_{co} = co_{AF}(\{\mathcal{L}_1, \dots, \mathcal{L}_n\})$ be the credulous outcome labelling. It holds that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{co}$.*

*Proof* First of all, it can be observed that $\mathcal{L}_{si} \sqsubseteq \mathcal{L}_{ci}$. This can be seen as follows. Let $A$ be labelled `in` by $\mathcal{L}_{si}$. Then $A$ is labelled `in` by every $\mathcal{L}_i$ $(i \in \{1, \dots, n\})$. It then also trivially follows that $A$ is labelled `in` or `undec` by every $\mathcal{L}_i$ $(i \in \{1, \dots, n\})$ so $A$ is labelled `in` by $\mathcal{L}_{ci}$. So $\mathtt{in}(\mathcal{L}_{si}) \subseteq \mathtt{in}(\mathcal{L}_{ci})$. Similarly, let $A$ be labelled `out` by $\mathcal{L}_{si}$. Then $A$ is labelled `out` by every $\mathcal{L}_i$ $(i \in \{1, \dots, n\})$. It then also trivially follows that $A$ is labelled `out` or `undec` by every $\mathcal{L}_i$ $(i \in \{1, \dots, n\})$ so $A$ is labelled `out` by $\mathcal{L}_{ci}$. So $\mathtt{out}(\mathcal{L}_{si}) \subseteq \mathtt{out}(\mathcal{L}_{ci})$. From the facts that $\mathtt{in}(\mathcal{L}_{si}) \subseteq \mathtt{in}(\mathcal{L}_{ci})$ and $\mathtt{out}(\mathcal{L}_{si}) \subseteq \mathtt{out}(\mathcal{L}_{ci})$ it follows that $\mathcal{L}_{si} \sqsubseteq \mathcal{L}_{ci}$. This, together with the fact that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{si}$, implies that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{ci}$ (because "$\sqsubseteq$" is transitive). So, $\mathcal{L}_{so}$ is an admissible labelling with $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{ci}$. From the fact that $\mathcal{L}_{co}$ is the (unique) biggest admissible labelling such that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_{ci}$ it then follows that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{co}$.

Theorem 14, together with the fact that the super credulous outcome is bigger or equal to the credulous outcome (which follows directly from Definition 25 and Definition 24) implies that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{co} \sqsubseteq \mathcal{L}_{sco}$. An interesting question is whether one could define an aggregated outcome that is even bigger than the super credulous outcome, while at the same time remaining compatible ("$\approx$") with each input labelling $\mathcal{L}_i$. Although there are possibilities of doing so, the practical use would be quite limited. First of all, we cannot simply define such an "ultra credulous" result as a maximal admissible (or maximal complete) labelling that is compatible with each $\mathcal{L}_i$, because such a labelling may not need to be uniquely defined. As an example, consider again the argumentation framework shown in Figure 6. Let $\mathcal{L}_1 = (\{A\}, \{B\}, \{C, D, E, F\})$ and $\mathcal{L}_2 = (\{B\}, \{A\}, \{C, D, E, F\})$. Then both $(\{C\}, \{D\}, \{A, B, E, F\})$ and $(\{D, F\}, \{C, E\}, \{A, B\})$ are maximal admissible (and maximal complete) labellings that are compatible with $\mathcal{L}_1$ and $\mathcal{L}_2$. Another possibility for defining an ultra credulous aggregation operator would be first to take all maximal admissible (or maximal complete) labellings that are compatible with each $\mathcal{L}_i$, and then to calculate their (sceptical, credulous or super credulous) outcome. Although this approach would yield a unique result (which is bigger or equal to $\mathcal{L}_{sc}$) it still suffers from a problem. Consider the argumentation framerwork $(Ar, def)$ with $Ar = \{A, B\}$ and $def = \{(A, B), (B, A), (B, B)\}$ and let $\mathcal{L}_1 = \mathcal{L}_2 = (\emptyset, \emptyset, \{A, B\})$. Then, our ultra credulous approach would yield the outcome $(\{A\}, \{B\}, \emptyset)$, although there is actually nothing in the input labellings that provides a proper justification for such an outcome. Based on these observations, we believe that the super credulous aggregation operator is probably the most credulous aggregation operator that can still be perceived as reasonable.

## 9.2 Relevance of the participants' input

There exists another interesting difference between the sceptical outcome and the credulous outcome. Consider the following scenario. There is a faculty meeting of which Martin and Gabriella are invited. We assume that Martin has a more cautious view of the world than Gabriella. That is, Martin's position is smaller than Gabriella's position. If the meeting applies the sceptical outcome procedure, then Gabriella might as well stay at home, because her participation will not influence the outcome of the meeting. However, if the meeting applies the credulous (or super credulous) procedure, then it is Martin who might as well stay at home, because his participation will not influence the outcome of the meeting. This is formalized by the following two theorems.

**Theorem 15** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}$ be admissible labellings of argumentation framework $AF = (Ar, def)$ such that $\mathcal{L}_i \sqsubseteq \mathcal{L}_{n+1}$ for some $i \in \{1, \ldots, n\}$. It holds that $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.*

*Proof* We first prove that $sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.
"$\sqsubseteq$": Let $A \in \mathtt{in}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$. Then $A \in \mathtt{in}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n+1\}$, so also $A \in \mathtt{in}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n\}$. It then follows that $A \in \mathtt{in}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}))$.
Let $A \in \mathtt{out}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$. Then $A \in \mathtt{out}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n+1\}$, so also $A \in \mathtt{out}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n\}$. It then follows that $A \in \mathtt{out}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}))$.
"$\sqsupseteq$": Let $A \in \mathtt{in}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}))$. Then $A \in \mathtt{in}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n\}$. From the fact that $\mathcal{L}_i \sqsubseteq \mathcal{L}_{n+1}$ for some $i \in \{1, \ldots, n\}$, it then follows that $A \in \mathtt{in}(\mathcal{L}_{n+1})$. So $A \in \mathtt{in}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n+1\}$. So $A \in \mathtt{in}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$.
Let $A \in \mathtt{out}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}))$. Then $A \in \mathtt{out}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n\}$. From the fact that $\mathcal{L}_i \sqsubseteq \mathcal{L}_{n+1}$ for some $i \in \{1, \ldots, n\}$, it then follows that $A \in \mathtt{out}(\mathcal{L}_{n+1})$. So $A \in \mathtt{out}(\mathcal{L}_i)$ for each $i \in \{1, \ldots, n+1\}$. So $A \in \mathtt{out}(sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$.
From the thus proved fact that $sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$, it then follows that $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

**Theorem 16** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}$ $(1 \leq n)$ be admissible labellings of argumentation framework $AF = (Ar, def)$ such that $\mathcal{L}_{n+1} \sqsubseteq \mathcal{L}_i$ for some $i \in \{1, \ldots, n\}$. It holds that $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.*

*Proof* We first prove that $cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.
"$\sqsubseteq$": Let $A \in \mathtt{in}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$. Then $\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{in}$ and $\neg\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{out}$. We distinguish two cases:

- $A \in \mathtt{in}(\mathcal{L}_{n+1})$. In that case, since $\mathcal{L}_{n+1} \sqsubseteq \mathcal{L}_i$ for some $i \in \{1, \ldots, n\}$, it follows that $A \in \mathtt{in}(\mathcal{L}_i)$ for some $i \in \{1, \ldots, n\}$.
- $A \notin \mathtt{in}(\mathcal{L}_{n+1})$. Then, from the fact that $\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{in}$, it follows that $A \in \mathtt{in}(\mathcal{L}_i)$ for some $i \in \{1, \ldots, n\}$.

In both cases, we have that $A \in \mathtt{in}(\mathcal{L}_i)$ for some $i \in \{1, \ldots, n\}$. This, together with the fact that $\neg\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{out}$ implies that $\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{in}$ and $\neg\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{out}$, so $A \in \mathtt{in}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}))$.
Let $A \in \mathtt{out}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$. Then $\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{out}$ and $\neg\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{in}$. We distinguish two cases:

- $A \in \mathtt{out}(\mathcal{L}_{n+1})$. In that case, since $\mathcal{L}_{n+1} \sqsubseteq \mathcal{L}_i$ for some $i \in \{1, \ldots, n\}$, it follows that $A \in \mathtt{out}(\mathcal{L}_i)$ for some $i \in \{1, \ldots, n\}$.
- $A \notin \mathtt{out}(\mathcal{L}_{n+1})$. Then, from the fact that $\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{out}$, it follows that $A \in \mathtt{out}(\mathcal{L}_i)$ for some $i \in \{1, \ldots, n\}$.

In both cases, we have that $A \in \mathtt{out}(\mathcal{L}_i)$ for some $i \in \{1, \ldots, n\}$. This, together with the fact that $\neg\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{in}$ implies that $\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{out}$ and $\neg\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{in}$, so $A \in \mathtt{out}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}))$.
"$\sqsupseteq$": Let $A \in \mathtt{in}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. Then $\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{in}$ and $\neg\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{out}$. From the fact that $\mathcal{L}_{n+1} \sqsubseteq \mathcal{L}_i$ for some $i \in \{1, \ldots, n\}$, it follows that $\mathcal{L}_{n+1}(A) \neq \mathtt{out}$, so $\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{in}$ and $\neg\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{out}$, so $A \in \mathtt{in}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$.
Let $A \in \mathtt{out}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. Then $\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{out}$ and $\neg\exists i \in \{1, \ldots, n\} : \mathcal{L}_i(A) = \mathtt{in}$. From the fact that $\mathcal{L}_{n+1} \sqsubseteq \mathcal{L}_i$ for some $i \in \{1, \ldots, n\}$, it follows that $\mathcal{L}_{n+1}(A) \neq \mathtt{in}$, so $\exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{out}$ and $\neq \exists i \in \{1, \ldots, n+1\} : \mathcal{L}_i(A) = \mathtt{in}$, so $A \in \mathtt{out}(cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}))$.
From the thus derived facts that $cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) \sqsubseteq cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ and $cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}) \sqsubseteq cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\})$, together with the fact that "$\sqsubseteq$" is anti-symmetric, it follows that $cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = cio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. It then follows that also $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n, \mathcal{L}_{n+1}\}) = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

### 9.3 Characterizing the preferred and grounded labellings

Using the theory that has been developed in the previous part of this paper, we are now able to provide different ways of characterizing the preferred labellings as well as the grounded labelling. We start with distinguishing four different ways of characterizing a preferred labelling (Theorem 17); notice that the first description of a preferred labelling is equivalent with the description in Definition 10.

**Theorem 17** *Let $\mathcal{L}$ be a labelling of argumentation framework $AF = (Ar, def)$. The following statements are equivalent.*

1. *$\mathcal{L}$ is a complete labelling where $\mathtt{in}(\mathcal{L})$ is maximal (w.r.t. $\subseteq$)*
2. *$\mathcal{L}$ is a complete labelling where $\mathtt{out}(\mathcal{L})$ is maximal (w.r.t. $\subseteq$)*
3. *$\mathcal{L}$ is a maximal complete labelling (w.r.t. $\sqsubseteq$)*
4. *$\mathcal{L}$ is a maximal admissible labelling (w.r.t. $\sqsubseteq$)*

*Proof* Equivalence between 1 and 2 follows almost directly from Lemma 1.
We now show that 1 implies 3. Let $\mathcal{L}'$ be a complete labelling with $\mathcal{L} \sqsubseteq \mathcal{L}'$. That is, $\mathtt{in}(\mathcal{L}) \subseteq \mathtt{in}(\mathcal{L}')$ and $\mathtt{out}(\mathcal{L}) \subseteq \mathtt{out}(\mathcal{L}')$. From the fact that $\mathcal{L}$ is a complete labelling

where $\mathtt{in}(\mathcal{L})$ is maximal and where (by equivalence with 2) $\mathtt{out}(\mathcal{L})$ is maximal, it follows that $\mathtt{in}(\mathcal{L}') = \mathtt{in}(\mathcal{L})$ and $\mathtt{out}(\mathcal{L}') = \mathtt{out}(\mathcal{L})$. Therefore also $\mathtt{undec}(\mathcal{L}') = \mathtt{undec}(\mathcal{L})$, so $\mathcal{L}' = \mathcal{L}$.
We now show that 3 implies 1. Let $\mathcal{L}'$ be a complete labelling with $\mathtt{in}(\mathcal{L}) \subseteq \mathtt{in}(\mathcal{L}')$. Then, from Lemma 1 it follows that $\mathtt{out}(\mathcal{L}) \subseteq \mathtt{out}(\mathcal{L}')$, so it holds that $\mathcal{L} \sqsubseteq \mathcal{L}'$. From the fact that $\mathcal{L}$ is a maximal complete labelling it then follows that $\mathcal{L}' = \mathcal{L}$.
We now prove that 3 implies 4. Let $\mathcal{L}$ be a maximal complete labelling. We now prove that for any admissible labelling $\mathcal{L}'$, if $\mathcal{L} \sqsubseteq \mathcal{L}'$ then $\mathcal{L} = \mathcal{L}'$. Let $\mathcal{L}'$ be an admissible labelling with $\mathcal{L} \sqsubseteq \mathcal{L}'$. Let $\mathcal{L}''$ be the up-complete labelling of $\mathcal{L}'$. Naturally, it holds that $\mathcal{L}''$ is a complete labelling with $\mathcal{L}' \sqsubseteq \mathcal{L}''$. This and the fact that $\mathcal{L} \sqsubseteq \mathcal{L}'$ implies that $\mathcal{L} \sqsubseteq \mathcal{L}''$. But since $\mathcal{L}$ is a maximal complete labelling, it then follows that $\mathcal{L} = \mathcal{L}''$, and therefore also that $\mathcal{L} = \mathcal{L}'$.
We now prove that 4 implies 3. Let $\mathcal{L}$ be a maximal admissible labelling. Let $\mathcal{L}'$ be the up-complete labelling of $\mathcal{L}$. Naturally, it holds that $\mathcal{L}'$ is a complete labelling with $\mathcal{L} \sqsubseteq \mathcal{L}'$. From the fact that $\mathcal{L}'$ is also an admissible labelling, and that $\mathcal{L}$ is a maximal admissible labelling, it follows that $\mathcal{L} = \mathcal{L}'$. This means that $\mathcal{L}$ is a complete labelling. We now prove that $\mathcal{L}$ is also a *maximal* complete labelling. Suppose $\mathcal{L}''$ is a complete labelling with $\mathcal{L} \sqsubseteq \mathcal{L}''$. Then $\mathcal{L}''$ is also an admissible labelling, and the fact that $\mathcal{L}$ is a maximal admissible labelling then implies that $\mathcal{L} = \mathcal{L}''$.

It is also possible to distinguish five different ways of characterizing the grounded labelling (Theorem 18). Notice that the first characterization of the grounded labelling is equivalent with the description in Definition 10. The uniqueness of the grounded labelling follows from description 5 in Theorem 18.

**Theorem 18** *Let $\mathcal{L}$ be a labelling of argumentation framework $AF = (Ar, def)$. The following statements are equivalent.*

1. *$\mathcal{L}$ is a complete labelling where $\mathtt{in}(\mathcal{L})$ is minimal (w.r.t. $\subseteq$)*
2. *$\mathcal{L}$ is a complete labelling where $\mathtt{out}(\mathcal{L})$ is minimal (w.r.t. $\subseteq$)*
3. *$\mathcal{L}$ is a complete labelling where $\mathtt{undec}(\mathcal{L})$ is maximal (w.r.t. $\subseteq$)*
4. *$\mathcal{L}$ is the smallest complete labelling (w.r.t. $\sqsubseteq$)*
5. *$\mathcal{L}$ is the up-complete labelling of the all-undec labelling.*

*Proof* Equivalence between 1 and 2 follows almost directly from Lemma 1. Also equivalence between 4 and 5 is straightforward. From 4, it directly follows that 1 and 2. From 1 and 2 together, it directly follows that 4.
We now show that 4 implies 3. Let $\mathcal{L}$ be the smallest complete labelling. Let $\mathcal{L}'$ be a labelling with $\mathtt{undec}(\mathcal{L}) \subseteq \mathtt{undec}(\mathcal{L}')$. It holds that $\mathcal{L} \sqsubseteq \mathcal{L}'$, so $\mathtt{in}(\mathcal{L}) \subseteq \mathtt{in}(\mathcal{L}')$ and $\mathtt{out}(\mathcal{L}) \subseteq \mathtt{out}(\mathcal{L}')$, so $\mathtt{undec}(\mathcal{L}) \supseteq \mathtt{undec}(\mathcal{L}')$, so $\mathtt{undec}(\mathcal{L}) = \mathtt{undec}(\mathcal{L}')$.
We now prove that 3 implies 4. Let $\mathcal{L}$ be a complete labelling where $\mathtt{undec}(\mathcal{L})$ is maximal. Let $\mathcal{L}'$ be the smallest complete labelling. Naturally, it holds that $\mathcal{L}' \sqsubseteq \mathcal{L}$, so $\mathtt{in}(\mathcal{L}') \subseteq \mathtt{in}(\mathcal{L})$ and $\mathtt{out}(\mathcal{L}') \subseteq \mathtt{out}(\mathcal{L})$. Therefore, $\mathtt{undec}(\mathcal{L}') \supseteq \mathtt{undec}(\mathcal{L})$. From the fact that $\mathtt{undec}(\mathcal{L}$ is maximal it then follows that $\mathtt{undec}(\mathcal{L}') = \mathtt{undec}(\mathcal{L})$. From the fact that $\mathtt{in}(\mathcal{L}') \subseteq \mathtt{in}(\mathcal{L})$ and $\mathtt{out}(\mathcal{L}') \subseteq \mathtt{out}(\mathcal{L})$ it then follows that $\mathcal{L}' = \mathcal{L}$.

## 9.4 On the scalability of the aggregation operators

Since the proposed aggregation operators are not based on majority voting but instead tend to make decisions on particular forms of (weak) consensus, it is interesting to examine the behavior when it comes to large groups in which there is a big diversity of opinions. In particular, it is interesting to see the behavior of the sceptical, credulous and super credulous aggregation operators when all admissible, complete or preferred labellings are taken into account. It will turn out that several of these possibilities converge to a single non-trivial outcome labelling.

We start the discussion by introducing the following lemma.

**Lemma 8** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be admissible labellings of argumentation framework $AF = (Ar, def)$ such that $\mathcal{L}_1 \approx \mathcal{L}_2$, and let $\mathcal{L}_3$ be $\mathcal{L}_1 \sqcup \mathcal{L}_2$. It holds that $\mathcal{L}_3$ is an admissible labelling with $\mathcal{L}_1 \sqsubseteq \mathcal{L}_3$ and $\mathcal{L}_2 \sqsubseteq \mathcal{L}_3$.*

*Proof* From the fact that $\mathcal{L}_1 \approx \mathcal{L}_2$ it follows that $\texttt{in}(\mathcal{L}_1) \cap \texttt{out}(\mathcal{L}_2) = \emptyset$ and $\texttt{out}(\mathcal{L}_1) \cap \texttt{in}(\mathcal{L}_2) = \emptyset$. So $\texttt{in}(\mathcal{L}_3) = \texttt{in}(\mathcal{L}_1) \cup \texttt{in}(\mathcal{L}_2)$ and $\texttt{out}(\mathcal{L}_3) = \texttt{out}(\mathcal{L}_1) \cup \texttt{out}(\mathcal{L}_2)$. It then immediately follows that $\mathcal{L}_1 \sqsubseteq \mathcal{L}_3$ and $\mathcal{L}_2 \sqsubseteq \mathcal{L}_3$. We now prove that $\mathcal{L}_3$ is also admissible.

- Let $A$ be an argument labelled `in` by $\mathcal{L}_3$. From the fact that $\texttt{in}(\mathcal{L}_3) = \texttt{in}(\mathcal{L}_1) \cup \texttt{in}(\mathcal{L}_2)$ it follows that $A$ is labelled `in` by $\mathcal{L}_1$ or $\mathcal{L}_2$. Assume without loss of generality that $A$ is labelled `in` by $\mathcal{L}_1$ (the case of $\mathcal{L}_2$ goes similar). Then the fact that $\mathcal{L}_1$ is an admissible labelling implies that all defeaters of $A$ are labelled `out` by $\mathcal{L}_1$. And because $\texttt{out}(\mathcal{L}_3) = \texttt{out}(\mathcal{L}_1) \cup \texttt{out}(\mathcal{L}_2)$ it follows that all defeaters of $A$ are also labelled `out` by $\mathcal{L}_3$. Therefore $A$ is legally `in` in $\mathcal{L}_3$.
- Let $A$ be an argument labelled `out` by $\mathcal{L}_3$. From the fact that $\texttt{out}(\mathcal{L}_3) = \texttt{out}(\mathcal{L}_1) \cup \texttt{out}(\mathcal{L}_2)$ it follows that $A$ is labelled `out` by $\mathcal{L}_1$ or $\mathcal{L}_2$. Assume without loss of generality that $A$ is labelled `out` by $\mathcal{L}_1$ (the case of $\mathcal{L}_2$ goes similar). Then the fact that $\mathcal{L}_1$ is an admissible labelling implies that $A$ has a defeater that is labelled `in` by $\mathcal{L}_1$. And because $\texttt{in}(\mathcal{L}_3) = \texttt{in}(\mathcal{L}_1) \cup \texttt{in}(\mathcal{L}_2)$ it follows that this defeater is also labelled `in` by $\mathcal{L}_3$. Therefore $A$ is legally `out` in $\mathcal{L}_3$.

The first theorem to be proved is that for the set of all preferred labellings, the sceptical outcome is the same as the credulous outcome.

**Theorem 19** *Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$ be the set of all preferred labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{so}$ be $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ and $\mathcal{L}_{co}$ be $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{so} = \mathcal{L}_{co}$.*

*Proof* Theorem 10 states that $\mathcal{L}_{co} \approx \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Let $\mathcal{L}'_i$ be $\mathcal{L}_{co} \sqcup \mathcal{L}_i$ (for an arbitrary $i \in \{1, \ldots, n\}$). From Lemma 8 it follows that $\mathcal{L}'_i$ is an admissible labelling with $\mathcal{L}_i \sqsubseteq \mathcal{L}'_i$. However, from the fact that $\mathcal{L}_i$ is a preferred labelling, it follows that $\mathcal{L}_i$ is a maximal admissible labelling (Theorem 17). It then follows that $\mathcal{L}_i = \mathcal{L}'_i$ and therefore also that $\mathcal{L}_{co} \sqcup \mathcal{L}_i = \mathcal{L}_i$. We now show that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_i$.

- Let $A \in \texttt{in}(\mathcal{L}_{co})$. That is, $A$ is labelled `in` by $\mathcal{L}_{co}$. Then $A$ cannot be labelled `out` by $\mathcal{L}_i$ (otherwise $A$ would have to be labelled `undec` by $\mathcal{L}_{co} \sqcup \mathcal{L}_i$, so $\mathcal{L}_{co} \sqcup \mathcal{L}_i \neq \mathcal{L}_i$: contradiction). Also $A$ cannot be labelled `undec` by $\mathcal{L}_i$ (otherwise $A$ would be labelled `in` by $\mathcal{L}_{co} \sqcup \mathcal{L}_i$, so $\mathcal{L}_{co} \sqcup \mathcal{L}_i \neq \mathcal{L}_i$: contradiction). Therefore, $A$ must be labelled `in` by $\mathcal{L}_i$. That is, $A \in \texttt{in}(\mathcal{L}_i)$.
- Let $A \in \texttt{out}(\mathcal{L}_{co})$. That is, $A$ is labelled `out` by $\mathcal{L}_{co}$. Then $A$ cannot be labelled `in` by $\mathcal{L}_i$ (otherwise $A$ would have to be labelled `undec` by $\mathcal{L}_{co} \sqcup \mathcal{L}_i$, so $\mathcal{L}_{co} \sqcup \mathcal{L}_i \neq \mathcal{L}_i$: contradiction). Also $A$ cannot be labelled `undec` by $\mathcal{L}_i$ (otherwise $A$ would be labelled `out` by $\mathcal{L}_{co} \sqcup \mathcal{L}_i$, so $\mathcal{L}_{co} \sqcup \mathcal{L}_i \neq \mathcal{L}_i$: contradiction). Therefore, $A$ must be labelled `out` by $\mathcal{L}_i$. That is, $A \in \texttt{out}(\mathcal{L}_i)$.

From the fact that $\mathtt{in}(\mathcal{L}_{co}) \subseteq \mathtt{in}(\mathcal{L}_i)$ and $\mathtt{out}(\mathcal{L}_{co}) \subseteq \mathtt{out}(\mathcal{L}_i)$ it follows that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). This then implies that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_{sio}$ (where $\mathcal{L}_{sio} = sio_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. From the fact that $\mathcal{L}_{sio} \sqsubseteq \mathcal{L}_{so}$ it then follows that $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_{so}$. This, together with the fact that $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{co}$ (Theorem 14), implies that $\mathcal{L}_{so} = \mathcal{L}_{co}$.

It also holds that for the set of all preferred labellings, the credulous outcome is the same as the super credulous outcome.

**Theorem 20** *Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$ be the set of all preferred labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{co}$ be $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ and $\mathcal{L}_{sco}$ be $sco_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{co} = \mathcal{L}_{sco}$.*

*Proof* From the fact that each $\mathcal{L}_i$ $(i \in \{1, \ldots, n\})$ is a preferred labelling, it directly follows that each $\mathcal{L}_i$ is a complete labelling. From Theorem 8 it then follows that $\mathcal{L}_{so}$ is a complete labelling. Since $\mathcal{L}_{so} = \mathcal{L}_{co}$ (Theorem 19) it the follows that $\mathcal{L}_{co}$ is a complete labelling. It then directly follows that the up-complete labelling of $\mathcal{L}_{co}$ is $\mathcal{L}_{co}$ itself. That is, $\mathcal{L}_{sco} = \mathcal{L}_{co}$.

The credulous outcome of the set of all preferred labellings is the same as the credulous outcome of the set of all complete labellings, which is also the same as the credulous outcome of all admissible labellings. Recall that for any argumentation framework $AF$ it holds that $\mathcal{T}_{preferred}(AF) \subseteq \mathcal{T}_{complete}(AF) \subseteq \mathcal{T}_{admissible}(AF)$.

**Theorem 21** *Let $AF = (Ar, def)$ be an argumentation framework and let $\{\mathcal{L}_1, \ldots, \mathcal{L}_k\}$ be the set of all its preferred labellings, $\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m\}$ be the set of all its complete labellings and $\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m, \ldots, \mathcal{L}_n\}$ be the set of all its admissible labellings. Let $\mathcal{L}_{co-pref} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k\})$, $\mathcal{L}_{co-comp} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots \mathcal{L}_m\})$ and $\mathcal{L}_{co-adm} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{co-pref} = \mathcal{L}_{co-comp} = \mathcal{L}_{co-adm}$.*

*Proof* Theorem 17 states that preferred labellings are maximal admissible labellings. Together with the fact that complete labellings are also preferred labellings, this implies that for each $\mathcal{L}_i$ $(i \in \{1, \ldots, n\})$ there exists an $\mathcal{L}_j$ $(j \in \{1, \ldots, k\})$ with $\mathcal{L}_i \sqsubseteq \mathcal{L}_j$. Using Theorem 16 it then follows that $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k\}) = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m\})$ as well as that $co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k\}) = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m, \ldots, \mathcal{L}_n\})$. That is, $\mathcal{L}_{co-pref} = \mathcal{L}_{co-comp} = \mathcal{L}_{co-adm}$.

The super credulous outcome of the set of all preferred labellings is the same as the super credulous outcome of the set of all complete labellings, which is also the same as the super credulous outcome of all admissible labellings.

**Theorem 22** *Let $AF = (Ar, def)$ be an argumentation framework and let $\{\mathcal{L}_1, \ldots, \mathcal{L}_k\}$ be the set of all its preferred labellings, $\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m\}$ be the set of all its complete labellings and $\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m, \ldots, \mathcal{L}_n\}$ be the set of all its admissible labellings. Let $\mathcal{L}_{sco-pref} = sco_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k\})$, $\mathcal{L}_{sco-comp} = sco_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots \mathcal{L}_m\})$ and $\mathcal{L}_{sco-adm} = sco_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{sco-pref} = \mathcal{L}_{sco-comp} = \mathcal{L}_{sco-adm}$.*

*Proof* Let $\mathcal{L}_{co-pref} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k\})$, $\mathcal{L}_{co-comp} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots \mathcal{L}_m\})$ and $\mathcal{L}_{co-adm} = co_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_k, \ldots, \mathcal{L}_m, \ldots, \mathcal{L}_n\})$. Since it holds that

$\mathcal{L}_{co-pref} = \mathcal{L}_{sco-comp} = \mathcal{L}_{sco-adm}$ (Theorem 21), it follows that the up-complete labelling of $\mathcal{L}_{co-pref}$ is equal to the up-complete labelling of $\mathcal{L}_{co-comp}$, which is then also equal to the up-complete labelling of $\mathcal{L}_{co-adm}$. That is, $\mathcal{L}_{sco-pref} = \mathcal{L}_{sco-comp} = \mathcal{L}_{sco-adm}$.

The sceptical outcome of the set of all complete labellings is the grounded labelling.

**Theorem 23** *Let $\mathcal{L}_G$ be the grounded labelling and $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$ be the set of complete labellings of argumentation framework $AF = (Ar, def)$. It holds that $\mathcal{L}_G = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.*

*Proof* Since the grounded labelling is also a complete labelling, it holds that $\mathcal{L}_G \in \{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$. Since the grounded labelling is the smallest complete labelling (Theorem 18) it directly follows that $\mathcal{L}_G \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Using Theorem 15, we then obtain that $so_{AF}(\{\mathcal{L}_G\}) = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$, and since $so_{AF}(\{\mathcal{L}_G\}) = \mathcal{L}_G$, it then follows that $\mathcal{L}_G = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

The sceptical outcome of the set of all admissible labellings is the all-undec labelling.

**Theorem 24** *Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$ be the set of admissible labellings of argumentation framework $AF = (Ar, def)$. It holds that $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is the all-undec labelling of $AF$.*

*Proof* Let $\mathcal{L}_U$ be the all-undec labelling of $AF$. It holds that $\mathcal{L}_U$ is an admissible labelling. Moreover, it also holds that $\mathcal{L}_U$ is smaller or equal to any admissible labelling. That is $\mathcal{L}_U \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$). Using Theorem 15 we then obtain that $so_{AF}(\{\mathcal{L}_U\}) = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$, and since $so_{AF}(\{\mathcal{L}_U\}) = \mathcal{L}_U$, we then obtain that $\mathcal{L}_U = so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$.

We will use the term *ideal labelling* for the sceptical, credulous and super credulous outcome of all preferred labellings, which is, as has been shown, also equal to the credulous and super credulous outcome of all complete or admissible labellings. It turns out that there is an interesting similarity between the concept of an ideal labelling and the concept of a maximal ideal set described by Dung, Mancarella and Toni [14].

**Definition 26 ([14])** A set of arguments is ideal iff it is admissible and contained in every preferred extension.

**Theorem 25 ([14])**

1. *every argumentation framework has a unique maximal ideal set of arguments*
2. *the maximal ideal set of arguments is complete*

We now prove that the ideal labelling can be seen as the labelling version of the maximal ideal set.

**Theorem 26** *Let $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$ be the set of all preferred labellings of argumentation framework $AF = (Ar, def)$. Let $\mathcal{L}_{ideal}$ be $so_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$. It holds that $\mathcal{L}_{ideal}$ is the maximal ideal set.*

*Proof* From Theorem 7 it follows that $\mathcal{L}_{ideal}$ is the biggest admissible labelling such that $\mathcal{L}_{ideal} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$) from which it immediately follows that $\mathtt{in}(\mathcal{L}_{ideal}) \subseteq \mathtt{in}(\mathcal{L}_i)$ ($i \in \{1, \ldots, n\}$). Since each $\mathcal{L}_i$ is a preferred labelling, it follows that $\mathtt{in}(\mathcal{L}_i)$ is a preferred extension (Theorem 4). Also the fact that $\mathcal{L}_{ideal}$ is an admissible labelling implies that $\mathtt{in}(\mathcal{L}_{ideal})$ is an admissible set (Theorem 1). So $\mathtt{in}(\mathcal{L}_{ideal})$ is an admissible set that is a subset of each preferred extension, hence $\mathtt{in}(\mathcal{L}_{ideal})$ is an ideal set. We now prove that it is also the unique biggest ideal set. Let $\mathcal{A}rgs \supseteq \mathtt{in}(\mathcal{L}_{ideal})$ be an admissible set that is a subset of each preferred extension. That is, let $\mathcal{A}rgs \supseteq \mathtt{in}(\mathcal{L}_{ideal})$ (for each $i \in \{1, \ldots, n\}$). Let $\mathcal{L}'_{ideal}$ be a labelling with $\mathtt{in}(\mathcal{L}'_{ideal}) = \mathcal{A}rgs$. By definition, it then holds that $\mathtt{in}(\mathcal{L}'_{ideal}) \subseteq \mathtt{in}(\mathcal{L}_i)$ (for each $i \in \{1, \ldots, n\}$). From Lemma 1 it then follows that $\mathtt{out}(\mathcal{L}'_{ideal}) \subseteq \mathtt{out}(\mathcal{L}_i)$, so $\mathcal{L}'_{ideal} \sqsubseteq \mathcal{L}_i$. From the fact that $\mathcal{A}rgs \supseteq \mathtt{in}(\mathcal{L}_{ideal})$ it directly follows that $\mathtt{in}(\mathcal{L}'_{ideal}) \supseteq \mathtt{in}(\mathcal{L}_{ideal})$, so (Lemma 1) $\mathtt{out}(\mathcal{L}'_{ideal}) \supseteq \mathtt{out}(\mathcal{L}_{ideal})$, so $\mathcal{L}'_{ideal} \sqsupseteq \mathcal{L}_{ideal}$. This, together with the fact that $\mathcal{L}_{ideal}$ is the *biggest* admissible labelling satisfying $\mathcal{L}_{ideal} \sqsubseteq \mathcal{L}_i$ (for each $i \in \{1, \ldots, n\}$) implies that $\mathcal{L}'_{ideal} = \mathcal{L}_{ideal}$. Therefore, $\mathcal{A}rgs = \mathtt{in}(\mathcal{L}_{ideal})$.

When one takes all semi-stable labellings as input, then the sceptical, credulous and super credulous outcome are all equal to each other. This can be proved by slightly modifying the proofs of Theorem 19 and 20, and taking into account that each semi-stable labelling is also a preferred labelling. We will call this outcome the *eager labelling* ($\mathcal{L}_{eager}$) and it can be shown that $\mathtt{in}(\mathcal{L}_{eager})$ is equal to the *eager extension* as described in [4].

The overall results of the discussion are represented in Table 3.

| | sceptical outcome | credulous outcome | super credulous outcome |
|---|---|---|---|
| all semi-stables | eager | eager | eager |
| all preferreds | ideal | ideal | ideal |
| all completes | grounded | ideal | ideal |
| all admissibles | all-undec | ideal | ideal |

**Table 3** Scalability of aggregation operators

Overall, we can observe that although the credulous outcome is more or equally committed (bigger or equal) than the sceptical outcome, and the super credulous outcome is more or equally committed (bigger or equal) than the credulous outcome, these outcomes in many cases tend to converge to the same result if the group size is large enough and opinions are diverse enough. In most of these cases, the outcome is the ideal or eager labelling, in one case (sceptical complete) it is the grounded labelling, and in only one case (sceptical admissible) the result collapses in the form of the all-undec labelling.

## 10 Discussion and Related Work

In this paper, we have declaratively defined the sceptical, credulous and super credulous outcome for judgment aggregation in the context of abstract argumentation. Moreover, we also have defined proof procedures to obtain these outcomes, using the concept of contraction and expansion sequences. The super credulous outcome is more committed than the credulous outcome, which is in its turn more committed than the sceptical outcome. The sceptical outcome is less committed than the individual positions that were used to generate it. For the credulous and super credulous outcome, this is in general not the case, although these do satisfy the weaker condition that at least the outcome is compatible with each individual position that was used to generate it. For the sceptical outcome, input labellings that are more committed than other input labellings do not influence the outcome, and can simply be omitted. For the credulous and super credulous outcome, on the other hand, input labellings that are less committed than other input labellings do not influence the outcome and can be omitted. The proposed aggregation outcomes (sceptical, credulous and super credulous) can be described as consensus oriented instead of as majority based, which raises the question of how they behave in large groups with a great variety of opinions. However, it turns out that in many cases, the outcome converges to a non-trivial well-defined position (like the eager, ideal or grounded labelling).

Our paper is the first that applies abstract argumentation to judgment aggregation, and that introduces operators that do not violate any of the members' views. The only two works that have investigated the aggregation of individual defeat relations into a social one are by Coste-Marquis *et al.* [7] and Tohmé *et al.* [38]. These two proposals differ substantially. In [7], an approach to merge Dung's argumentation frameworks is presented. The argumentation frameworks to be merged may be different, that is agents may ignore arguments put forward by other agents. Conflicts between argumentation frameworks are solved using merging techniques [21], in particular a distance-based merging operator. The intuition is to minimize the distance between the profile and the collective outcome. Typically, more than one argumentation system minimizes the chosen distance. Hence, the final step consists in asking the individuals to vote on the selected extensions to obtain the final group argumentation framework. Their approach is shown to preserve at the collective level all the evaluations on which the individuals do not disagree.

In Tohmé *et al.* the aggregation of individual attack relations is linked to the aggregation of individual preferences in social choice. However, preferences and attack relations are different: preferences are usually assumed to be weak orders, while such restriction is not imposed on attack relations. Another difference is that, unlike attack relations, preference relations have maximal elements. It is exactly in virtue of these differences that Tohmé *et al.* can apply Arrow's theorem conditions to argumentation without this necessarily implying an impossibility result. They indeed show that, by assuming argumentation frameworks in which the attack relations are acyclic, it is possible to define an aggregation operator that satisfies Arrow's theorem conditions.

Results in judgment aggregation usually assume complete judgment sets both at the individual and collective level. However, in argumentation theory individuals are allowed to be undecided regarding some arguments. Hence, in our framework individuals can abstain. Gärdenfors [17] and Dokow and Holzman [12] have considered judgment aggregation with incomplete judgment sets. Gärdenfors was the first to criticize the completeness of judgment sets as being a too strong and unrealistic assumption. In his approach, voters are allowed to abstain from expressing judgments on some propositions in the agenda. He proves that, if the judgment sets may not be complete (but logically closed and consistent), then every aggregation function that is independent and Paretian, must be oligarchic. An aggregation function is *oligarchic* if, for every issue in the agenda, the group accepts or rejects that proposition if and only if all the members of a subset of the group (the oligarchy) accept or reject that proposition. Clearly, when there is only one member in the oligarchy, oligarchy corresponds to dictatorship. Gärdenfors' framework requires the agenda to have a very rich logical structure (with an infinite number of issues). Dokow and Holzman [12] showed that Gärdenfors' result holds also for finite agendas. Hence, when individuals are allowed to abstain, oligarchy replaces dictatorship. However, as we have seen in Section 6, our aggregation operators are not Paretian, since they cannot guarantee unanimity preservation.

In our search for aggregation procedures that select outcomes that do not oppose the judgments of any of the group members, we had to rule out the majority rule. We have observed that proposition-wise majority voting may return a consistent combination of reasons and conclusion that no member supported. All the members would then be forced to take positions against their beliefs, and we have presented procedures that avoid such pitfall. There have been recent discussions on whether it is appropriate that an individual changes his mind in view of the different opinion of the majority. The source of the debate is in the fact that, besides being perceived as a 'fair' aggregation rule, the majority rule received also an epistemic justification from the Condorcet Jury Theorem. This theorem states that under majority rule, if the individuals of a group are independent of each other and are better than chance on some yes/no judgment, then the probability that the whole group will select the right answer will approach infallibility as the group size increases. So, in virtue of the Condorcet Jury Theorem, an individual should surrender his judgment to the majority when he disagrees with it. However, the normative value of such requirement has been questioned [15, 6, 29, 37], also with respect specifically to judgment aggregation by Pettit [33]. Criticizing the view that requires an individual to surrender to the opinion of the majority, Pettit [33] distinguishes between more or less embedded beliefs. You may be sure that the car used by the robbers to flee was black. Nevertheless, when confronted with several other equally reliable testimonies who are sure that the car was in fact blue, you may reasonably be persuaded to be wrong and so to revise your belief. This is because perceptual beliefs are weakly entrenched in our web of beliefs. Different is the situation in which we express our opinion on issues that are strongly embedded with other beliefs we hold, such as moral belief like whether the death penalty should be permissible, or political and strategical views like whether our country should declare war to another.

Some other work on argument labellings and social preferences has been done by Rahwan and Larson [36]. In their approach, each agent has a preference relation between the various complete labellings. An example of such a preference relation would be to try to maximize the `in`-labelling of a set of argument an individual agent cares about. Thus, their work differs from ours in that an agent does not have a single labelling but a preference ordering between labellings. Rahwan and Larson show that different types of preference orderings result in different types of labellings becoming Pareto optimal [36].

## 11 Conclusions

In this paper we have studied judgment aggregation as the aggregation of individual labellings of a given argumentation framework. At the best of our knowledge, this is the first time that argumentation theory is applied to judgment aggregation, bringing in this way judgment aggregation into a nonmonotonic framework. Argumentation is a well-established field, where the different notions of argumentation semantics allow us to be flexible when defining what kind of input and output labelling an aggregation operator should accept and return.

Unlike previous contributions in judgment aggregation, our analysis did not focus on the discursive dilemma. Collective irrationality is not the only problem that may arise when individual judgments are aggregated. Our concern here was to define social outcomes that any individual participating at the decision could subscribe while guaranteeing collective rationality. The operators we proposed are specifically meant to capture the new property that the social outcome must not go against any judgment that the individuals were asked to disclose in the deliberation phase. The participation to a group should not put an individual in danger of being forced to support decisions that go against his own beliefs. That is what we mean to be a 'compatible' collective outcome.

We have presented three operators, each of which satisfies both the above property and collective rationality. In particular, the operators define degrees in which the collective judgment does not go against any of the individual judgments: the sceptical operator is much stricter than the super credulous operator. Moreover, two definitions of 'compatibility' between social outcome and individual judgments have been introduced.

However, one property that is not satisfied is that of preservation of a unanimously supported outcome. As our discussion on the Paretian dilemma and on the failure for $so_{AF}$ to satisfy both unanimity and collective coherence illustrated, this is because the aggregated judgment is not merely the sum of the individual judgments. It is very well possible that the same argument is accepted by different participants for different reasons, but that these reasons cancel each other out when being put together. Hence, failing to preserve unanimity should neither be seen as the consequence of the use of unreasonable aggregation rules nor as the individuals being paternalized by a superior entity (the group) that replaces their opinions. It is rather a new entity, the group, that emerges.

In our work we did not consider majority-based social rules. The aim of the paper is to define group outcomes compatible with the view of any group member while guaranteeing collective rationality. As seen in Section 2, the problem of aggregating individual judgments is not restricted to majority voting but affects all aggregation rules that satisfy some seemingly reasonable conditions. However, majority voting is an appealing rule when many agents take part to the decision process. We leave the investigation of majority-based social rules in an argumentation framework for future work.

In this paper we have given an example of how to map a judgment aggregation problem into an argumentation framework. However, whether such mapping exists for *all* kinds of judgment aggregation problems is still an open question, which we plan to address in future work. Judgment aggregation assumes that the individual and collective judgments are made on logically connected propositions. Such logical connection may be already explicit in the agenda (like in $\{p, \neg p, p \rightarrow q, \neg(p \rightarrow q), q, \neg q\}$) or can be given as an additional constraint, like in the Hortis Bank example. How can these different agendas and additional constraints be represented in an argumentation framework?

## Acknowledgments

We thank Leon van der Torre and three anonymous reviewers for insightful comments, and Yining Wu for helping to develop the concepts of an ideal labelling.


## References


1. K. Arrow. *Social Choice and Individual Values.* Cowles Foundation Monograph Series, 1963.
2. M.W.A. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence; 10th European Conference, JELIA 2006*, pages 111–123. Springer, 2006. LNAI 4160.
3. M.W.A. Caminada. An algorithm for computing semi-stable semantics. In *Proceedings of the 9th European Conference on Symbolic and Quantitalive Approaches to Reasoning with Uncertainty (ECSQARU 2007)*, number 4724 in Springer Lecture Notes in AI, pages 222–234, Berlin, 2007. Springer Verlag.
4. M.W.A. Caminada. Comparing two unique extension semantics for formal argumentation: ideal and eager. In Mohammad Mehdi Dastani and Edwin de Jong, editors, *Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence (BNAIC 2007)*, pages 81–87, 2007.
5. M.W.A. Caminada and Y. Wu. An argument game of stable semantics. *Logic Journal of IGPL*, 17(1):77–90, 2009.
6. D. Coady. When experts disagree. *Episteme: A Journal of Social Epistemology*, 3(1):68–79, 2006.
7. S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis. On the merging of dung's argumentation systems. *Artificial Intelligence*, 171(10-15):730–753, 2007.
8. F. Dietrich. A generalized model of judgment aggregation. *Social Choice and Welfare*, 28(4):529–565, 2007.
9. F. Dietrich and C. List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007.
10. F. Dietrich and C. List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19(4):391–424, 2007.
11. E. Dokow and R. Holzman. Aggregation of binary evaluations. *Working paper - Technion Israel University of Technology*, 2005.
12. E. Dokow and R. Holzman. Aggregation of binary evaluations with abstentions. *Working paper*, 2007.
13. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and $n$-person games. *Artificial Intelligence*, 77:321–357, 1995.
14. P. M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171(10-15):642–674, 2007.
15. D. Estlund. Beyond fairness and deliberation: The epistemic dimension of democratic authority. In William Rehg James Bohman, editor, *Deliberative democracy: Essays on Reason and Politics*, pages 173–204. MIT Press, 1997.
16. A.J. García and G.R. Simari. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.
17. P. Gärdenfors. A representation theorem for voting with logical consequences. *Economics and Philosophy*, 22:181–190, 2006.
18. G. Governatori, M.J. Maher, G. Antoniou, and D. Billington. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14(5):675–702, 2004.
19. J. Horty. Skepticism and floating conclusions. *Artificial Intelligence*, 135:55–72, 2002.
20. H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of logic and computation*, 9(2):215–261, 1999.
21. S. Konieczny and R. Pino Perez. On the logic of merging. In *Proceedings of KR'98*, pages 488–498, 1998.
22. L.A. Kornhauser and L.G. Sager. Unpacking the court. *Yale Law Journal*, 96:82–117, 1986.

23. L.A. Kornhauser and L.G. Sager. The one and the many: Adjudication in collegial courts. *California Law Review*, 81:1–51, 1993.
24. C. List and P. Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18:89–110, 2002.
25. C. List and P. Pettit. Aggregating sets of judgments: Two impossibility results compared. *Synthese*, 140(1-2):207–235, 2004.
26. C. List and C. Puppe. Judgment aggregation: A survey. In P. Anand, P. Pattanaik, and C. Puppe, editors, *Oxford Handbook of Rational and Social Choice*. Oxford University Press (New York), forthcoming.
27. D. M. Makinson and K. Schlechta. Floating conclusions and zombie paths: two deep difficulties in the "directly sceptical" approach to defeasible inheritance nets. *Artificial Intelligence*, 48:199–209, 1991.
28. S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.
29. M. A. Moffett. Reasonable disagreement and rational group inquiry. *Episteme: A Journal of Social Epistemology*, 4:352–367, 2007.
30. K. Nehring. The (im)possibility of a Paretian rational. Economics working papers, Institute for Advanced Study, School of Social Science, Nov 2005.
31. K. Nehring and C. Puppe. The structure of strategy-proof social choice. part i: General characterization and possibility results on median spaces. *Journal of Economic Theory*, 135:269–305, 2007.
32. M. Pauly and M. van Hees. Logical constraints on judgment aggregation. *Journal of Philosophical Logic*, 35:569–585, 2006.
33. P. Pettit. When to defer to majority testimony - and when not. *Analysis*, 66(3):179–187, 2006.
34. G. Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006.
35. J. L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, 1995.
36. Iyad Rahwan and Kate Larson. Pareto optimality in abstract argumentation. In Dieter Fox and Carla P. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, pages 150–155. AAAI Press, 2008.
37. E. Sosa. The epistemology of disagreement. Unpublished, 2007.
38. F.A. Tohmé, G. A. Bodanza, and G. R. Simari. Aggregation of attack relations: A social-choice theoretical analysis of defeasibility criteria. In S. Hartmann and G. Kern-Isberner, editors, *Proceedings of Foundations of Information and Knowledge Systems, FoIKS 2008*, pages 8–23, Pisa, Italy, 2008.
39. Bart Verheij. A labeling approach to the computation of credulous acceptance in argumentation. In Manuela M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India*, pages 623–628, 2007.
40. G. A. W. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA-00)*, number 1919 in Springer Lecture Notes in AI, pages 239–253, Berlin, 2000. Springer Verlag.
41. G.A.W. Vreeswijk. An algorithm to compute minimally grounded and admissible defence sets in argument systems. In P.E. Dunne and TJ.M. Bench-Capon, editors, *Computational Models of Argument; Proceedings of COMMA 2006*, pages 109–120. IOS, 2006.