Pareto optimality and strategy-proofness in group argument evaluation

Edmond Awad, Martin Caminada, Gabriella Pigozzi, Mikolaj Podlaszewski, Iyad Rahwan

Pareto Optimality and Strategy Proofness in Group Argument Evaluation (Extended Version)

Edmond Awad[1,2], Martin Caminada[3], Gabriella Pigozzi[4], Mikołaj Podlaszewski[5], and Iyad Rahwan[1,2,†]

[1] *The Media Lab, Massachusetts Institute of Technology, USA*
[2] *Masdar Institute, UAE*
[3] *School of Computer Science & Informatics, Cardiff University, UK*
[4] *Université Paris-Dauphine, PSL Research University, CNRS, UMR [7243], LAMSADE, 75016 Paris, France*
[5] *University of Luxembourg, Luxembourg*
[†] *Correspondence should be addressed to* `irahwan@mit.edu`

Abstract

An inconsistent knowledge base can be abstracted as a set of arguments and a defeat relation among them. There can be more than one consistent way to evaluate such an argumentation graph. Collective argument evaluation is the problem of aggregating the opinions of multiple agents on how a given set of arguments should be evaluated. It is crucial not only to ensure that the outcome is logically consistent, but also satisfies measures of social optimality and immunity to strategic manipulation. This is because agents have their individual preferences about what the outcome ought to be. In the current paper, we analyze three previously introduced argument-based aggregation operators with respect to Pareto optimality and strategy proofness under different general classes of agent preferences. We highlight fundamental trade-offs between strategic manipulability and social optimality on one hand, and classical logical criteria on the other. Our results motivate further investigation into the relationship between social choice and argumentation theory. The results are also relevant for choosing an appropriate aggregation operator given the criteria that are considered more important, as well as the nature of agents' preferences.

1 Introduction

Argumentation has recently become one of the main approaches for non-monotonic reasoning and multi-agent interaction in artificial intelligence and computer science [4, 7, 41]. The most prominent approach in argumentation models is probably the abstract argumentation framework (AAF) by Dung [24]. In AAF, the contents of the arguments are abstracted from and the framework can be represented as a directed graph in which nodes represent arguments (a set $\mathcal{A}$), and arcs between these nodes represent binary *defeat* relations (denoted as $\rightharpoonup$) over them.

An important question is which arguments to accept. In his seminal paper, Dung has defined extension-based semantics which correspond to different criteria of acceptability of arguments. For example, if we have two arguments that defeat each other, we cannot accept both. We may accept only one of them. Another equivalent labeling-based semantics is proposed by Caminada [18, 14]. Using this approach, an argument is labeled `in` (i.e. accepted), `out` (i.e. rejected), or `undec` (i.e. undecided).

One of the essential properties, that is common, is the condition of *admissibility*: that accepted arguments must not attack one another, and must defend themselves against counter-arguments, by attacking them back. A stronger notion is called *completeness*, and is captured, in terms of labelings, in the following two conditions:

1. An argument is labeled *accepted* (or *in*) if and only if all its defeaters are rejected (or *out*).

2. An argument is labeled *rejected* (or *out*) if and only if at least one of its defeaters is accepted (or *in*).

In all other cases, an argument should be labeled *undecided* (or *undec*). Thus, evaluating a set of arguments amounts to labeling each argument using a labeling function $\mathcal{L} : \mathcal{A} \rightarrow \{\texttt{in}, \texttt{out}, \texttt{undec}\}$ to capture these three possible labels. Any labeling that satisfies the above conditions is a *legal labeling*, and corresponds to a complete labeling (to be discussed in more detail below). Every complete (*i.e.* legal) labeling represents a consistent self-defending point of view. We will use *legal labeling* and *complete labeling* interchangably.

Since there can be different reasonable positions regarding the evaluation of an argumentation graph, choosing one legal labeling above another is not a trivial task. Therefore, in a multi-agent setting, different agents can subscribe to different positions. Hence, a group of agents with an argumentation graph would need to find a collective labeling that best reflects the opinion of the group. Consider the following example which is depicted in Figure 1.

Example 1 (A Murder Case). *A murder case is under investigation. There is an argument that the suspect is innocent, which suggests that he should be set free (A). However, there is some evidence that the suspect was at the crime scene during the crime time, which suggests that the suspect is not innocent (B). Weirdly enough, a witness confirmed that she saw someone who looks like the suspect in a bar during the crime time, which suggests that the suspect is innocent (C).*

Clearly, B and C defeat each other since they support negating conclusions. Also, B defeats A since it provides enough evidence to nullify it.

A team of four jurors has been assigned to decide on this case. They have been provided with the previous information. Figure 1 shows the three possible legal labelings. Each juror's judgment can correspond to only one of these labelings. Suppose they voted as shown in Figure 1 (the four thumbs-ups), what would be a labeling that best reflects the opinion of the team?

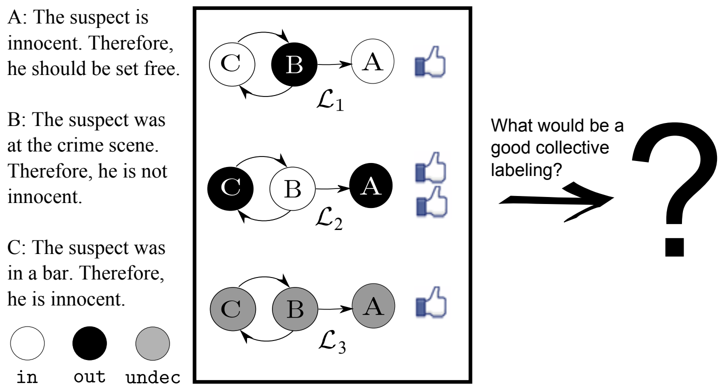

Figure 1: Three different labelings with the votes on each labeling.

Despite the apparent simplicity of the problem, the aggregation of individual evaluations can result into an inconsistent group outcome i.e. even when each individual submits a legal labeling, the aggregation outcome might not be a legal labeling. This problem of aggregating labelings can be compared to preference aggregation (PA) [1, 2, 26, 45], judgment aggregation (JA) [33, 31, 32, 30], and non-binary judgment aggregation [22, 23]. These areas have so far blossomed around impossibility results. There exist many differences between labelings and preference relations stemming from their corresponding order-theoretic characterizations. Labeling aggregation differ from JA in that arguments (which are the counterparts of propositions) can have three values instead of two traditionally considered in JA. Considering the general framework in [23], our settings can be considered as focusing on special classes of feasible evaluations, which are the conditions imposed by the legal labelling (or other semantics). Additionally, the possible evaluations of each issue (argument, in our case) are to accept (labels as `in`), reject (labels as `out`), or be undecided (labels as `undec`). However, translation of results between labeling aggregation and non-binary JA amounts to encoding argument semantics in propositional logic, which is not a trivial task [5, 6].

Recently, the problem of aggregating valid labelings has been the topic of some studies [42, 3, 15, 9, 11]. In [42, 3], the argument-wise plurality rule (AWPR) which chooses the collective label of each argument by plurality, independently from other arguments, was defined and analyzed. On the other hand, Caminada and Pigozzi [15] showed how judgment aggregation concepts can be applied to formal argumentation in a different way. They proposed three possible operators for aggregating labelings, namely the skeptical operator, the credulous operator, and the super credulous operator. These operators guarantee not only a well-formed outcome but also a compatible one, that is, it does not go against the judgment of any individual.

In order to assess the three operators, we assume that individuals have preferences over the outcomes. Although the outcomes of the three aggregation operators proposed in [15] are compatible with every individual's labeling, this does not mean that they are the most desirable given individuals' preferences. It is possible that other compatible labelings are more desirable. Moreover, it is possible that some agents submit an insincere opinion in order to get more desirable outcomes. Given that, it is interesting to study the following two questions:

1. Are the social outcomes of the aggregation operators in [15] Pareto optimal if preferences between different outcomes are also taken into consideration?

2. How robust are these operators against strategic manipulation? And what are the effects of strategic manipulation from the perspective of social welfare?

The first question studies the Pareto optimality of the outcomes of these operators. A Pareto optimal outcome (given individuals preferences) cannot be replaced with another outcome that is more preferred by all individuals and is strictly more preferred by at least one individual. Pareto optimality is a fundamental concept in any social choice setting and a clearly desirable property for any aggregation operator.

The second question studies the strategy proofness of the operators. Strategy proofness is fundamental in any realistic multi-agent setting. A strategy-proof operator is one that produces outcomes where individuals have no incentive to misrepresent their votes (i.e. to lie). Unfortunately, as we will see later, most strategy proofness results for the three operators are negative. However, we show later that lies do not always have bad effects on other agents.

One can realize that individuals' preferences (over all the labelings) play a vital role in answering the previous two questions. However, aggregation operators usually do not give the chance for individuals to disclose these preferences. The labeling an agent submits is the only information available about agent's preferences. It seems a natural choice to assume that the submitted labeling is the most preferred one according to agents' individual preference. Moreover we assume that the rest of agent's preferences can be modeled using distance from the most preferred one. For example, if the top preferred outcome for agent i is the outcome O_1 (*i.e.* $\forall O_j, O_1 \succeq_i O_j$), then $O_2 \succ_i O_3$ iff $dist(O_1, O_2) < dist(O_1, O_3)$ where $dist(O_1, O_2)$ is the distance between the two outcomes O_1 and O_2.

In this work, we investigate different classes of preferences based on different distance measures, and use them to analyze the three aggregation operators proposed in [15] with respect to the aforementioned two questions.

This paper makes three distinct contributions. First, it introduces the first thorough study of Pareto optimality and strategy proofness for aggregation operators in the context of argumentation. In doing so, the paper highlights that considering argumentation in multi-agent conflict resolution calls for criteria other than logical consistency such as social optimality and strategic manipulation.

Second, the paper introduces different families of agents' preferences. Building on the new concept of *issue*, proposed by Booth et al. [10], we define a new class of agents preferences. We also define a new class of preferences which consider the label `undec` as a middle label between `in` and `out`. These new families of preferences capture the intuitions, are more natural, and broaden the scope of analysis of preferences.

The third contribution of this paper is establishing relations between the different classes of preferences. Some of these relations hold for any aggregation operator and others for some special aggregation operators. Additionally, we provide a full comparison for three previously introduced labeling aggregation operators with respect to the proposed classes of preferences. Moreover, we also consider cases where agents do not share the same classes of preference. Our results are based on two fundamental criteria, namely Pareto optimality and strategy proofness. For most classes of preferences we establish the superiority of the skeptical operator. However, we also characterize situations where the credulous and super credulous operators are as good as the skeptical operator. This highlights a trade-off between the two criteria on one hand, and seeking more committed outcomes on the other hand.

Our results bridge a gap in our understanding of the social optimality and strategic manipulation of labeling aggregation operators. As for the Pareto optimality, we show the persistence of the superiority of the skeptical operator. However, there are situations where the credulous and super credulous operators are as good as the skeptical operator. This has an implication on the choice of the appropriate aggregation operator given the criteria that is considered more important, as well as, the nature of agents preferences.

As for the strategy proofness, we establish the fragility of the three operators against strategic manipulation. This negative result is consistent even for a wide range of individual agent preference criteria (except for two cases). This highlights a major limitation of these otherwise attractive approaches to collective argument evaluation.

Despite the negative results, our results show that lies with the skeptical operator are always benevolent i.e. every strategic lie by an agent does not hurt others, but rather improves their welfare. Furthermore, we show that this effect is surprisingly consistent for a wide range of individual agent preference criteria. This shows an important advantage for such an approach to labeling aggregation.[1]

2 Preliminaries

2.1 Abstract Argumentation Framework (AAF).[2]

The seminal paper by Dung [24] introduced the fundamental notion of abstract argumentation framework that can be represented as a directed graph where the vertices represent arguments (ignoring details about their contents) and the directed arcs represent the defeat relations between these arguments.[3] For example, in Figure 2, argument A_1 is defeated by arguments A_2 and A_4 which are, in turn, defeated by arguments A_3 and A_5.

Definition 1 (Argumentation framework [24])**.** *An* argumentation framework *is a pair* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ *where* $\mathcal{A}$ *is a finite set of arguments and* $\rightharpoonup \subseteq \mathcal{A} \times \mathcal{A}$ *is a defeat relation. We say that an argument* A defeats *an argument* B *if* $(A,B) \in \rightharpoonup$ *(sometimes written* $A \rightharpoonup B$*).*

There are two approaches to define semantics that assess the acceptability of arguments. One of them is extension-based semantics by Dung [24], which produces a set of arguments that are

[1] Part of the results of this paper have been presented in [16].

[2] Readers familiar with AAF can skip this part.

[3] We will use "argumentation graph" and "argumentation framework" interchangeably.

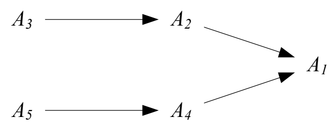

Figure 2: A simple argumentation graph

accepted together. Another equivalent labeling-based semantics is proposed by Caminada [18, 14], which gives a labeling for each argument. With argument labelings, we can accept arguments (by labeling them as `in`), reject arguments (by labeling them as `out`), and abstain from deciding whether to accept or reject (by labeling them as `undec`). As [15] employed the labeling approach, so we continue to use it here.

Definition 2 (Argument labeling [18, 14])**.** *Let* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ *be an argumentation framework. An* argument labeling *is a total function* $\mathcal{L} : \mathcal{A} \to \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$.

For the purposes of this paper, we use the following marking convention, as shown in Figure 3, arguments labeled `in` are shown in white, `out` in black, and `undec` in gray.

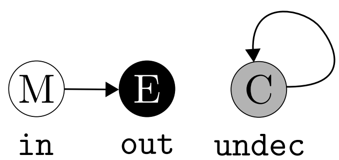

Figure 3: A labeling of an argumentation graph.

We write $\mathtt{in}(\mathcal{L})$ for the set of arguments that are labeled `in` by $\mathcal{L}$, $\mathtt{out}(\mathcal{L})$ for the set of arguments that are labeled `out` by $\mathcal{L}$, and $\mathtt{undec}(\mathcal{L})$ for the set of arguments that are labeled `undec` by $\mathcal{L}$. A labeling $\mathcal{L}$ can be represented as $\mathcal{L} = (\mathtt{in}(\mathcal{L}),\mathtt{out}(\mathcal{L}),\mathtt{undec}(\mathcal{L}))$. Equivalently, we also denote a labeling $\mathcal{L}$ as: $\mathcal{L} = \{(A,l) | \mathcal{L}(A) = l$ for all $A \in \mathcal{A}, l \in \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}\}$.

However, labelings should follow some given conditions. If an argument is labeled `in` then all of its defeaters are labeled `out`. If an argument is labeled `out` then at least one of its defeaters is labeled `in`. We call a labeling that follows the previous two conditions an *admissible labeling*.

Definition 3 (Admissible labeling [18, 14])**.** *Let* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ *be an argumentation framework. An admissible labeling is a mapping* $\mathcal{L} : \mathcal{A} \to \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$ *such that for each* $A \in \mathcal{A}$ *it holds that:*

if $\mathcal{L}(A) = \mathtt{in}$ *then* $\forall B \in \mathcal{A} : (B \rightharpoonup A \Rightarrow \mathcal{L}(B) = \mathtt{out})$, *and*
if $\mathcal{L}(A) = \mathtt{out}$ *then* $\exists B \in \mathcal{A} : (B \rightharpoonup A \wedge \mathcal{L}(B) = \mathtt{in})$.

Some examples for *admissible* labelings, in Figure 2, can include the following: $(\{A_1,A_3,A_5\}, \{A_2,A_4\}, \emptyset)$, $(\{A_3,A_5\}, \{A_2,A_4\}, \{A_1\})$, $(\{A_3\}, \{A_2\}, \{A_1,A_4,A_5\})$, $(\{A_5\}, \emptyset, \{A_1,A_2,A_3,A_4\})$, and $(\emptyset, \emptyset, \{A_1,A_2,A_3,A_4,A_5\})$.

One can realize that in an admissible labeling, unlike `in`-labeled and `out`-labeled arguments, `undec`-labels do not need to be justified i.e. an argument can be labeled `undec` under an admissible labeling without any condition.

The *complete* semantics, however, force undec-labels to be also justified. A *complete* labeling is an admissible labeling with the following extra condition: If an argument is labeled undec then there is no defeating argument that is labeled in (that is, there is insufficient ground to label the argument out) and not all defeating arguments are labeled out (that is, there is insufficient ground to label the argument in). We call a labeling which follows these rules a *complete labeling*.

Definition 4 (Complete labeling [18, 14])**.** *Let $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework. A complete labeling is a mapping $\mathcal{L} : \mathcal{A} \to \{\texttt{in}, \texttt{out}, \texttt{undec}\}$ such that for each $A \in \mathcal{A}$ it holds that:*
if $\mathcal{L}(A) = \texttt{in}$ *then* $\forall B \in \mathcal{A} : (B \rightharpoonup A \Rightarrow \mathcal{L}(B) = \texttt{out})$,
if $\mathcal{L}(A) = \texttt{out}$ *then* $\exists B \in \mathcal{A} : (B \rightharpoonup A \wedge \mathcal{L}(B) = \texttt{in})$*, and*
if $\mathcal{L}(A) = \texttt{undec}$ *then:*
$\neg[\forall B \in \mathcal{A} : (B \rightharpoonup A \Rightarrow \mathcal{L}(B) = \texttt{out})] \wedge \neg[\exists B \in \mathcal{A} : (B \rightharpoonup A \wedge \mathcal{L}(B) = \texttt{in})]$

As an example for a *complete* labeling, in Figure 2, we have only one complete labeling, namely $(\{A_1, A_3, A_5\}, \{A_2, A_4\}, \emptyset)$.

2.2 Aggregation Operators

Perhaps, the most common aggregation rules are the majority rules, in which an alternative is chosen if and only if it receives a number of votes that exceeds some presepecified threshold $k > 0.5 \times N$, where N is the number of voters. However, these rules are not always appropriate. One example is in juries, when the legal or the moral responsibility of the outcome is shared by all individuals. Indeed Ronnegard [43] argued that the attribution of moral responsibility to all members of a committee is legitimate when the decision is taken through unanimous voting, while it is not necessarily the case otherwise. Another example is when the outcome of the decision can potentially harm some individuals. It was shown in [8] that people show a preference for more conservative aggregation procedures when the outcome of the decision may involve the infliction of personal harm. Aiming to address such specific scenarios, Caminada and Pigozzi [15] proposed three aggregation rules that ensure the compatibility of the outcome with all individuals votes.

Before introducing the aggregation operators that were defined in [15], we first define the problem of aggregation. The problem of labeling aggregation can be formulated as a set of individuals that collectively decide how an argumentation framework $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ must be labelled.

Definition 5 (Labeling aggregation problem [3])**.** *Let $Ag = \{1, \ldots, n\}$ be a finite non-empty set of agents, and $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework. A labeling aggregation problem is a pair $\mathcal{LAP} = \langle Ag, \mathcal{AF} \rangle$.*

Each individual $i \in Ag$ has a labeling $\mathcal{L}_i$ which expresses the evaluation of $\mathcal{AF}$ by this individual. A labeling profile P is a set of the labelings submitted by agents in Ag: $P = \{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$.[4]

[4] We follow [15] in assuming that the profile is a set of labelings instead of a list of labelings. Although this is not common in judgment aggregation literature where the number of votes matter in many operators, it is not the case for the three operators considered in this study, since they focus on compatibility instead of cardinality. As such, although we list n labelings in the profile, it is possible that a profile has less than n elements, since agents can submit similar labelings.

A labeling aggregation operator is a function that maps a set of n labelings, chosen from the set of all labelings, $\mathtt{Labs}$, into a collective labeling.[5]

Definition 6 (Labeling aggregation operator $O_{\mathcal{AF}}$ [15]). *Let $\mathcal{LAP} = \langle Ag, \mathcal{AF} \rangle$ be a labeling aggregation problem. A labeling aggregation operator for $\mathcal{LAP}$ is a function $O_{\mathcal{AF}} : 2^{\mathtt{Labs}} \setminus \{\emptyset\} \rightarrow \mathtt{Labs}$ such that $O_{\mathcal{AF}}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}) = \mathcal{L}_{Coll}$, where $\mathcal{L}_{Coll}$ is the collective labeling.*

A labeling $\mathcal{L}_1$ is said to be *less or equally committed* than another labeling $\mathcal{L}_2$ if and only if every argument that is labeled $\mathtt{in}$ by $\mathcal{L}_1$ is also labeled $\mathtt{in}$ by $\mathcal{L}_2$ and every argument that is labeled $\mathtt{out}$ by $\mathcal{L}_1$ is also labeled $\mathtt{out}$ by $\mathcal{L}_2$.

Definition 7 (Less or equally committed $\sqsubseteq$ [15]). *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two labelings of argumentation framework $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We say that $\mathcal{L}_1$ is less or equally committed as $\mathcal{L}_2$ ($\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$) iff* $(\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2)) \wedge (\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2))$.

Two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ are said to be *compatible* with each other if and only if for every argument, there is no $\mathtt{in} - \mathtt{out}$ conflict between the two. In other words, every argument that is labeled $\mathtt{in}$ by $\mathcal{L}_1$ is not labeled $\mathtt{out}$ by $\mathcal{L}_2$ and every argument that is labeled $\mathtt{out}$ by $\mathcal{L}_1$ is not labeled $\mathtt{in}$ by $\mathcal{L}_2$.

Definition 8 (Compatible labelings $\approx$ [15]). *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two labelings of argumentation framework $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We say that $\mathcal{L}_1$ is compatible with $\mathcal{L}_2$ ($\mathcal{L}_1 \approx \mathcal{L}_2$) iff* $(\mathtt{in}(\mathcal{L}_1) \cap \mathtt{out}(\mathcal{L}_2) = \emptyset) \wedge (\mathtt{out}(\mathcal{L}_1) \cap \mathtt{in}(\mathcal{L}_2) = \emptyset)$

We now define a compatible operator as the following:

Definition 9 (Compatible operator). *Let $\mathcal{LAP} = \langle Ag, \mathcal{AF} \rangle$ be a labeling aggregation problem, and let $O_{\mathcal{AF}}$ be a labeling aggregation operator for $\mathcal{LAP}$. We say $O_{\mathcal{AF}}$ is a compatible operator if given any labeling profile $P = \{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$, $O_{\mathcal{AF}}(P) \approx \mathcal{L}_i, \forall i \in Ag$ i.e. the outcome of $O_{\mathcal{AF}}$ is compatible with each individual's labeling.*

In [15], Caminada and Pigozzi proposed three different aggregation operators, namely the skeptical operator, the credulous operator and the super credulous operator. Each of these operators maps a set of labelings, that are submitted by individuals, into a collective labeling. The following two definitions are used in the definition of these operators:

Definition 10 (Initial operators $\sqcap$, $\sqcup$ [15]). *Let $\mathcal{LAP} = \langle Ag, \mathcal{AF} \rangle$ be a labeling aggregation problem. The skeptical initial $\sqcap$ and credulous initial $\sqcup$ operators are labeling aggregation operators for $\mathcal{LAP}$ defined as the following:*

- $\sqcap(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}) = \{(A, \mathtt{in}) | \forall i \in Ag : \mathcal{L}_i(A) = \mathtt{in}\} \cup \{(A, \mathtt{out}) | \forall i \in Ag : \mathcal{L}_i(A) = \mathtt{out}\} \cup \{(A, \mathtt{undec}) | \exists i \in Ag : \mathcal{L}_i(A) \neq \mathtt{in} \wedge \exists j \in Ag : \mathcal{L}_j(A) \neq \mathtt{out}\}$

[5]Although it would be more precise to use $\mathtt{Labs}^S_{\mathcal{AF}}$ to denote the set of all labelings for $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ according to semantics S, we will often drop $\mathcal{AF}$ and S, and use $\mathtt{Labs}$ instead when there is no ambiguity about the argumentation framework. The same goes for all other notations (e.g. $O_{\mathcal{AF}}$) that were defined for an $\mathcal{AF}$, when there is no ambiguity about the argumentation framework.

- $\sqcup(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}) = \{(A,\texttt{in}) | \exists i \in Ag : \mathcal{L}_i(A) = \texttt{in} \wedge \neg\exists j \in Ag : \mathcal{L}_j(A) = \texttt{out}\} \cup \{(A,\texttt{out}) | \exists i \in Ag : \mathcal{L}_i(A) = \texttt{out} \wedge \neg\exists j \in Ag : \mathcal{L}_j(A) = \texttt{in}\} \cup \{(A,\texttt{undec}) | \forall i \in Ag : \mathcal{L}_i(A) = \texttt{undec} \vee (\exists j \in Ag : \mathcal{L}_j(A) = \texttt{in} \wedge \exists k \in Ag : \mathcal{L}_k(A) = \texttt{out})\}$ [6]

Definition 11 (Down-admissible $\downarrow$ and up-complete $\uparrow$ labelings [15])**.** *Let $\mathcal{L}$ be a labeling of argumentation framework $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. The down-admissible labeling of $\mathcal{L}$, denoted as $\mathcal{L}\downarrow$, is the biggest element of the set of all admissible labelings that are less or equally committed than $\mathcal{L}$:*

$$\forall \mathcal{L}' \in \texttt{Adms} \ : \ (\mathcal{L}' \sqsubseteq \mathcal{L} \Rightarrow \mathcal{L}' \sqsubseteq (\mathcal{L}\downarrow) \sqsubseteq \mathcal{L})$$

where Adms *is the set of all admissible labelings for $\mathcal{AF}$. The up-complete labeling of $\mathcal{L}$, denoted as $\mathcal{L}\uparrow$, is the smallest element of the set of all complete labelings that are bigger or equally committed than $\mathcal{L}$.*

$$\forall \mathcal{L}' \in \texttt{Comps} \ : \ (\mathcal{L} \sqsubseteq \mathcal{L}' \Rightarrow \mathcal{L} \sqsubseteq (\mathcal{L}\uparrow) \sqsubseteq \mathcal{L}')$$

Now, we provide the definitions of the three operators:

Definition 12 (Skeptical $so_{\mathcal{AF}}$, Credulous $co_{\mathcal{AF}}$ and Super Credulous $sco_{\mathcal{AF}}$ operators [15])**.** *Let $\mathcal{LAP} = \langle Ag, \mathcal{AF} \rangle$ be a labeling aggregation problem. The skeptical $so_{\mathcal{AF}}$, the credulous $co_{\mathcal{AF}}$ and super credulous $sco_{\mathcal{AF}}$ operators are labeling aggregation operators for $\mathcal{LAP}$ defined as the following:*

- $so_{\mathcal{AF}}(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}) = (\sqcap(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}))\downarrow$.
- $co_{\mathcal{AF}}(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}) = (\sqcup(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}))\downarrow$.
- $sco_{\mathcal{AF}}(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}) = ((\sqcup(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\}))\downarrow)\uparrow$.

Given the set of all admissible labelings Adms for some argumentation framework, it is shown that the outcome of the skeptical aggregation operator is the biggest element in Adms that is less or equally committed to every individual's labeling.

Theorem 1 ([15])**.** *Let $\mathcal{L}_1,\ldots,\mathcal{L}_n$ ($n \geq 1$) be labelings of argumentation framework $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. Let $\mathcal{L}_{SO}$ be $so_{\mathcal{AF}}(\{\mathcal{L}_1,\ldots,\mathcal{L}_n\})$. It holds that $\mathcal{L}_{SO}$ is the biggest admissible labeling such that for every $i \in Ag : \mathcal{L}_{SO} \sqsubseteq \mathcal{L}_i$.*

2.3 Distance Measures

In this part, we define the family of distance measures that we use to define preferences. Each of the distance measures we consider is characterized by three choices:

- Individual arguments vs. Issues (set of arguments).
- Set inclusion vs. Quantitative distance.
- Uniform vs. undec in the middle.

[6] We will often use $sio_{\mathcal{AF}}$ and $cio_{\mathcal{AF}}$ to refer to the skeptical initial and credulous initial operators, respectively.

The combination of all of these choices produces eight different distance measures. We start from the third choice. The uniform vs. `undec` in the middle choice captures the intuition that the distance between accepting an argument (`in`) and rejecting it (`out`) may be set as equal or superior to the distance of accepting (or rejecting) an argument (`in` or `out`) and abstaining on the same argument (`undec`). In other words, an `in`/`out` disagreement may be as serious or more serious (depending on the contexts) than a `in`/`undec` (or a `out`/`undec`) disagreement.

Thus, we consider the following two cases. First, `in`, `out`, and `undec` are equally distant from each other. In other words, $dist(\texttt{in},\texttt{out}) = dist(\texttt{dec},\texttt{undec})$, where $dist(.)$ is the difference between two labels for one argument, and `dec` is either `in` or `out`. In the other case, we assume that `undec` is in the middle between `in` and `out`. Thus, we differentiate between two types of disagreement. One between `in` and `out`, and the other between `dec` and `undec`. When considering distance, we assume $dist(\texttt{in},\texttt{out}) > dist(\texttt{dec},\texttt{undec})$.

2.3.1 Case 1: `in`, `out`, and `undec` are Equally Distant from Each Other

Hamming Set and Hamming Distance
The Hamming set between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the set of arguments that these two labelings disagree upon.

Definition 13 (Hamming Set $\ominus$)**.** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We define the Hamming set between these two labelings as:*

$$\mathcal{L}_1 \ominus \mathcal{L}_2 = \{A \in \mathcal{A} | \mathcal{L}_1(A) \neq \mathcal{L}_2(A)\} \tag{1}$$

The Hamming distance between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the number of arguments that these two labelings disagree upon.

Definition 14 (Hamming Distance $|\ominus|$)**.** *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We define the Hamming distance between these two labelings as:*

$$\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_2 = |\mathcal{L}_1 \ominus \mathcal{L}_2| \tag{2}$$

Issue-wise Set and Issue-wise Distance.
The label of an argument depends on the labels of the defeating arguments. Therefore, measuring the distance by treating arguments independently might not give an accurate sense of how far two labelings are from each other. Consider the example in Figure 4. Using Hamming distance, we have $\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_2 = \mathcal{L}_1 \,|\ominus|\, \mathcal{L}_3 = 4$.

However, one can argue that $\mathcal{L}_3$ is closer (than $\mathcal{L}_2$) to $\mathcal{L}_1$. Intuitively speaking, if $\mathcal{L}_1$ and $\mathcal{L}_3$ further agreed on the labeling of C (or D), then they would have been equivalent. On the other hand, $\mathcal{L}_1$ and $\mathcal{L}_2$ should further agree on E (or F) and G (or H) in order to become equivalent. In other words, the number of arguments whose labelings need to be switched in order to make the two labelings be equivalent is less between $\mathcal{L}_1$ and $\mathcal{L}_3$ than between $\mathcal{L}_1$ and $\mathcal{L}_2$.

Motivated by this example, Booth et al. [10] proposed a new distance method, using the notion of "issue", which they defined. This distance method captures the idea in the previous example,

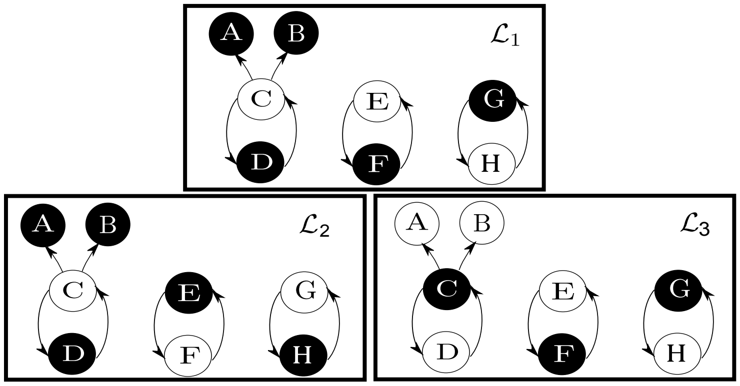

Figure 4: An example showing how Hamming distance is not a suitable distance measure for argument labeling [10].

while satisfying a set of axiomatic properties which they listed as essential for any distance measure.

Crucial to the definition of the "issue" is the concept of "in-sync". We say that two arguments A and B are *in-sync* if for any pair of labelings $\mathcal{L}, \mathcal{L}' \in \texttt{Labs}$, $\mathcal{L}(A)$ cannot be changed to $\mathcal{L}'(A)$ without causing a change of equal magnitude when moving from $\mathcal{L}(B)$ to $\mathcal{L}'(B)$, and vice versa.

Definition 15 (in-Sync $\equiv$ for semantics S [10])**.** *Let* $\texttt{Labs}^S$ *be the set of all labelings according to semantics S for argumentation framework* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$*. We say that two arguments* $A, B \in \mathcal{A}$ *are in-sync for semantics S (*$A \equiv^S B$*):*

$$A \equiv^S B \text{ iff } (A \equiv_1^S B \vee A \equiv_2^S B) \tag{3}$$

where:

- $A \equiv_1^S B$ *iff* $\forall \mathcal{L} \in \texttt{Labs}^S : \mathcal{L}(A) = \mathcal{L}(B)$.
- $A \equiv_2^S B$ *iff* $\forall \mathcal{L} \in \texttt{Labs}^S : (\mathcal{L}(A) = \texttt{in} \Leftrightarrow \mathcal{L}(B) = \texttt{out}) \wedge (\mathcal{L}(A) = \texttt{out} \Leftrightarrow \mathcal{L}(B) = \texttt{in})$

In-sync is an equivalence relation. We can partition the set of arguments in any argumentation framework $\mathcal{AF}$ into the *in-sync* equivalence classes, which form what is called *issues*.[7]

Definition 16 (Issue [10])**.** *Given the argumentation framework* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$*, a set of arguments* $\mathcal{B} \subseteq \mathcal{A}$ *is called an issue iff it forms an equivalence class of the relation* in-Sync ($\equiv$).

The Issue-wise set between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the set of issues that these two labelings disagree upon.

Definition 17 (Issue-wise Set $\ominus_{\mathcal{W}}$)**.** *Let* $\mathcal{L}_1$*,* $\mathcal{L}_2$ *be two labelings of* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ *and let* $\mathcal{I}$ *be the set of all issues in* $\mathcal{AF}$*. We define the Issue-wise set between these two labelings as:*

$$\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2 = \{\mathcal{B} \in \mathcal{I} | \mathcal{L}_1(A) \neq \mathcal{L}_2(A) \textit{ for some (equiv. all) } A \in \mathcal{B}\} \tag{4}$$

[7]The definition of *issue*, along with all the definitions depending on it, can be defined for semantics S (as the case for "in-sync"). However, from now on, we will restrict all of these definitions to the *complete* semantics, and drop the letter S. Thus, "issues" in what follows refers to the equivalnce classes of *in-sync* for the *complete* semantics.

Note that the sentence "for some (equiv. all)" follows from the definition of issues. One can realize that:

$$\forall \mathcal{L}_1, \mathcal{L}_2 \in \texttt{Labs}, \forall \mathcal{B} \in \mathcal{I} : (\exists A \in \mathcal{B}\ s.t.\ \mathcal{L}_1(A) \neq \mathcal{L}_2(A) \Leftrightarrow \forall A \in \mathcal{B} :\ \mathcal{L}_1(A) \neq \mathcal{L}_2(A)) \tag{5}$$

The Issue-wise distance between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the number of issues that these two labelings disagree upon.

Definition 18 (Issue-wise Distance $|\ominus_{\mathcal{W}}|$)**.** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We define the Issue-wise distance between these two labelings as:*

$$\mathcal{L}_1 \,|\ominus_{\mathcal{W}}|\, \mathcal{L}_2 = |\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2| \tag{6}$$

For example, in Figure 4, the Issue-wise sets between $\mathcal{L}_1$ and the other two labellings are:

$$\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2 = \{\{E,F\},\{G,H\}\}$$

$$\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_3 = \{\{A,B,C,D\}\}$$

While the corresponding Issue-wise distances are:

$$\mathcal{L}_1 \,|\ominus_{\mathcal{W}}|\, \mathcal{L}_2 = |\{\{E,F\},\{G,H\}\}| = 2$$

$$\mathcal{L}_1 \,|\ominus_{\mathcal{W}}|\, \mathcal{L}_3 = |\{\{A,B,C,D\}\}| = 1$$

2.3.2 Case 2: `undec` is in the Middle between `in` and `out`

In this section, we consider the case where `undec` is in the middle between `in` and `out`. Thus, we differentiate between two types of disagreement: 1) `in`/`out` disagreement, and 2) `dec`/`undec` disagreement. When considering distance, we assume $dist(\texttt{in},\texttt{out}) = 2 \times dist(\texttt{dec},\texttt{undec}) = 2$. [8]

To illustrate the difference from the previous case, consider the example shown in Figure 5. In this example, one can realize that the labelings $\mathcal{L}_2$ and $\mathcal{L}_3$ are equally distant from labeling $\mathcal{L}_1$ when considering Hamming set/distance or Issue-wise set/distance.

However, one can argue that $\mathcal{L}_3$ is closer than $\mathcal{L}_2$ to $\mathcal{L}_1$. Consider the arguments in Figure 5. Labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ seem to be on completely different sides regarding their evaluations for A and B. On the other hand, the difference between $\mathcal{L}_1$ and $\mathcal{L}_3$ is less drastic, because $\mathcal{L}_3$ abstains from taking any position about A and B.

We use IUO (short for In-Undec-Out i.e. Undec is in the middle) to denote this class of preferences.

[8] The use of 2 here is chosen carefully to satisfy the triangle inequality. However, the use of any α s.t. $1 < \alpha \leq 2$ would not affect the results of this paper. We just use 2 here for simplicity.

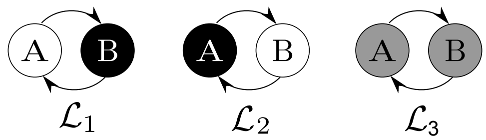

Figure 5: An example showing the need for considering `undec` as a middle labeling between `in` and `out`.

IUO Hamming Sets and IUO Hamming Distance
The $\texttt{in}-\texttt{out}$ Hamming set ($\ominus^{io}$) between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the set of arguments that both labelings label as decided (i.e. `in` or `out`), but on which they disagree upon. The $\texttt{dec}-\texttt{undec}$ Hamming set ($\ominus^{du}$) between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the set of arguments that one of the two labelings labels as decided (whether `in` or `out`) and the other labels as undecided.

Definition 19 (IUO Hamming sets $\ominus^{\mathcal{M}}$)**.** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We define the IUO Hamming sets as a pair $\ominus^{\mathcal{M}} = (\ominus^{io}, \ominus^{du})$, where $\ominus^{io}$ is $\texttt{in}-\texttt{out}$ Hamming set and $\ominus^{du}$ is $\texttt{dec}-\texttt{undec}$ Hamming set:*

$$\mathcal{L}_1 \ominus^{io} \mathcal{L}_2 = \{A \in \mathcal{A} | (\mathcal{L}_1(A) = \texttt{in} \wedge \mathcal{L}_2(A) = \texttt{out}) \vee (\mathcal{L}_1(A) = \texttt{out} \wedge \mathcal{L}_2(A) = \texttt{in})\} \tag{7}$$

$$\mathcal{L}_1 \ominus^{du} \mathcal{L}_2 = \{A \in \mathcal{A} | (A \in \texttt{dec}(\mathcal{L}_1) \wedge \mathcal{L}_2(A) = \texttt{undec}) \vee (\mathcal{L}_1(A) = \texttt{undec} \wedge A \in \texttt{dec}(\mathcal{L}_2))\} \tag{8}$$

where $\texttt{dec}(\mathcal{L}_1)$ is the set of decided (`in` or `out`) arguments according to the labeling $\mathcal{L}_1$.

The IUO Hamming distance between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the number of arguments in $\mathcal{L}_1 \ominus^{du} \mathcal{L}_2$ added to twice the number of the arguments in $\mathcal{L}_1 \ominus^{io} \mathcal{L}_2$.

Definition 20 (IUO Hamming Distance $\left|\ominus^{\mathcal{M}}\right|$)**.** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We define the IUO Hamming distance between these two labelings as:*

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_2 = 2 \times |\mathcal{L}_1 \ominus^{io} \mathcal{L}_2| + |\mathcal{L}_1 \ominus^{du} \mathcal{L}_2| \tag{9}$$

IUO Issue-wise Sets and IUO Issue-wise Distance
The $\texttt{in}-\texttt{out}$ Issue-wise set ($\ominus^{io}_{\mathcal{W}}$) between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the set of issues that both of the two labelings label as decided, but on which they disagree upon. The $\texttt{dec}-\texttt{undec}$ Issue-wise set ($\ominus^{du}_{\mathcal{W}}$) between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the set of issues that one of the two labelings labels as decided and the other labels as undecided.

Definition 21 (IUO Issue-wise sets $\ominus_{\mathcal{W}}^{\mathcal{M}}$)**.** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ and let $\mathcal{I}$ be the set of all issues in $\mathcal{AF}$. We define the IUO Issue-wise sets as $\ominus_{\mathcal{W}}^{\mathcal{M}} = (\ominus_{\mathcal{W}}^{io}, \ominus_{\mathcal{W}}^{du})$, where $\ominus_{\mathcal{W}}^{io}$ is the* $\texttt{in} - \texttt{out}$ *Issue-wise set and $\ominus_{\mathcal{W}}^{du}$ is the* $\texttt{dec} - \texttt{undec}$ *Issue-wise set:*

$$\mathcal{L}_1 \ominus_{\mathcal{W}}^{io} \mathcal{L}_2 = \{\mathcal{B} \in \mathcal{I} | (\mathcal{L}_1(A) = \texttt{in} \wedge \mathcal{L}_2(A) = \texttt{out}) \vee \\ (\mathcal{L}_1(A) = \texttt{out} \wedge \mathcal{L}_2(A) = \texttt{in}) \textit{ for some (equiv. all) } A \in \mathcal{B}\} \tag{10}$$

$$\mathcal{L}_1 \ominus_{\mathcal{W}}^{du} \mathcal{L}_2 = \{\mathcal{B} \in \mathcal{I} | (A \in \texttt{dec}(\mathcal{L}_1) \wedge \mathcal{L}_2(A) = \texttt{undec}) \vee \\ (\mathcal{L}_1(A) = \texttt{undec} \wedge A \in \texttt{dec}(\mathcal{L}_2)) \textit{ for some (equiv. all) } A \in \mathcal{B}\} \tag{11}$$

Note that given the definition of issues, for every labeling $\mathcal{L}$, an issue is either decided (all arguments in it are labeled `in` or `out` by $\mathcal{L}$) or undecided (all arguments in it are labeled undecided by $\mathcal{L}$):

$$\forall \mathcal{L} \in \texttt{Labs}, \forall \mathcal{B} \in \mathcal{I} : (\exists A \in \mathcal{B} \; s.t. \; A \in \texttt{dec}(\mathcal{L}) \Leftrightarrow \forall A \in \mathcal{B} : \; A \in \texttt{dec}(\mathcal{L})) \tag{12}$$

The IUO Issue-wise distance between two labelings $\mathcal{L}_1$ and $\mathcal{L}_2$ is the number of issues in $\mathcal{L}_1 \ominus_{\mathcal{W}}^{du} \mathcal{L}_2$ added to twice the number of the issues in $\mathcal{L}_1 \ominus_{\mathcal{W}}^{io} \mathcal{L}_2$.

Definition 22 (IUO Issue-wise Distance $\left|\ominus_{\mathcal{W}}^{\mathcal{M}}\right|$)**.** *Let $\mathcal{L}_1$, $\mathcal{L}_2$ be two labelings of $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$. We define the IUO Issue-wise distance between these two labelings as:*

$$\mathcal{L}_1 \left|\ominus_{\mathcal{W}}^{\mathcal{M}}\right| \mathcal{L}_2 = 2 \times |\mathcal{L}_1 \ominus_{\mathcal{W}}^{io} \mathcal{L}_2| + |\mathcal{L}_1 \ominus_{\mathcal{W}}^{du} \mathcal{L}_2| \tag{13}$$

For example, in Figure 5, the IUO Issue-wise sets between $\mathcal{L}_1$ and the other two labellings are:

$$\mathcal{L}_1 \ominus_{\mathcal{W}}^{io} \mathcal{L}_2 = \{\{A, B\}\}, \mathcal{L}_1 \ominus_{\mathcal{W}}^{du} \mathcal{L}_2 = \{\}$$

$$\mathcal{L}_1 \ominus_{\mathcal{W}}^{io} \mathcal{L}_3 = \{\}, \mathcal{L}_1 \ominus_{\mathcal{W}}^{du} \mathcal{L}_3 = \{\{A, B\}\}$$

While the corresponding IUO Issue-wise distances are:

$$\mathcal{L}_1 \left|\ominus_{\mathcal{W}}^{\mathcal{M}}\right| \mathcal{L}_2 = 2 \times |\{\{A, B\}\}| + 0 = 2$$

$$\mathcal{L}_1 \left|\ominus_{\mathcal{W}}^{\mathcal{M}}\right| \mathcal{L}_3 = 2 \times 0 + |\{\{A, B\}\}| = 1$$

Table 1 summarizes the distance measures we consider.

		Uniform	IUO
Hamming	**Set**	Hamming Set $\ominus$	IUO Hamming Sets $\ominus^{\mathcal{M}}$
	Distance	Hamming Distance $\lvert\ominus\rvert$	IUO Hamming Distance $\lvert\ominus^{\mathcal{M}}\rvert$
Issue-wise	**Set**	Issue-wise Set $\ominus_{\mathcal{W}}$	IUO Issue-wise Sets $\ominus_{\mathcal{W}}^{\mathcal{M}}$
	Distance	Issue-wise Distance $\lvert\ominus_{\mathcal{W}}\rvert$	IUO Issue-wise Distance $\lvert\ominus_{\mathcal{W}}^{\mathcal{M}}\rvert$

Table 1: Full family of distance measures.

2.4 Preferences

Given the distance measures defined earlier, we define agents' preferences. We say an agent's preferences are x-based, if her preferences are calculated using the distance measure x (e.g. Hamming distance based preferences). We use $\succeq_{i,x}$ to denote a *weak preference* relation by agent i whose preferences are x-based i.e. for any pair $\mathcal{L}_1, \mathcal{L}_2 \in \texttt{Labs}$, $\mathcal{L}_1 \succeq_{i,x} \mathcal{L}_2$ denotes that $\mathcal{L}_1$ is more or equally preferred than $\mathcal{L}_2$ by agent i with x-based preferences. Further, we use $\succ_{i,x}$ to denote a *strict preference* relation ($\mathcal{L}_1 \succ_{i,x} \mathcal{L}_2$ iff $(\mathcal{L}_1 \succeq_{i,x} \mathcal{L}_2) \wedge \neg(\mathcal{L}_2 \succeq_{i,x} \mathcal{L}_1)$), $\sim$ to denote an *incomparability* relation ($\mathcal{L}_1 \sim_{i,x} \mathcal{L}_2$ iff $\neg(\mathcal{L}_1 \succeq_{i,x} \mathcal{L}_2) \wedge \neg(\mathcal{L}_2 \succeq_{i,x} \mathcal{L}_1)$), and $\cong$ to denote an *indifference* relation ($\mathcal{L}_1 \cong_{i,x} \mathcal{L}_2$ iff $(\mathcal{L}_1 \succeq_{i,x} \mathcal{L}_2) \wedge (\mathcal{L}_2 \succeq_{i,x} \mathcal{L}_1)$).

We define the subset relation over pairs of sets as the following.

Definition 23 (Subset Over Pairs $\subseteq$)**.** *Let A_1, A_2, B_1, B_2 be four sets, and Let $S_1 = (A_1, B_1)$, $S_2 = (A_2, B_2)$ be two pairs of sets. We use $S_1 \subseteq S_2$ to denote the subset relation over pairs of subsets:*

$$S_1 \subseteq S_2 \textit{ iff } A_1 \subseteq A_2 \wedge B_1 \subseteq B_2 \tag{14}$$

Given a set measure $\otimes \in \{\ominus, \ominus^{\mathcal{M}}, \ominus_{\mathcal{W}}, \ominus_{\mathcal{W}}^{\mathcal{M}}\}$, an agent i, who has $\otimes$-set based preferences (and whose top preference is $\mathcal{L}_i$), would prefer a labelling $\mathcal{L}$ over another labeling $\mathcal{L}'$ if and only if the set of arguments in $\mathcal{L}_i \otimes \mathcal{L}$ is a subset of $\mathcal{L}_i \otimes \mathcal{L}'$ (where "subset" here refers to the standard definition of subset as well as the definition of "subset over pairs" defined above). Note that the set based preference yields a partial order over the labelings.[9]

Definition 24 (Set Based Preference $\succeq_{i,\otimes}$)**.** *We say that agent i's preferences are $\otimes$-set based w.r.t $\mathcal{L}_i$ iff:*

$$\forall \mathcal{L}, \mathcal{L}' \in \texttt{Labs} : \mathcal{L} \succeq_{i,\otimes} \mathcal{L}' \Leftrightarrow \mathcal{L} \otimes \mathcal{L}_i \subseteq \mathcal{L}' \otimes \mathcal{L}_i \tag{15}$$

where $\mathcal{L}_i$ is agent i's most preferred labeling and $\otimes \in \{\ominus, \ominus^{\mathcal{M}}, \ominus_{\mathcal{W}}, \ominus_{\mathcal{W}}^{\mathcal{M}}\}$. Note that $\otimes$-set based preferences is read Hamming set based preferences when $\otimes = \ominus$, Issue-wise set based preferences when $\otimes = \ominus_{\mathcal{W}}$, ...etc.

Given a distance measure $|\otimes| \in \{|\ominus|, |\ominus^{\mathcal{M}}|, |\ominus_{\mathcal{W}}|, |\ominus_{\mathcal{W}}^{\mathcal{M}}|\}$, an agent i, who has $|\otimes|$-distance based preferences (and whose top preference is $\mathcal{L}_i$), would prefer a labelling $\mathcal{L}$ over another labeling $\mathcal{L}'$ if and only if $\mathcal{L}_i \,|\otimes|\, \mathcal{L}$ is less than $\mathcal{L}_i \,|\otimes|\, \mathcal{L}'$. Note that the distance based preference yields a total pre-order over the labelings.

[9] Although formally, the set-based criteria are not measures but mappings to sets, we will slightly abuse terminology and refer to all criteria (set based and distance based) as *set and distance measures* for easy reference.

We now define the classes of preferences which are based on different distance measures, that we defined earlier.

Definition 25 (Distance Based Preference $\succeq_{i,|\otimes|}$). *We say that agent i's preferences are $|\otimes|$-distance based w.r.t $\mathcal{L}_i$ iff:*

$$\forall \mathcal{L}, \mathcal{L}' \in \texttt{Labs} : \mathcal{L} \succeq_{i,|\otimes|} \mathcal{L}' \Leftrightarrow \mathcal{L} \,|\otimes|\, \mathcal{L}_i \leq \mathcal{L}' \,|\otimes|\, \mathcal{L}_i \tag{16}$$

where $\mathcal{L}_i$ is agent i's most preferred labeling and $|\otimes| \in \{|\ominus|, |\ominus^{\mathcal{M}}|, |\ominus_{\mathcal{W}}|, |\ominus^{\mathcal{M}}_{\mathcal{W}}|\}$. Note that $|\otimes|$-distance based preferences is read Hamming distance based preferences when $|\otimes| = |\ominus|$, Issue-wise distance based preferences when $|\otimes| = |\ominus_{\mathcal{W}}|$, ...etc.

To illustrate the set and distance based preferences, we use Hamming set and Hamming distance based preferences for their simplicity. Consider the example in Figure 6 with four possible complete labelings. The Hamming sets between $\mathcal{L}_1$ and the other three labelings are:

$$\mathcal{L}_1 \ominus \mathcal{L}_2 = \{A, B\}$$

$$\mathcal{L}_1 \ominus \mathcal{L}_3 = \{C, D, E\}$$

$$\mathcal{L}_1 \ominus \mathcal{L}_4 = \{A, B, C, D, E\}$$

Consequently, the Hamming distance values between $\mathcal{L}_1$ and the other three labelings are the cardinality values of the Hamming sets between $\mathcal{L}_1$ and the other three labelings.

$$\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_2 = |\{A, B\}| = 2$$

$$\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_3 = |\{C, D, E\}| = 3$$

$$\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_4 = |\{A, B, C, D, E\}| = 5$$

Assume we have agents with Hamming set based preferences. Hence, an arbitrary agent i who prefers $\mathcal{L}_1$ the most, would have the following preferences: $\mathcal{L}_1 \succ \mathcal{L}_2 \succ \mathcal{L}_4$ and $\mathcal{L}_1 \succ \mathcal{L}_3 \succ \mathcal{L}_4$ (neither $\mathcal{L}_1 \ominus \mathcal{L}_2$ nor $\mathcal{L}_1 \ominus \mathcal{L}_3$ is a subset of the other). However, if agents have Hamming distance based preferences, an agent who prefers $\mathcal{L}_1$ the most, would have the following preferences: $\mathcal{L}_1 \succ \mathcal{L}_2 \succ \mathcal{L}_3 \succ \mathcal{L}_4$.

We can now examine the examples in Figure 4 and 5 in the light of preferences. The example in Figure 4 shows how an agent i whose top preference is $\mathcal{L}_i = \mathcal{L}_1$ would have different opinions about other labelings given the different distance measures used. If agent i has Hamming distance based preferences, then:

$$\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_2 = |\{E, F, G, H\}| = 4$$

$$\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_3 = |\{A, B, C, D\}| = 4$$

then, her preferences would be $\mathcal{L}_1 \succ_{i,|\ominus|} \mathcal{L}_2 \cong_{i,|\ominus|} \mathcal{L}_3$, while if she has Issue-wise distance based preferences, then:

$$\mathcal{L}_1 \,|\ominus_{\mathcal{W}}|\, \mathcal{L}_2 = |\{\{E, F\}, \{G, H\}\}| = 2$$

$$\mathcal{L}_1 \,|\ominus_{\mathcal{W}}|\, \mathcal{L}_3 = |\{\{A, B, C, D\}\}| = 1$$

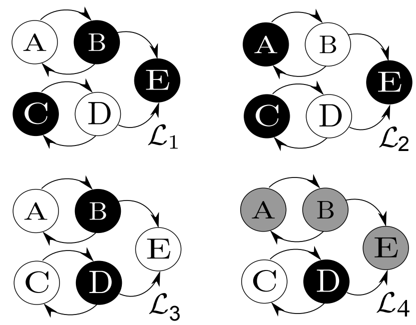

Figure 6: An argumentation graph with four possible complete labelings.

then, her preferences would be $\mathcal{L}_1 \succ_{i,|\ominus_{\mathcal{W}}|} \mathcal{L}_3 \succ_{i,|\ominus_{\mathcal{W}}|} \mathcal{L}_2$. Hence, it is interesting to introduce the "Issue-wise" concept to define a new class of preferences.

The example in Figure 5 shows how an agent i (whose top preference is $\mathcal{L}_1$) would have different preferences depending on whether `in`, `out`, and `undec` are equally distant, or `undec` is in the middle between `in` and `out`. In the former case:

$$\mathcal{L}_1 \, |\ominus| \, \mathcal{L}_2 = |\{A,B\}| = 2$$

$$\mathcal{L}_1 \, |\ominus| \, \mathcal{L}_3 = |\{A,B\}| = 2$$

and

$$\mathcal{L}_1 \, |\ominus_{\mathcal{W}}| \, \mathcal{L}_2 = |\{\{A,B\}\}| = 1$$

$$\mathcal{L}_1 \, |\ominus_{\mathcal{W}}| \, \mathcal{L}_3 = |\{\{A,B\}\}| = 1$$

hence, $\mathcal{L}_1 \succ_{i,|\ominus|} \mathcal{L}_2 \cong_{i,|\ominus|} \mathcal{L}_3$ and $\mathcal{L}_1 \succ_{i,|\ominus_{\mathcal{W}}|} \mathcal{L}_2 \cong_{i,|\ominus_{\mathcal{W}}|} \mathcal{L}_3$. In the latter case:

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_2 = 2 \times |\{A,B\}| + 0 = 4$$

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_3 = 2 \times 0 + |\{A,B\}| = 2$$

and

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}_{\mathcal{W}}\right| \mathcal{L}_2 = 2 \times |\{\{A,B\}\}| + 0 = 2$$

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}_{\mathcal{W}}\right| \mathcal{L}_3 = 2 \times 0 + |\{\{A,B\}\}| = 1$$

hence, $\mathcal{L}_1 \succ_{i,|\ominus^{\mathcal{M}}|} \mathcal{L}_3 \succ_{i,|\ominus^{\mathcal{M}}|} \mathcal{L}_2$ and $\mathcal{L}_1 \succ_{i,|\ominus^{\mathcal{M}}_{\mathcal{W}}|} \mathcal{L}_3 \succ_{i,|\ominus^{\mathcal{M}}_{\mathcal{W}}|} \mathcal{L}_2$. Thus, it is interesting to define Hamming and Issue based preferences with `undec` as a middle labeling.

3 Pareto Optimality

In this section, we study the Pareto optimality of the outcomes of the three operators given different variations of the preferences. Pareto optimality is one of the fundamental concepts that ensures that, given a profile, the social outcome selected by the aggregation procedure cannot be improved.

A labeling $\mathcal{L}_1$ Pareto dominates $\mathcal{L}_2$ if and only if for any agent i, i would prefer $\mathcal{L}_1$ at least as much as she prefers $\mathcal{L}_2$, and for at least one agent j, j would strictly prefer $\mathcal{L}_1$ over $\mathcal{L}_2$.

Definition 26 (Pareto dominance)**.** *Let $Ag = \{1,\ldots,n\}$ be a set of agents with preferences $\succeq_i$, $i \in Ag$. $\mathcal{L}$ Pareto dominates $\mathcal{L}'$ iff $\forall i \in Ag$, $\mathcal{L} \succeq_i \mathcal{L}'$ and $\exists j \in Ag$, $\mathcal{L} \succ_j \mathcal{L}'$.*

A labeling is Pareto optimal in a set, if it is not Pareto dominated by any other labeling from that set.

Definition 27 (Pareto optimality of a labeling in $\mathcal{S}$)**.** *Let $\mathcal{S}$ be a set of labelings. A labeling $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ if there is no labeling $\mathcal{L}' \in \mathcal{S}$ such that $\mathcal{L}'$ Pareto dominates $\mathcal{L}$.*

In our results, the set $\mathcal{S}$ will mainly refer to a set of *admissible* (or *complete*) labelings that are *compatible with* (or *smaller or equal to*) each of the participants' labelings. Moreover, whenever we refer to an operator as Pareto optimal (in a set $\mathcal{S}$) we mean that it only produces Pareto optimal outcomes (in $\mathcal{S}$).

Definition 28 (Pareto optimality of an operator in $\mathcal{S}$)**.** *Let $\mathcal{S}$ be a set of labelings. An operator is Pareto optimal in $\mathcal{S}$ if it only produces Pareto optimal (in $\mathcal{S}$) outcomes.*

3.1 Connections between Classes of Preferences

3.1.1 General Connections

We start our analysis by noticing that some of the defined types are in fact equivalent. Consider the following lemma which establishes the equivalence between some types of preferences:

Lemma 1. *Issue-wise set based preferences coincide with Hamming set based preferences, and IUO Issue-wise sets based preferences coincide with IUO Hamming sets based preferences.*

Further, we notice that Pareto optimality carries over from each of the distance-based preferences to its corresponding set-based preferences.

Theorem 2. *Let $\otimes \in \{\ominus, \ominus^{\mathcal{M}}, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}_{\mathcal{W}}\}$ be a set measure and $|\otimes|$ be its corresponding distance measure (i.e. if $\otimes = \ominus^{\mathcal{M}}$ then $|\otimes| = |\ominus^{\mathcal{M}}|$). If a labeling[10] is Pareto optimal in a set $\mathcal{S}$ given agents with $|\otimes|$-based preferences, then it is Pareto optimal in $\mathcal{S}$ given agents with $\otimes$-based preferences.*

Unfortunately, these connections are only one-way. A counterexample for the opposite way is provided in the Appendix.

[10]Note that since an operator is Pareto optimal in a set if and only if all of its outcomes are Pareto optimal in that set, then one can see that in this theorem, and others as well, 'labeling' can be substituted with 'operator'.

3.1.2 Connections in Special Contexts

The implications shown in the previous part hold without restrictions. However, when all labelings in $\mathcal{S}$ are admissible labelings and are compatible ($\approx$) with each of the individuals' labelings, one can find further connections.

Theorem 3. *Let $\mathcal{X}$ be the set of all admissible labelings that are compatible ($\approx$) with each of the participants' individual labelings. Let $\mathcal{S}$ be any arbitrary set such that $\mathcal{S} \subseteq \mathcal{X}$. A labeling from $\mathcal{S}$ is Pareto optimal in $\mathcal{S}$ when individual preferences are Hamming set (resp. distance) based iff it is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Hamming sets (resp. distance) based.*

Theorem 4. *Let $\mathcal{X}$ be the set of all admissible labelings that are compatible ($\approx$) with each of the participants' individual labelings. Let $\mathcal{S}$ be any arbitrary set such that $\mathcal{S} \subseteq \mathcal{X}$. A labeling from $\mathcal{S}$ is Pareto optimal in $\mathcal{S}$ when individual preferences are Issue-wise set (resp. distance) based iff it is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Issue-wise sets (resp. distance) based.*

Note that unlike in the previous part where connections are one-way (from distance based to set based, but not vice versa), the connections in this part hold in both directions.

3.1.3 Failed Connections

Given the findings so far, one might wonder about the existence of other connections among the eight classes of preferences. Unfortunately, other than the ones found above, there exists no connection, even after considering further restrictions, similar to the ones in the previous part. In the Appendix, we provide counterexamples for the connections that do not hold between the classes of preferences.

Basically, we provide an example for an argumentation framework in which Pareto optimality is satisfied when agents' preferences are Hamming set based, Issue-wise set based, Issue-wise distance based, IUO Hamming set based, IUO Issue-wise set based, and IUO Issue-wise distance based. However, Pareto optimality is not satisfied when agents' preferences are Hamming distance based, or IUO Hamming distance based. This shows that Pareto optimality given Hamming distance based preferences and IUO Hamming distance based preferences cannot be inferred from the other six classes of preferences.

In a similar way, we provide an example for an argumentation framework which shows that Pareto optimality given Issue-wise distance based preferences and IUO Issue-wise distance based preferences cannot be inferred from the other six classes of preferences.

We summarize all the findings in Table 2. For each cell in the table, a Y means that for any operator, Pareto optimality in any arbitrary set $\mathcal{S}$ carries over from the preference class in the row to the preference class in the column (in the same set $\mathcal{S}$), while a Y^* means that it only holds for compatible operators (i.e. that produce compatible outcomes).

Now we turn to studying the Pareto optimality of the three operators: the skeptical, the credulous and the super credulous, with respect to the eight classes of preferences.

	HS/IwS	HD	IwD	IUO HS/IwS	IUO HD	IUO IwD
Hamming set (HS) **Issue-wise set (IwS)**	Y	N	N	Y*	N	N
Hamming dist. (HD)	Y	Y	N	Y*	Y*	N
Issue-wise dist. (IwD)	Y	N	Y	Y*	N	Y*
IUO Hamming sets (IUO HS) **IUO Issue-wise sets (IUO IwS)**	Y*	N	N	Y	N	N
IUO Hamming dist. (IUO HD)	Y*	Y*	N	Y	Y	N
IUO Issue-wise dist. (IUO IwD)	Y*	N	Y*	Y	N	Y

Table 2: Pareto optimality relations between the different preference classes. A Y means Pareto optimality carries over from the class in the row to the class in the column, and a Y^* means it only carries over if the operator only produces compatible labelings.

3.2 Case 1: `in`, `out`, and `undec` are Equally Distant from Each Other

3.2.1 Hamming Set and Hamming Distance

In this part, we establish the first advantage of the skeptical operator over the credulous and super credulous operators. When all individuals' preferences are Hamming set based, or all are Hamming distance based, the skeptical operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each individual's labeling.

Theorem 5. *If individual preferences are Hamming set based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings.*

Theorem 6. *If individual preferences are Hamming distance based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings.*

On the other hand, the credulous and super credulous operators are only Pareto optimal when individuals have Hamming set based preferences, and they fail to produce Pareto optimal outcomes when the preferences are Hamming distance based.

Theorem 7. *If individual preferences are Hamming set based, then the credulous aggregation operator is Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Theorem 8. *If individual preferences are Hamming set based, then the super credulous aggregation operator is Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

Observation 1. *If individual preferences are Hamming distance based, then the credulous (resp. the super credulous) aggregation operator is not Pareto optimal in the set of admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings.*

3.2.2 Issue-wise Set and Issue-wise Distance

Given Lemma 1, we can substitute "Hamming set" with "Issue-wise set" in Theorem 5, Theorem 7 and Theorem 8.

Unfortunately, Lemma 1 only concerns the implication from Hamming set to Issue-wise set based preferences and vice versa. One might wonder if a similar lemma can be shown for the case with Hamming distance and Issue-wise distance based preferences. However, we show in Examples 1 and 2 in the Appendix, that this is not necessarily the case. As such, one has to show whether Pareto optimality results hold or not for Issue-wise distance based preferences independently from those of the Hamming distance based preferences.

Doing so confirms the superiority of the skeptical operator for the Issue-wise distance-based preferences. As we show next, when agents preferences are Issue-wise distance based, only the skeptical aggregation operator is guaranteed to produce Pareto optimal outcomes.

Theorem 9. *If individual preferences are Issue-wise distance based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal (w.r.t $\sqsubseteq$) to each of the participants' labelings.*

Observation 2. *If individual preferences are Issue-wise distance based, then the credulous (resp. the super credulous) aggregation operator is not Pareto optimal in the set of admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings.*

3.3 Case 2: `undec` is in the Middle between `in` and `out`

We now analyze the Pareto optimality for the three operators given the classes of preferences that assume `undec` to be in the middle between `in` and `out` ($dist(\texttt{dec}, \texttt{undec}) < dist(\texttt{in}, \texttt{out})$). We show that for the three considered operators, it is possible to show an equivalence with the results of Section 3.2.

3.3.1 IUO Hamming Sets and IUO Hamming Distance

IUO Hamming sets (resp. IUO Hamming distance) differs from the Hamming set (resp. Hamming distance) in that the former separates `in/out` disagreement from `dec/undec` disagreement. We use Theorem 3 to show the results for this part.

Proposition 1. *If individual preferences are IUO Hamming sets (resp. distance) based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings.*

As for the credulous and super credulous operators, their results given IUO Hamming set and distance based preferences echo their results with the Hamming set and distance based preferences. Both credulous and super credulous produce Pareto optimal outcomes given Hamming set based preferences, but can fail to produce Pareto optimal outcomes given Hamming distance based preferences.

Proposition 2. *If individual preferences are IUO Hamming sets based, then the credulous aggregation operator is Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proposition 3. *If individual preferences are IUO Hamming distance based, then the credulous aggregation operator is not Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proposition 4. *If individual preferences are IUO Hamming sets based, then the super credulous aggregation operator is Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proposition 5. *If individual preferences are IUO Hamming distance based, then the super credulous aggregation operator is not Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

3.3.2 IUO Issue-wise Sets and IUO Issue-wise Distance

Again, given the discussion earlier as a result of Lemma 1 (i.e. IUO Issue-wise sets based preferences coincide with IUO Hamming sets based preferences), we can substitute "IUO Hamming sets" with "IUO Issue-wise sets" in Proposition 1, Proposition 2 and Proposition 4.

Next, we show the results for the IUO Issue-wise distance based preferences.

Proposition 6. *If individual preferences are IUO Issue-wise distance based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings.*

We now turn to the credulous aggregation operator.

Proposition 7. *If individual preferences are IUO Issue-wise distance based, then the credulous aggregation operator is not Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Finally, we turn to the super credulous aggregation operator.

Proposition 8. *If individual preferences are IUO Issue-wise distance based, then the super credulous aggregation operator is not Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

Table 3 summarizes the Pareto optimality results for the three operators given all the eight classes of preferences.

	Skeptical Operator	**Credulous Operator**	**Super Credulous Operator**
Hamming set Issue-wise set	Yes (Thm. 5)	Yes (Thm. 7)	Yes (Thm. 8)
Hamming dist.	Yes (Thm. 6)	No (Obs. 1)	No (Obs. 1)
Issue-wise dist.	Yes (Thm. 9)	No (Obs. 2)	No (Obs. 2)
IUO Hamming sets IUO Issue-wise sets	Yes (Prop. 1)	Yes (Prop. 2)	Yes (Prop. 4)
IUO Hamming dist.	Yes (Prop. 1)	No (Prop. 3)	No (Prop. 5)
IUO Issue-wise dist.	Yes (Prop. 6)	No (Prop. 7)	No (Prop. 8)

Table 3: Pareto optimality in a set $\mathcal{S}$ of the aggregation operators depending on the type of preference. The set $\mathcal{S}$ differs for each operator. For the skeptical operator, it is the set of all admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings. For the credulous (resp. super credulous) operator, it is the set of all admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings.

3.4 Heterogeneous Preferences

The previous subsections have all considered the case where agents have homogeneous preferences i.e. agents share the same class of preferences (e.g. all agents have Hamming set based preferences). However, there can be some scenarios where this assumption does not hold. In this part, we study the effect of removing this assumption.

Let $\mathcal{F}$ be the set of all classes of preferences, $\mathcal{R}$ be some arbitrary subset of $\mathcal{F}$, and $c : Ag \to \mathcal{F}$ be a function defining the class of preferences for each agent. We say that the set of agents Ag have *homogeneous* preferences from $\mathcal{R}$ if $\forall i, j \in Ag : c(i) = c(j) \in \mathcal{R}$. We say Ag have *heterogeneous* preferences from $\mathcal{R}$ if $\forall i \in Ag : c(i) \in \mathcal{R}$ and $\exists i, j \in Ag$ s.t. $c(i) \neq c(j)$.

Let $\mathcal{R}$ be an arbitrary set of classes of preferences. In general, if a labeling $\mathcal{L}$ is Pareto optimal in a set $\mathcal{S}$ given that Ag have *homogeneous* preferences from $\mathcal{R}$, then $\mathcal{L}$ might not be Pareto optimal if Ag have *heterogeneous* preferences from $\mathcal{R}$ (see Example 3 in the Appendix).

However, one can show that some of the classes of preferences that we defined enjoy special relations with each others that make Pareto optimality carry over from homogeneous preference of each of those classes to heterogeneous preferences that combine all of those classes. Consider the following theorem.

Theorem 10. *Let $\mathcal{R} = \{\ominus, \ominus^{\mathcal{M}}\}$ be a set of preference classes, Ag be a set of agents, $\mathcal{S}$ be the set of all admissible labelings that are compatible with each individual's labeling, and $\mathcal{L}$ be a labeling from $\mathcal{S}$. If $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ given that Ag have* homogeneous *preferences from $\mathcal{R}$, then $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ given that Ag have* heterogeneous *preferences from $\mathcal{R}$.*

Also, given our discussion in Lemma 1, we can add Issue-wise set $\ominus_{\mathcal{W}}$ and IUO Issue-wise sets $\ominus^{\mathcal{M}}_{\mathcal{W}}$ to the set $\mathcal{R}$ in the previous result. For our three operators, we have the following corollaries.

Corollary 1. *Let* $\mathcal{R} = \{\ominus, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}, \ominus^{\mathcal{M}}_{\mathcal{W}}\}$*. The skeptical operator is Pareto optimal in the set of all admissible labelings that are smaller or equal (*$\sqsubseteq$*) to each of the participants' labelings given that individuals have heterogeneous preferences from* $\mathcal{R}$*.*

Corollary 2. *Let* $\mathcal{R} = \{\ominus, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}, \ominus^{\mathcal{M}}_{\mathcal{W}}\}$*. The credulous operator is Pareto optimal in the set of all admissible labelings that are compatible (*$\approx$*) with each of the participants' labelings given that individuals have heterogeneous preferences from* $\mathcal{R}$*.*

Corollary 3. *Let* $\mathcal{R} = \{\ominus, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}, \ominus^{\mathcal{M}}_{\mathcal{W}}\}$*. The super credulous operator is Pareto optimal in the set of all complete labelings that are compatible (*$\approx$*) with each of the participants' labelings given that individuals have heterogeneous preferences from* $\mathcal{R}$*.*

We showed earlier that the skeptical operator is always Pareto optimal no matter which class of preferences the individuals have, as long as all agents have the same class i.e. homogenous preferences (as Table 3 shows). We show here even a stronger result, that is even when agents preferences are hererogenous, and no matter what the combination of classes of preferences that they have, the skeptical operator sustains Pareto optimality. This establishes the robustness of the skeptical operator when it comes to Pareto optimality.

Theorem 11. *Let* $\mathcal{R} = \{\ominus, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}, \ominus^{\mathcal{M}}_{\mathcal{W}}, |\ominus|, |\ominus_{\mathcal{W}}|, |\ominus^{\mathcal{M}}|, |\ominus^{\mathcal{M}}_{\mathcal{W}}|\}$*. The skeptical operator is Pareto optimal in the set of all admissible labelings that are smaller or equal (*$\sqsubseteq$*) to each of the participants' labelings given that individuals have heterogeneous preferences from* $\mathcal{R}$*.*

4 Strategy Proofness

Strategic manipulability is usually an undesirable property in which an agent, upon knowing the preferences of other individuals, has incentive to misrepresent her own true opinion in order to force a collective outcome which is closer to her true opinion. A strategic lie is what an agent can say if and when she has the opportunity to vote strategically.

Definition 29 (Strategic lie)**.** *Let* P *be a profile and* $\mathcal{L}_k \in P$ *be the most preferred labeling of an agent with preference* $\succeq_k$*. Let* Op *be any aggregation operator. A labeling* $\mathcal{L}'_k$ *such that* $Op(P_{\mathcal{L}_k/\mathcal{L}'_k}) \succ_k Op(P)$ *is called a strategic lie. Where* $P_{\mathcal{L}_k/\mathcal{L}'_k}$ *is the profile that results from the profile* P *after agent* k *changes her vote from* $\mathcal{L}_k$ *to* $\mathcal{L}'_k$*.*

A strategy proof operator is one where individuals have no incentive to make strategic lies.

Definition 30 (Strategy proof operator)**.** *An aggregation operator* Op *is strategy proof if strategic lies are not possible.*

Despite the fact that, as we shall see, for most classes of preference, the aggregation operators turned out to be vulnerable to strategic manipulation, a novel type of lie emerged: the benevolent lie. Unlike the malicious lie, the benevolent lie has positive effects on some of the other agents and no negative effects on any agent.

Definition 31 (Malicious lie)**.** *Let Op be some aggregation operator and P be a profile of labelings. We say that a strategic lie $\mathcal{L}'_k$ is malicious iff, for some agent $j \neq k$, $Op(P) \succ_j Op(P_{\mathcal{L}_k/\mathcal{L}'_k})$.*

Definition 32 (Benevolent lie)**.** *Let Op be some aggregation operator and P be a profile of labelings. We say that a strategic lie $\mathcal{L}'_k$ is benevolent iff, for any agent i $Op(P_{\mathcal{L}_k/\mathcal{L}'_k}) \succeq_i Op(P)$ and there exists an agent $j \neq k$, $Op(P_{\mathcal{L}_k/\mathcal{L}'_k}) \succ_j Op(P)$.*

4.1 Connections between Classes of Preferences

Similar to the previous section, we start by showing general connections. First, note that the results of Lemma 1 holds in this case. This means that the benevolence property (that is, all strategic lies are benevolent) carries over between Hamming set and Issue-wise set based preferences. The same goes for IUO Hamming set and IUO Issue-wise set based preferences. Fruther, we show that this benevolence property carries over from Hamming distance to IUO Hamming distance based preferences, and from Issue-wise distance to IUO Issue-wise distance based preferences.

Theorem 12. *Consider an operator Op that only produces labelings that are compatible ($\approx$) with each individual's labeling. If all strategic lies are benevolent when agents have Hamming distance based preferences then all strategic lies are benevolent when agents have IUO Hamming distance based preferences.*

Theorem 13. *Consider an operator Op that only produces labelings that are compatible ($\approx$) with each individual's labeling. If all strategic lies are benevolent when agents have Issue-wise distance based preferences then all strategic lies are benevolent when agents have IUO Issue-wise distance based preferences.*

Now we turn to studying the strategy proofness of the three operators: the skeptical, the credulous and the super credulous.

4.2 Case 1: `in`, `out`, and `undec` are Equally Distant from Each Other

4.2.1 Hamming Set and Hamming Distance

Following, we show that none of the three operators is strategy proof given Hamming set (resp. Hamming distance) based preferences.

Observation 3. *The skeptical aggregation operator is not strategy proof for neither Hamming set nor Hamming distance based preferences.*

Observation 4. *The credulous (resp. super credulous) aggregation operator is not strategy proof for neither Hamming set nor Hamming distance based preferences.*

For the skeptical aggregation operator, however, every strategic lie is benevolent, given Hamming set (resp. Hamming distance) based preferences. Unfortunately, this is not the case for the credulous or the super credulous operators.

Theorem 14. *Consider the skeptical aggregation operator and Hamming set based preferences. For any agent, her strategic lies are benevolent.*

Theorem 15. *Consider the skeptical aggregation operator and Hamming distance based preferences. For any agent, her strategic lies are benevolent.*

Note that the previous two theorems raise an interesting point. Given the Pareto optimality of the skeptical operator for Hamming set/distance based preferences, one would expect that benevolent lies are not possible. Otherwise, it means there exists another labeling that is more preferred by every agent and strictly preferred by at lease one agent. This contradicts the Pareto optimality result found earlier.

However, it is important to remember that the Pareto optimality results found earlier are all with respect to the sets of labelings that are smaller or equal (or compatible in the case of the other operators) to each individuals' labelings. On the other hand, an outcome given a benevolent lie is not compatible with every individual's labeling i.e. while the skeptical operator does produce labelings that are compatible with each individual's true labeling, it does so for the submitted labelings only. Hence, when an agent k lies and submits $\mathcal{L}'_k$ instead of $\mathcal{L}_k$, the outcome $\mathcal{L}'_{SO}$ (which is the outcome when k submits $\mathcal{L}'_k$) is compatible with $\mathcal{L}'_k$ but not necessarily to $\mathcal{L}_k$. As a result, the labeling $\mathcal{L}'_{SO}$ does not belong to the set of labelings that $\mathcal{L}_{SO}$ is compared to when studying Pareto optimality.

This highlights another interesting point that can be implied by the benevolence and Pareto optimality of the skeptical operator. When using the skeptical operator, whenever an agent k considers lying in order to get a closer labeling to $\mathcal{L}_k$, she is faced with an inevitable trade-off between getting a less or equally committed outcome and getting a closer (i.e. more preferred) outcome.

4.2.2 Issue-wise Set and Issue-wise Distance

Similar to the Hamming based preferences, none of the three operators is strategy proof given Issue-wise set (resp. Issue-wise distance) based preferences.

Observation 5. *The skeptical aggregation operator is not strategy proof for neither Issue-wise set nor Issue-wise distance based preferences.*

Observation 6. *The credulous (resp. super credulous) aggregation operator is not strategy proof for neither Issue-wise set nor Issue-wise distance based preferences.*

Again, similar to the Hamming based preference, only the skeptical aggregation operator has the benevolent property (every strategic lie is benevolent), given Issue-wise set (resp. Issue-wise distance) based preferences. As for the Issue-wise set based preferences, we can use the result of Lemma 1 (that is, Issue-wise set based preferences coincide with Hamming set based preferences) to substitute "Hamming set" with "Issue-wise set" in Theorem 14. As for the Issue-wise distance based preferences, it is shown in the following theorem.

Theorem 16. *Consider the skeptical aggregation operator and Issue-wise distance based preferences. For any agent, her strategic lies are benevolent.*

4.3 Case 2: `undec` is in the Middle between `in` and `out`

In this part, we analyze the strategy proofness for the three operators given the classes of preferences that assume `undec` is in the middle between `in` and `out` ($dist(\texttt{dec}, \texttt{undec}) < dist(\texttt{in}, \texttt{out})$).

4.3.1 IUO Hamming Sets and IUO Hamming Distance

Following, we show the strongest result for this section. The skeptical operator is strategy proof given the IUO Hamming sets based preferences. This result also holds for IUO Issue-wise sets given the discussion for Lemma 1.

Theorem 17. *The skeptical aggregation operator is strategy proof when individuals have IUO Hamming sets based preferences.*

The previous result does not hold for the credulous or the super credulous operators. Further, none of the three operators is strategy-proof when individuals have IUO Hamming distance based preferences. However, as was the case with other classes of preferences, lies with the skeptical operators are always benevolent, unlike those with the credulous or the super credulou operators.

Observation 7. *The skeptical aggregation operator is not strategy proof when individuals have IUO Hamming distance based preferences.*

Observation 8. *The credulous (resp. super credulous) aggregation operator is not strategy proof for neither IUO Hamming sets nor IUO Hamming distance based preferences.*

Proposition 9. *Consider the skeptical aggregation operator and IUO Hamming distance based preferences. For any agent, her strategic lies are benevolent.*

4.3.2 IUO Issue-wise Sets and IUO Issue-wise Distance

Similar to the results in the previous part, the skeptical operator is strategy-proof given IUO Issue-wise sets based preferences (by substituting "IUO Hamming sets" with "IUO Issue-wise sets" in Theorem 17, given the discussion for Lemma 1), unlike the credulous or the super credulous operators for which lies are possible and might not be benevolent. Further, none of the three operators is strategy-proof when individuals have IUO Hamming distance based preferences, but lies with the skeptical operators are always benevolent, unlike those with the credulous or the super credulou operators.

Observation 9. *The skeptical aggregation operator is not strategy proof when individuals have IUO Issue-wise distance based preferences.*

Observation 10. *The credulous and super credulous aggregation operators are not strategy proof when individuals have IUO Issue-wise sets (resp. distance) based preferences.*

Proposition 10. *Consider the skeptical aggregation operator and IUO Issue-wise distance based preferences. For any agent, her strategic lies are benevolent.*

Table 4 summarizes the strategy proofness results for the three operators given all the eight classes of preferences.

	Skeptical Operator	**Credulous Operator**	**Super Credulous Operator**
Hamming set **Issue-wise set**	No (Obs. 3 and 5) but benev. (Thm. 14)	No, and not benev. (Obs. 4 and 6)	No, and not benev. (Obs. 4 and 6)
Hamming dist.	No (Obs. 3) but benev. (Thm. 15)	No, and not benev. (Obs. 4)	No, and not benev. (Obs. 4)
Issue-wise dist.	No (Obs. 5) but benev. (Thm. 16)	No, and not benev. (Obs. 6)	No, and not benev. (Obs. 6)
IUO Hamming sets **IUO Issue-wise sets**	Yes (Thm. 17)	No, and not benev. (Obs. 8 and 10)	No, and not benev. (Obs. 8 and 10)
IUO Hamming dist.	No (Obs. 7) but benev. (Prop. 9)	No, and not benev. (Obs. 8)	No, and not benev. (Obs. 8)
IUO Issue-wise dist.	No (Obs. 9) but benev. (Prop. 10)	No, and not benev. (Obs. 10)	No, and not benev. (Obs. 10)

Table 4: Strategy proofness of operators depending on the type of preferences.

4.4 Heterogeneous Preferences

Following Subsection 3.4, we do a similar analysis for the case where agents have heterogeneous preferences. Since strategy proofness is usually considered given other agents's preferences are fixed, it is easy to show the result for the heterogeneous preferences given the homogeneous preferences.

Theorem 18. *Let $\mathcal{F}$ be the set of all possible classes of preferences, $\mathcal{R}$ be some set s.t. $\mathcal{R} \subseteq \mathcal{F}$, and Ag be the set of agents. If an operator is strategy proof given that Ag have homogeneous preferences from $\mathcal{R}$, then it is strategy proof given that Ag have heterogeneous preferences from $\mathcal{R}$.*

Corollary 4. *Let $\mathcal{R} = \{\ominus^{\mathcal{M}}, \ominus^{\mathcal{M}}_{\mathcal{W}}\}$. The skeptical aggregation operator is strategy proof given that agents have heterogeneous preferences from $\mathcal{R}$.*

5 Discussion and Future Work

In order to apply argumentation to multi-agent conflict resolution, it is crucial to take into account not only postulates about logical consistency, but also measures of social optimality and strategic manipulation. Two key criteria are Pareto optimality and strategy proofness, which are fundamental in any social choice and multi-agent setting. In this study, we have analyzed and compared three aggregation operators, namely the skeptical, the credulous and the super credulous operators with respect to a wide range of classes of preferences. Our comparison is based on two fundamental criteria, namely Pareto optimality and strategy proofness. Eight different classes of preferences were considered by using eight different distance methods. Additionally, we established relations between the different classes of preferences. Some of these relations hold for any aggregation operator and others for some special aggregation operators. Moreover, we also consider cases where agents do not share the same classes of preference.

We showed that the skeptical operator guarantees Pareto optimal outcomes given all the different classes of preferences, while the credulous and super credulous operators only guarantee Pareto optimal outcomes given the set-based preferences. Since more committed outcomes might be more desirable in general, credulous and super credulous operators will be preferable if individuals' preferences are known to be set-based. However, if the individuals preferences are unknown or are known to be distance-based, then there is a trade-off between Pareto optimality and the more committed outcomes. As for the strategy proofness, the three operators are vulnerable to manipulation given most classes of preferences. However, the skeptical operator guarantees benevolent lies. Understandably, unlike malicious lies, benevolent ones are not harmful to the group. Hence, there is another trade-off in choosing an appropriate operator between avoiding the malicious lies and choosing the more committed outcomes.

Few studies have considered Pareto optimality and strategy proofness with argument based aggregation. Rahwan and Larson [39] defined a set of simplistic agents preferences over argumentation outcomes, and studied the Pareto optimality of different argument evaluation rules defined using classical semantics (e.g. *complete*,...etc.) given agents with these simple types of preferences. Unlike Rahwan and Larson, we study the Pareto optimality of labeling aggregation operators that produce a collective evaluation given many different evaluations. Another difference is that we consider more realistic, distance-based preferences. As for strategy proofness, since the Gibbard-Satterthwaite theorem [27, 44], much research has been done towards analyzing strategic manipulation of preference aggregation (PA) rules [28, 29, 35, 19, 36, 25]. Strategy proofness of judgment aggregation (JA) operators have been first studied by Dietrich and List [20, 21]. In the former, Dietrich mentioned some *independence* conditions that make the rule strategy proof. In the latter, Dietrich and List showed equivalence between satisfying strategy proofness and satisfying both the *independence* and *monotonicity* postulates.

The first study of strategy proofness of labeling aggregation operator has been done by Rahwan and Tohmé [42] in the context of a specific labeling aggregation operator (argument-wise plurality rule). They showed the strategy proofness of this operator given agents with a particular class of preferences, dubbed focal set preferences. Our work considers different labeling aggregation operators, and we provide the first broad analysis for strategy proofness of labeling aggregation operators given a wide variety of preferences. Strategic manipulation in argumentation has also been studied by Rahwan, Larson and others [40, 38, 37], when arguments are distributed among agents, and where these agents may choose to show or hide arguments. Thus, that work focuses on how agents contribute to the construction of the argument graph itself, which is then evaluated centrally by the mechanism (e.g. a judge). In contrast, our present paper is concerned with strategic manipulation of labeling aggregation operators which take as input different evaluations of a given fixed graph. In fact, the difference between our problem and theirs can be analogized to jury versus litigators. The former are provided with shared information, while the latter propose their information on a dialog basis.

We believe that one of the strengths of the argumentation approach is that it is able to provide a dialectical specification of non-monotonic inference. That is, whether or not an argument is accepted (w.r.t. a specific argumentation semantics) can be assessed using dialectical proof procedures. For instance whether an argument is labelled `in` by the grounded labelling can be assessed using either the *Standard Grounded Game* [34] or the more recently defined *Grounded Persuasion Game* [17]. Another example is the *Admissibility Game* [13] that assesses whether an argument A is in an admissible set (equiv. in a preferred extension or a complete extension). The seminal work by Dung with model-based semantics such as the grounded and preferred semantics laid the building blocks for these games. Likewise, we believe our work and the work of Caminada and Pigozzi [15] lays the building blocks for dialectical preference aggregation by studying the aggregation of opinions using model-based semantics. For instance, by using modified versions of the above mentioned games, one can define a game for the down-admissible operator and a game for the up-complete operator. Then, using these two games, one can provide dialectical proof procedures for each of the JA-operators (skeptical, credulous and super-credulous) studied in this work. The interested reader can refer to the technical report [12] for more details about how the previously mentioned games can be modified to define games for the studied aggregation operators.

Acknowledgements. *Part of this paper was written during a visit of Edmond Awad to the University of Luxembourg which was generously supported by SINTELNET. The research of Martin Caminada was supported by the Engineering and Physical Sciences Research Council (EPSRC, UK), grant ref. EP/J012084/1 (SAsSY project). The research of Gabriella Pigozzi benefited from the support of the project AMANDE ANR-13-BS02-0004 of the French National Research Agency (ANR). The research of Mikołaj Podlaszewski was supported by the National Research Fund, Luxembourg (LAAMIcomp project).*

References

[1] Kenneth J. Arrow. *Social choice and individual values*. Wiley, New York NY, USA, 1951.

[2] Kenneth J. Arrow, A. K. Sen, and K. Suzumura, editors. *Handbook of Social Choice and Welfare*, volume 1. Elsevier Science Publishers (North-Holland), 2002.

[3] Edmond Awad, Richard Booth, Fernando Tohme, and Iyad Rahwan. Judgment aggregation in multi-agent argumentation. *Journal of Logic and Computation*, 2015.

[4] Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10–15):619–641, 2007.

[5] Philippe Besnard and Sylvie Doutre. Checking the acceptability of a set of arguments. In *NMR*, volume 4, pages 59–64, 2004.

[6] Philippe Besnard, Sylvie Doutre, and Andreas Herzig. Encoding argument graphs in logic. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 345–354. Springer, 2014.

[7] Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, Cambridge MA, USA, 2008.

[8] Jean-François Bonnefon. Behavioral evidence for framing effects in the resolution of the doctrinal paradox. *Social choice and welfare*, 34(4):631–641, 2010.

[9] Richard Booth, Edmond Awad, and Iyad Rahwan. Interval methods for judgment aggregation in argumentation. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR 2014)*, pages 594–597, 2014.

[10] Richard Booth, Martin Caminada, Mikołaj Podlaszewski, and Iyad Rahwan. Quantifying disagreement in argument-based reasoning. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 493–500. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[11] Richard Booth and Mikołaj Podlaszewski. Using distances for aggregation in abstract argumentation. In *Proceedings of the 26th Benelux Conference on Artificial Intelligence (BNAIC2014)*, pages 17–24, 2014.

[12] Martin Caminada and Richard Booth. Towards dialogue games for the down-admissible and up-complete procedures. Technical Report http://hdl.handle.net/10993/16504, University of Luxembourg, April 2014.

[13] Martin Caminada, Wolfgang Dvorák, and Srdjan Vesic. Preferred semantics as socratic discussion. *Journal of Logic and Computation*, 2014.

[14] Martin Caminada and Dov M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.

[15] Martin Caminada and Gabriella Pigozzi. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102, 2011.

[16] Martin Caminada, Gabriella Pigozzi, and Mikołaj Podlaszewski. Manipulation in group argument evaluation. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume One*, pages 121–126. AAAI Press, 2011.

[17] Martin Caminada and Mikolaj Podlaszewski. Grounded semantics as persuasion dialogue. *COMMA*, 245:478–485, 2012.

[18] Martin W. A. Caminada. On the issue of reinstatement in argumentation. In Michael Fisher, Wiebe van der Hoek, Boris Konev, and Alexei Lisitsa, editors, *Proceedings of the 10th European Conference on Logics in Artificial Intelligence (JELIA)*, volume 4160 of *Lecture Notes in Computer Science*, pages 111–123. Springer, 2006.

[19] John R Chamberlin. A mathematical programming approach to assessing the manipulability of social choice functions. *Political Methodology*, 8(4):25–38, 1982.

[20] Franz Dietrich. Judgment aggregation:(im) possibility theorems. *Journal of Economic Theory*, 126(1):286–298, 2006.

[21] Franz Dietrich and Christian List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23:269–300, 2007.

[22] Elad Dokow and Ron Holzman. Aggregation of binary evaluations with abstentions. *Journal of Economic Theory*, 145(2):544–561, 2010.

[23] Elad Dokow and Ron Holzman. Aggregation of non-binary evaluations. *Advances in Applied Mathematics*, 45(4):487–504, 2010.

[24] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[25] Pierre Favardin, Dominique Lepelley, and Jérôme Serais. Borda rule, Copeland method and strategic manipulation. *Review of Economic Design*, 7(2):213–228, 2002.

[26] Wulf Gärtner. *A Primer on Social Choice Theory*. Oxford University Press, 2006.

[27] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–601, 1973.

[28] Allan Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45:665–681, 1977.

[29] Allan Gibbard. Straightforwardness of game forms with lotteries as outcomes. *Econometrica*, 46:595–614, 1978.

[30] Davide Grossi and Gabriella Pigozzi. *Judgment Aggregation: A Primer*. Morgan & Claypool, 2014.

[31] C. List. The theory of judgment aggregation: An introductory review. *Synthese*, 187(1):179–207, 2012.

[32] Christian List and Ben Polak. Introduction to judgment aggregation. *Journal of economic theory*, 145(2):441–466, 2010.

[33] Christian List and Clemens Puppe. Judgment aggregation: a survey. In Paul Anand, Clemens Puppe, and Prasanta Pattanaik, editors, *The handbook of rational and social choice*. Oxford University Press, Oxford, UK, 2009.

[34] Sanjay Modgil and Martin Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in artificial intelligence*, pages 105–129. Springer, 2009.

[35] Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[36] Shmuel Nitzan. The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice*, 47(2):349–370, 1985.

[37] Shengying Pan, Kate Larson, and Iyad Rahwan. Argumentation mechanism design for preferred semantics. *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA)*, pages 403–414, 2010.

[38] Iyad Rahwan and Kate Larson. Mechanism design for abstract argumentation. In L. Padgham, D. Parkes, J. Mueller, and S. Parsons, editors, *7th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS'2008, Estoril, Portugal*, pages 1031–1038, 2008.

[39] Iyad Rahwan and Kate Larson. Pareto optimality in abstract argumentation. In Dieter Fox and Carla Gomes, editors, *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI-2008)*, pages 150–155, Menlo Park CA, USA, 2008.

[40] Iyad Rahwan, Kate Larson, and Fernando Tohmé. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 251–256, 2009.

[41] Iyad Rahwan and Guillermo R. Simari, editors. *Argumentation in Artificial Intelligence*. Springer, 2009.

[42] Iyad Rahwan and Fernando Tohmé. Collective Argument Evaluation as Judgement Aggregation. In *9th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS'2010, Toronto, Canada*, 2010.

[43] David Rönnegard. *The Fallacy of Corporate Moral Agency*. Springer, 2015.

[44] M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.

[45] Philippe Vincke. Aggregation of preferences: a review. *European Journal of Operational Research*, 9(1):17–22, 1982.

Appendix

This part contains proofs and counterexamples for the results presented in the paper "Pareto Optimality and Strategy Proofness in Group Argument Evaluation"

1 General Lemmas

The following two lemmas are crucial for establishing the relations between the different classes of preferences. The first lemma implies a very interesting result. While Hamming distance and Issue-wise distance based preferences are different, as we showed earlier, Hamming set and Issue-wise set based preferences are equivalent. The same can also be said about IUO Hamming sets and IUO Issue-wise sets. Hence, all the results in this paper that hold (resp. do not hold) for Hamming set based preferences would also hold (resp. do not hold) for Issue-wise set based preferences. The same can be also said about IUO Hamming sets and IUO issue-wise set.

Lemma 1. *Issue-wise set based preferences coincide with Hamming set based preferences, and IUO Issue-wise sets based preferences coincide with IUO Hamming sets based preferences. Formally, let $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework, and let $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ be three labelings. Then:*

1. $\mathcal{L}_1 \ominus \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus \mathcal{L}_3 \Leftrightarrow \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_3$.
   (or equivalently $\mathcal{L}_2 \succeq_{1,\ominus} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,\ominus_{\mathcal{W}}} \mathcal{L}_3$)
2. $\mathcal{L}_1 \ominus^{\mathcal{M}} \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus^{\mathcal{M}} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_1 \ominus^{\mathcal{M}}_{\mathcal{W}} \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus^{\mathcal{M}}_{\mathcal{W}} \mathcal{L}_3$.
   (or equivalently $\mathcal{L}_2 \succeq_{1,\ominus^{\mathcal{M}}} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,\ominus^{\mathcal{M}}_{\mathcal{W}}} \mathcal{L}_3$

Proof. Let $\mathcal{I}$ be the set of all issues in $\mathcal{AF}$.

1. ($\Rightarrow$): From the definition of issues, we have that $\forall \mathcal{B} \in \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2$ (where $\mathcal{B} \in \mathcal{I}$):

   $$\forall A \in \mathcal{A} : A \in \mathcal{B} \Rightarrow A \in \mathcal{L}_1 \ominus \mathcal{L}_2$$

   Then, by assumption, we have $A \in \mathcal{L}_1 \ominus \mathcal{L}_3$. Hence, we have $\forall A \in \mathcal{B} : A \in \mathcal{L}_1 \ominus \mathcal{L}_3$. Then, $\mathcal{B} \in \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_3$.

   ($\Leftarrow$): Consider an arbitrary argument A s.t. $A \in \mathcal{L}_1 \ominus \mathcal{L}_2$. Let $\mathcal{B} \in \mathcal{I}$ be s.t. $A \in \mathcal{B}$. Then:

   $$\forall A' \neq A : A' \in \mathcal{B} \Rightarrow A' \in \mathcal{L}_1 \ominus \mathcal{L}_2$$

   from the definition of issues. This means that $\mathcal{B} \in \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2$. By assumption, $\mathcal{B} \in \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_3$. Then, $A \in \mathcal{L}_1 \ominus \mathcal{L}_3$.

2. Similar to (1), but instead of showing from $\ominus$ to $\ominus_{\mathcal{W}}$ and vice versa, it is enough to show from $\ominus^{io}$ and $\ominus^{du}$ to $\ominus^{io}_{\mathcal{W}}$ and $\ominus^{du}_{\mathcal{W}}$ and vice versa, respectively.

Using the definitions of set-based preferences yield $\mathcal{L}_2 \succeq_{1,\ominus} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,\ominus_{\mathcal{W}}} \mathcal{L}_3$ and $\mathcal{L}_2 \succeq_{1,\ominus^{\mathcal{M}}} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,\ominus^{\mathcal{M}}_{\mathcal{W}}} \mathcal{L}_3$.

■

The following lemma is important in the context of compatible operators. For each agent $i \in Ag$, let $\mathcal{L}_i = \mathcal{L}_1$. Then, provided the conditions below, the lemma says an individual's preference over $\mathcal{L}_2$ and $\mathcal{L}_3$ would coincide whether she has a Hamming set (resp. distance) or IUO Hamming sets (resp. distance). The same can be said about Issue-wise set (resp. distance) and IUO Issue-wise sets (resp. distance).

Lemma 2. *Let $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework. Let $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ be three labelings and let $\mathcal{L}_1 \approx \mathcal{L}_2$ and $\mathcal{L}_1 \approx \mathcal{L}_3$:*

1. $\mathcal{L}_1 \ominus \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus \mathcal{L}_3 \Leftrightarrow \mathcal{L}_1 \ominus^{\mathcal{M}} \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus^{\mathcal{M}} \mathcal{L}_3$, *and*
   $\mathcal{L}_1 \left|\ominus\right| \mathcal{L}_2 \leq \mathcal{L}_1 \left|\ominus\right| \mathcal{L}_3 \Leftrightarrow \mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_2 \leq \mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_3$.
   (or equivalently $\mathcal{L}_2 \succeq_{1,\ominus} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,\ominus^{\mathcal{M}}} \mathcal{L}_3$ and $\mathcal{L}_2 \succeq_{1,|\ominus|} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,|\ominus^{\mathcal{M}}|} \mathcal{L}_3$)
2. $\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_1 \ominus^{\mathcal{M}}_{\mathcal{W}} \mathcal{L}_2 \subseteq \mathcal{L}_1 \ominus^{\mathcal{M}}_{\mathcal{W}} \mathcal{L}_3$, *and*
   $\mathcal{L}_1 \left|\ominus_{\mathcal{W}}\right| \mathcal{L}_2 \leq \mathcal{L}_1 \left|\ominus_{\mathcal{W}}\right| \mathcal{L}_3 \Leftrightarrow \mathcal{L}_1 \left|\ominus^{\mathcal{M}}_{\mathcal{W}}\right| \mathcal{L}_2 \leq \mathcal{L}_1 \left|\ominus^{\mathcal{M}}_{\mathcal{W}}\right| \mathcal{L}_3$.
   (or equivalently $\mathcal{L}_2 \succeq_{1,\ominus_{\mathcal{W}}} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,\ominus^{\mathcal{M}}_{\mathcal{W}}} \mathcal{L}_3$ and $\mathcal{L}_2 \succeq_{1,|\ominus_{\mathcal{W}}|} \mathcal{L}_3 \Leftrightarrow \mathcal{L}_2 \succeq_{1,|\ominus^{\mathcal{M}}_{\mathcal{W}}|} \mathcal{L}_3$)

Proof. 1. Since $\mathcal{L}_1 \approx \mathcal{L}_2$ then:

$$\neg\exists A \in \mathcal{A} \text{ s.t. } (\mathcal{L}_1(A) = \texttt{in} \wedge \mathcal{L}_2(A) = \texttt{out}) \vee (\mathcal{L}_1(A) = \texttt{out} \wedge \mathcal{L}_2(A) = \texttt{in}) \tag{17}$$

Therefore $\mathcal{L}_1 \ominus^{io} \mathcal{L}_2 = \emptyset$ which implies:

$$\mathcal{L}_1 \ominus^{\mathcal{M}} \mathcal{L}_2 = \mathcal{L}_1 \ominus^{du} \mathcal{L}_2 = \mathcal{L}_1 \ominus \mathcal{L}_2 \tag{18}$$

and

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_2 = \left|\mathcal{L}_1 \ominus^{du} \mathcal{L}_2\right| = \mathcal{L}_1 \left|\ominus\right| \mathcal{L}_2 \tag{19}$$

Similarly, we can show that:

$$\mathcal{L}_1 \ominus^{\mathcal{M}} \mathcal{L}_3 = \mathcal{L}_1 \ominus^{du} \mathcal{L}_3 = \mathcal{L}_1 \ominus \mathcal{L}_3 \tag{20}$$

and

$$\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_3 = \left|\mathcal{L}_1 \ominus^{du} \mathcal{L}_3\right| = \mathcal{L}_1 \left|\ominus\right| \mathcal{L}_3 \tag{21}$$

Now, using Eq.18 and Eq.20 together, and using Eq.19 and Eq.21 together, we can prove both directions for the results in 1. Then, the equilvance follows from the definitions of set-based and distance-based preferences.

2. Let $\mathcal{I}$ be the set of all issues in $\mathcal{AF}$. Since $\mathcal{L}_1 \approx \mathcal{L}_2$ then:

$$\neg\exists \mathcal{B} \in \mathcal{I} \text{ s.t. } (\mathcal{L}_1(A) = \texttt{in} \wedge \mathcal{L}_2(A) = \texttt{out}) \vee (\mathcal{L}_1(A) = \texttt{out} \wedge \mathcal{L}_2(A) = \texttt{in}) \\ \text{for some (equiv. all) } A \in \mathcal{B} \tag{22}$$

The rest is similar.

■

The following lemmas are also crucial for the proofs of theorems in this paper. Since the labelings have only three values, we can use the following lemma.

Lemma 3. *Let* $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ *be an argumentation framework. Let* $\texttt{dec}(\mathcal{L}) = \texttt{in}(\mathcal{L}) \cup \texttt{out}(\mathcal{L})$ $\forall \mathcal{L} \in \texttt{Labs}$. *For any pair* $\mathcal{L}_1, \mathcal{L}_2 \in \texttt{Labs}$*:*

a) $\mathcal{L}_1 \ominus \mathcal{L}_2 = (\texttt{in}(\mathcal{L}_1) \cap \texttt{out}(\mathcal{L}_2)) \cup (\texttt{in}(\mathcal{L}_1) \cap \texttt{undec}(\mathcal{L}_2)) \cup (\texttt{out}(\mathcal{L}_1) \cap \texttt{in}(\mathcal{L}_2)) \cup (\texttt{out}(\mathcal{L}_1) \cap \texttt{undec}(\mathcal{L}_2)) \cup (\texttt{undec}(\mathcal{L}_1) \cap \texttt{in}(\mathcal{L}_2)) \cup (\texttt{undec}(\mathcal{L}_1) \cap \texttt{out}(\mathcal{L}_2))$

b) if $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$ *then* $\mathcal{L}_1 \ominus \mathcal{L}_2 = \texttt{undec}(\mathcal{L}_1) \cap \texttt{dec}(\mathcal{L}_2)$

c) if $\mathcal{L}_1 \approx \mathcal{L}_2$ *then* $\mathcal{L}_1 \ominus \mathcal{L}_2 = (\texttt{dec}(\mathcal{L}_1) \cap \texttt{undec}(\mathcal{L}_2)) \cup (\texttt{undec}(\mathcal{L}_1) \cap \texttt{dec}(\mathcal{L}_2))$

Proof.

a) This follows from the fact that $\texttt{in}(\mathcal{L})$, $\texttt{out}(\mathcal{L})$ and $\texttt{undec}(\mathcal{L})$ partition the domain of any labeling $\mathcal{L}$.

b) From $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$, the sets $(\texttt{in}(\mathcal{L}_1) \cap \texttt{out}(\mathcal{L}_2))$, $(\texttt{in}(\mathcal{L}_1) \cap \texttt{undec}(\mathcal{L}_2))$, $(\texttt{out}(\mathcal{L}_1) \cap \texttt{in}(\mathcal{L}_2))$, and $(\texttt{out}(\mathcal{L}_1) \cap \texttt{undec}(\mathcal{L}_2))$ are all empty sets. Then, we are left with the following:

$$(\texttt{undec}(\mathcal{L}_1) \cap \texttt{in}(\mathcal{L}_2)) \cup (\texttt{undec}(\mathcal{L}_1) \cap \texttt{out}(\mathcal{L}_2))$$

which can be written as:

$$\texttt{undec}(\mathcal{L}_1) \cap (\texttt{in}(\mathcal{L}_2) \cup \texttt{out}(\mathcal{L}_2))$$

and replacing $\texttt{in}(\mathcal{L}) \cup \texttt{out}(\mathcal{L})$ by $\texttt{dec}(\mathcal{L})$ would give the result.

c) From $\mathcal{L}_1 \approx \mathcal{L}_2$, the sets $(\texttt{in}(\mathcal{L}_1) \cap \texttt{out}(\mathcal{L}_2))$, and $(\texttt{out}(\mathcal{L}_1) \cap \texttt{in}(\mathcal{L}_2))$ are empty. The rest can be rearranged similarly to b), and replacing $\texttt{in}(\mathcal{L}) \cup \texttt{out}(\mathcal{L})$ by $\texttt{dec}(\mathcal{L})$ would give the result.

■

We now prove two lemmas establishing the relations between less or equally committed labelings and Hamming based preferences over labelings.

Lemma 4. *Let $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework. Let $\mathcal{L}$, $\mathcal{L}'$ and $\mathcal{L}_i$ be three labelings such that $\mathcal{L} \sqsubseteq \mathcal{L}' \sqsubseteq \mathcal{L}_i$. If $\mathcal{L}_i$ is the most preferred labeling of agent i and her preference is Hamming set or Hamming distance based, then $\mathcal{L}' \succeq_{i,\ominus} \mathcal{L}$ and $\mathcal{L}' \succeq_{i,|\ominus|} \mathcal{L}$ respectively.*

Proof. From $\mathcal{L} \sqsubseteq \mathcal{L}'$, we have that $\texttt{dec}(\mathcal{L}) \subseteq \texttt{dec}(\mathcal{L}')$, which is equivalent to $\texttt{undec}(\mathcal{L}') \subseteq \texttt{undec}(\mathcal{L})$ because $\texttt{undec}$ is the complement of $\texttt{dec}$. From this, it follows that $\texttt{undec}(\mathcal{L}') \cap \texttt{dec}(\mathcal{L}_i) \subseteq \texttt{undec}(\mathcal{L}) \cap \texttt{dec}(\mathcal{L}_i)$. Since $\mathcal{L} \sqsubseteq \mathcal{L}_i$ and $\mathcal{L}' \sqsubseteq \mathcal{L}_i$ (by assumption and transitivity of $\sqsubseteq$), we can use Lemma 3b to obtain $\mathcal{L}' \ominus \mathcal{L}_i \subseteq \mathcal{L} \ominus \mathcal{L}_i$. By definition we have that $\mathcal{L}' \succeq_{i,\ominus} \mathcal{L}$ and $\mathcal{L}' \succeq_{i,|\ominus|} \mathcal{L}$. ■

Lemma 5. *Let $\mathcal{AF} = \langle \mathcal{A}, \rightharpoonup \rangle$ be an argumentation framework. Let $\mathcal{L}$, $\mathcal{L}'$ and $\mathcal{L}_i$ be three labelings and let $\mathcal{L} \sqsubseteq \mathcal{L}_i$. If $\mathcal{L}_i$ is the most preferred labeling of agent i, her preference is Hamming set based and $\mathcal{L}' \succeq_{i,\ominus} \mathcal{L}$, then $\mathcal{L} \sqsubseteq \mathcal{L}'$.*

Proof. $\mathcal{L}' \succeq_{i,\ominus} \mathcal{L}$ implies $\mathcal{L}' \ominus \mathcal{L}_i \subseteq \mathcal{L} \ominus \mathcal{L}_i$ which implies $\mathcal{L}(A) = \mathcal{L}_i(A) \Rightarrow \mathcal{L}'(A) = \mathcal{L}_i(A)$ for any argument A (i). Now, $\mathcal{L} \sqsubseteq \mathcal{L}_i$ implies $\mathcal{L}(A) = \mathcal{L}_i(A)$ for any $A \in \texttt{dec}(\mathcal{L})$ (ii). From (i) and (ii) it follows that $\mathcal{L}(A) = \mathcal{L}'(A)$ for any $A \in \texttt{dec}(\mathcal{L})$. Hence $\mathcal{L} \sqsubseteq \mathcal{L}'$. ■

2 Pareto Optimality

Theorem 2. *Let $\otimes \in \{\ominus, \ominus^{\mathcal{M}}, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}_{\mathcal{W}}\}$ be a set measure and $|\otimes|$ be its corresponding distance measure (i.e. if $\otimes = \ominus^{\mathcal{M}}$ then $|\otimes| = |\ominus^{\mathcal{M}}|$). If a labeling is Pareto optimal in a set $\mathcal{S}$ given agents with $|\otimes|$-based preferences, then it is Pareto optimal in $\mathcal{S}$ given agents with $\otimes$-based preferences.*

Proof. Let $\mathcal{S}$ be a set of labelings, and $\mathcal{L}$ be a labeling that is Pareto optimal in $\mathcal{S}$ given agents with $|\otimes|$-based preferences. Suppose, towards a contradiction, that $\mathcal{L}$ is not Pareto optimal in $\mathcal{S}$ given agents with $\otimes$-based preferences. Then, $\exists \mathcal{L}_X \in \mathcal{S}$ such that:

$$(\forall i \in Ag\colon \mathcal{L}_X \succeq_{i,\otimes} \mathcal{L}) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}_X \succ_{j,\otimes} \mathcal{L}) \tag{23}$$

From the definition of strict preferences $\succ$:

$$(\forall i \in Ag\colon \mathcal{L}_X \succeq_{i,\otimes} \mathcal{L}) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}_X \succeq_{j,\otimes} \mathcal{L} \wedge \neg \mathcal{L} \succeq_{j,\otimes} \mathcal{L}_X) \tag{24}$$

From the definition of set-based preference:

$$(\forall i \in Ag\colon \mathcal{L}_X \otimes \mathcal{L}_i \subseteq \mathcal{L} \otimes \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}_X \otimes \mathcal{L}_j \subset \mathcal{L} \otimes \mathcal{L}_j) \tag{25}$$

This implies:

$$(\forall i \in Ag\colon \mathcal{L}_X |\otimes| \mathcal{L}_i \leq \mathcal{L} |\otimes| \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}_X |\otimes| \mathcal{L}_j < \mathcal{L} |\otimes| \mathcal{L}_j) \tag{26}$$

Which means that $\mathcal{L}$ is not Pareto optimal in $\mathcal{S}$ given agents with $|\otimes|$-based preferences. Contradiction. ■

Theorem 3. *Let $\mathcal{X}$ be the set of all admissible labelings that are compatible ($\approx$) with each of the participants' individual labelings. Let $\mathcal{S}$ be any arbitrary set such that $\mathcal{S} \subseteq \mathcal{X}$. A labeling from $\mathcal{S}$ is Pareto optimal in $\mathcal{S}$ when individual preferences are Hamming set (resp. distance) based iff it is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Hamming sets (resp. distance) based.*

Proof. ($\Rightarrow$): Let $\mathcal{L}$ and $\mathcal{L}'$ be two labelings in $\mathcal{S}$. Suppose $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ when agents have Hamming set based preferences. Then, there is no labeling $\mathcal{L}_X$ that Pareto dominates $\mathcal{L}$ w.r.t Hamming set based preferences:

$$\neg\exists\mathcal{L}_X \in \mathcal{S} \text{ s.t. } (\forall i \in Ag : \mathcal{L}_X \ominus \mathcal{L}_i \subseteq \mathcal{L} \ominus \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L} \ominus \mathcal{L}_j \nsubseteq \mathcal{L}_X \ominus \mathcal{L}_j) \tag{27}$$

Note that since $\mathcal{L}, \mathcal{L}_X \in \mathcal{S} \subseteq \mathcal{X}$, then $\mathcal{L} \approx \mathcal{L}_i$, and $\mathcal{L}_X \approx \mathcal{L}_i$, $\forall i \in Ag$, then by using Lemma 2 (1) for each label $\mathcal{L}_i$, this is equivalent to:

$$\neg\exists\mathcal{L}_X \in \mathcal{S} \text{ s.t. } (\forall i \in Ag : \mathcal{L}_X \ominus^{\mathcal{M}} \mathcal{L}_i \subseteq \mathcal{L} \ominus^{\mathcal{M}} \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L} \ominus^{\mathcal{M}} \mathcal{L}_j \nsubseteq \mathcal{L}_X \ominus^{\mathcal{M}} \mathcal{L}_j) \tag{28}$$

Similarly, suppose $\mathcal{L}'$ is Pareto optimal in $\mathcal{S}$ when agents have Hamming distance based preferences. Then, there is no labeling $\mathcal{L}_X$ that Pareto dominates $\mathcal{L}$ w.r.t Hamming distance based preferences:

$$\neg\exists\mathcal{L}_X \in \mathcal{S} \text{ s.t. } (\forall i \in Ag : \mathcal{L}_X |\ominus| \mathcal{L}_i \leq \mathcal{L}' |\ominus| \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}' |\ominus| \mathcal{L}_j \nleq \mathcal{L}_X |\ominus| \mathcal{L}_j) \tag{29}$$

Also by using Lemma 2 (1) for each label $\mathcal{L}_i$, this is equivalent to:

$$\neg\exists\mathcal{L}_X \in \mathcal{S} \text{ s.t. } (\forall i \in Ag : \mathcal{L}_X \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_i \leq \mathcal{L}' \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}' \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_j \nleq \mathcal{L}_X \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_j) \tag{30}$$

($\Leftarrow$): Similar to ($\Rightarrow$) ∎

Theorem 4. *Let $\mathcal{X}$ be the set of all admissible labelings that are compatible ($\approx$) with each of the participants' individual labelings. Let $\mathcal{S}$ be any arbitrary set such that $\mathcal{S} \subseteq \mathcal{X}$. A labeling from $\mathcal{S}$ is Pareto optimal in $\mathcal{S}$ when individual preferences are Issue-wise set (resp. distance) based iff it is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Issue-wise sets (resp. distance) based.*

Proof. ($\Rightarrow$): Let $\mathcal{I}$ be the set of all issues in $\mathcal{AF}$, and let $\mathcal{L}$ and $\mathcal{L}'$ be two labelings in $\mathcal{S}$. Suppose $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ when agents have Issue-wise set based preferences. Then, there is no labeling $\mathcal{L}_X$ that Pareto dominates $\mathcal{L}$ w.r.t Issue-wise set based preferences:

$$\neg\exists\mathcal{L}_X \in \mathcal{S} \text{ s.t. } (\forall i \in Ag : \mathcal{L}_X \ominus_W \mathcal{L}_i \subseteq \mathcal{L} \ominus_W \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L} \ominus_W \mathcal{L}_j \nsubseteq \mathcal{L}_X \ominus_W \mathcal{L}_j) \tag{31}$$

The rest is similar to the proof of Theorem 3 (using "issues" instead of "arguments" with the help of Lemma 2 (2)).

∎

The following two examples are used to show the results discussed in the subsection *3.1.3 Failed Connections*.

Example 1. *Consider Figure 7, let $Ag = \{1,2\}$ be the set of agents 1 and 2, whose preferred labelings are respectively $\mathcal{L}_1$ and $\mathcal{L}_2$, and let $\mathcal{S} = \{\mathcal{L}_{CO}, \mathcal{L}_X\}$. Note the following:*

- *$\mathcal{L}_{CO}$ is Pareto optimal in $\mathcal{S}$ given agents with Hamming (Issue-wise) set based preferences.*
- *$\mathcal{L}_{CO}$ is Pareto optimal in $\mathcal{S}$ given agents with Issue-wise distance based preferences.*
- *$\mathcal{L}_{CO}$ is not Pareto optimal in $\mathcal{S}$ given agents with Hamming distance based preferences.*

Since both $\mathcal{L}_{CO}$ and $\mathcal{L}_X$ are admissible labelings that are compatible ($\approx$) with both of $\mathcal{L}_1$ and $\mathcal{L}_2$, then using Theorems 3 and 4:

- *$\mathcal{L}_{CO}$ is Pareto optimal in $\mathcal{S}$ given agents with IUO Hamming (Issue-wise) set based preferences.*
- *$\mathcal{L}_{CO}$ is Pareto optimal in $\mathcal{S}$ given agents with IUO Issue-wise distance based preferences.*
- *$\mathcal{L}_{CO}$ is not Pareto optimal in $\mathcal{S}$ given agents with IUO Hamming distance based preferences.*

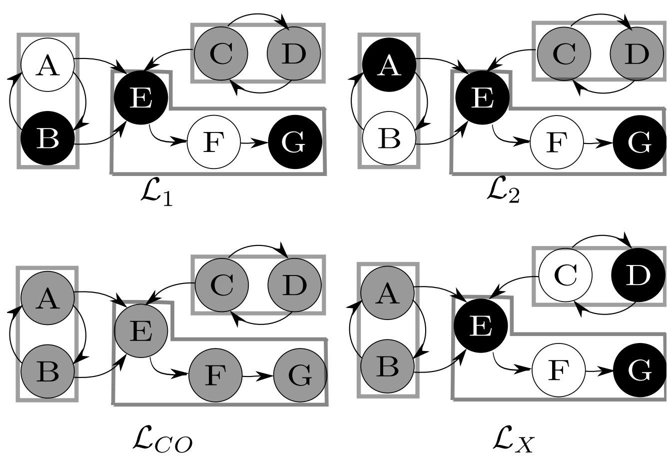

Figure 7: An example showing how Pareto optimality given agents with (IUO) Hamming distance based preferences cannot be inferred from other classes of preferences. The set of arguments located in one box form an issue.

Also consider the following example.

Example 2. *Consider Figure 8, let $Ag = \{1,2\}$ be the set of agents 1 and 2, whose preferred labelings are respectively $\mathcal{L}_1$ and $\mathcal{L}_2$, and let $\mathcal{S} = \{\mathcal{L}_{CO}, \mathcal{L}_X\}$. Note the following:*

- *$\mathcal{L}_{CO}$ is Pareto optimal in $\mathcal{S}$ given agents with Hamming (Issue-wise) set based preferences.*
- *$\mathcal{L}_{CO}$ is Pareto optimal in $\mathcal{S}$ given agents with Hamming distance based preferences.*

- $\mathcal{L}_{CO}$ *is not Pareto optimal in* $\mathcal{S}$ *given agents with Issue-wise distance based preferences.*

Also, since both $\mathcal{L}_{CO}$ *and* $\mathcal{L}_X$ *are admissible labelings that are compatible (*$\approx$*) with both of* $\mathcal{L}_1$ *and* $\mathcal{L}_2$*, then using Theorems 3 and 4:*

- $\mathcal{L}_{CO}$ *is Pareto optimal in* $\mathcal{S}$ *given agents with IUO Hamming (Issue-wise) set based preferences.*

- $\mathcal{L}_{CO}$ *is Pareto optimal in* $\mathcal{S}$ *given agents with IUO Hamming distance based preferences.*

- $\mathcal{L}_{CO}$ *is not Pareto optimal in* $\mathcal{S}$ *given agents with IUO Issue-wise distance based preferences.*

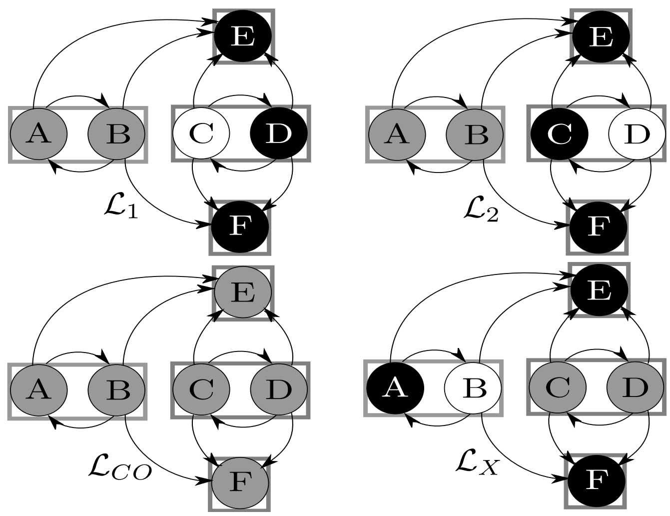

Figure 8: An example showing how Pareto optimality given agents with (IUO) Issue-wise distance based preferences cannot be inferred from other classes of preferences. The set of arguments located in one box form an issue.

Following, we show the results about Pareto optimality regarding the three operators: the skeptical, the credulous and the super credulous.

Theorem 5. *If individual preferences are Hamming set based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal (*$\sqsubseteq$*) to each of the participants' labelings.*

Proof. Let P be a profile of labelings, $\mathcal{L}_{SO} = so_{\mathcal{AF}}(P)$ and $\mathcal{L}_X$ some admissible labeling with the property $\forall i \in Ag, \mathcal{L}_X \sqsubseteq \mathcal{L}_i$. From Theorem 1 we know that $\mathcal{L}_{SO}$ is the biggest admissible labeling with such property, so $\mathcal{L}_X \sqsubseteq \mathcal{L}_{SO}$. So we have $\forall i \in Ag, \mathcal{L}_X \sqsubseteq \mathcal{L}_{SO} \sqsubseteq \mathcal{L}_i$. From Lemma 4 we have $\mathcal{L}_{SO} \succeq_i \mathcal{L}_X$ for any i. So no agent strictly prefers $\mathcal{L}_X$ and hence there is no labeling that Pareto dominates $\mathcal{L}_{SO}$. ■

Theorem 6. *If individual preferences are Hamming distance based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal (*$\sqsubseteq$*) to each of the participants' labelings.*

Proof. In the proof of Theorem 5, Hamming set may be replaced by Hamming distance because it is only used in Lemma 4, which works for Hamming distance as well. ■

Theorem 7. *If individual preferences are Hamming set based, then the credulous aggregation operator is Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Let P be a profile of labelings, $\mathcal{L}_{CO} = co_{\mathcal{AF}}(P)$, $\mathcal{L}_{CIO} = cio_{\mathcal{AF}}(P)$. Assume by contradiction that there exists some admissible labeling $\mathcal{L}_X$ with the property $\forall i \in Ag, \mathcal{L}_X \approx \mathcal{L}_i$ that Pareto dominates $\mathcal{L}_{CO}$.

First notice that compatibility ensures that there are no `in`/`out` conflicts between $\mathcal{L}_X$ and $\mathcal{L}_{CO}$. If there is an `in`/`out` conflict between agents' labelings on some argument, then both $\mathcal{L}_X$ and $\mathcal{L}_{CO}$ need to label it `undec`. If there exists an agent whose labeling decides on some argument and other agents' labelings agree or refrain from decision, $\mathcal{L}_{CO}$ and $\mathcal{L}_X$ also agree or refrain from decision. If all agents refrain from decision on some argument, $\mathcal{L}_{CO}$ by definition also refrains, and $\mathcal{L}_X$ may label freely.

Let us take $A \in \mathtt{dec}(\mathcal{L}_X)$. Then, there needs to be an agent with a labeling that agrees on A. Otherwise all agents' labelings would be undecided on such argument and, according to definition, $\mathcal{L}_{CO}$ would not decide either. But then all agents' labelings will agree on such argument with $\mathcal{L}_{CO}$ and disagree with $\mathcal{L}_X$, so no agent will strongly prefer $\mathcal{L}_X$, which contradicts with domination. So there exists at least one agent whose labeling agrees with $\mathcal{L}_X$ on A. Other agents' labelings also need to agree on A or label it `undec` because of the compatibility of $\mathcal{L}_X$. Then by definition $\mathcal{L}_{CIO}(A) = \mathcal{L}_X(A)$. This holds for any argument $A \in \mathtt{dec}(\mathcal{L}_X)$, so we have $\mathcal{L}_X \sqsubseteq \mathcal{L}_{CIO}$. But $\mathcal{L}_X$ is admissible and, by definition, $\mathcal{L}_{CO}$ is the biggest admissible labeling less or equally committed as $\mathcal{L}_{CIO}$. So we have $\mathcal{L}_X \sqsubseteq \mathcal{L}_{CO} \sqsubseteq \mathcal{L}_{CIO}$.

$\mathcal{L}_X$ must be different from $\mathcal{L}_{CO}$ to dominate it. Let A be an argument on which these labelings differ. From the previous considerations, it follows that $A \in \mathtt{undec}(\mathcal{L}_X)$ and $A \in \mathtt{dec}(\mathcal{L}_{CO})$. $\mathcal{L}_{CO}$ decides on an argument only if there exists an agent that decides on such argument. But then this agent will agree on A with $\mathcal{L}_{CO}$ and disagree with $\mathcal{L}_X$, so it will not prefer $\mathcal{L}_X$. This is in contradiction with dominance. Hence, such dominating labeling cannot exist. ■

Observation 11. *If individual preferences are Hamming distance based, then the credulous (resp. the super credulous) aggregation operator is not Pareto optimal in the set of admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings. An example is given in Figure 9 where $\mathcal{L}_{CO}$ represents the outcome of the credulous (or the super credulous) aggregation operator. Both labelings $\mathcal{L}_{CO}$ and $\mathcal{L}_X$ are compatible with both $\mathcal{L}_1$ and $\mathcal{L}_2$, but $\mathcal{L}_X$ is closer when applying Hamming distance. $\mathcal{L}_1 \ominus \mathcal{L}_{CO} = \mathcal{L}_2 \ominus \mathcal{L}_{CO} = \{A,B,E,F,G\}$, so the Hamming distance is 5, whereas $\mathcal{L}_1 \ominus \mathcal{L}_X = \mathcal{L}_2 \ominus \mathcal{L}_X = \{A,B,C,D\}$, so the Hamming distance is 4.*

Theorem 8. *If individual preferences are Hamming set based, then the super credulous aggregation operator is Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

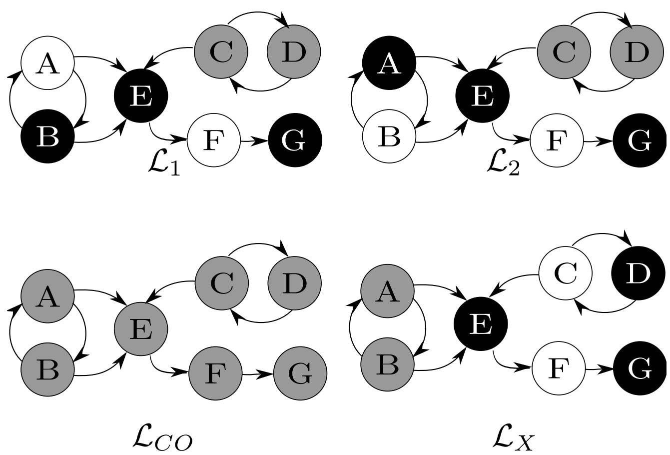

Figure 9: If individuals' preferences are Hamming distance based, the (super) credulous aggregation operator is not Pareto optimal in the set of admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings.

Proof. Let P be a profile of labelings, $\mathcal{L}_{CIO} = cio_{\mathcal{AF}}(P)$, $\mathcal{L}_{CO} = co_{\mathcal{AF}}(P)$, and $\mathcal{L}_{SCO} = sco_{\mathcal{AF}}(P)$. Suppose, towards a contradiction, that there exists a complete labeling $\mathcal{L}_X$ s.t. $\mathcal{L}_X \approx \mathcal{L}_i\ \forall i \in Ag$, and $\mathcal{L}_X$ dominates $\mathcal{L}_{SCO}$ (w.r.t $\succeq_{i,\ominus}$).

Let $A \in \mathtt{dec}(\mathcal{L}_{CO})$, then $\mathcal{L}_{SCO}$ agrees on A with $\mathcal{L}_{CO}$. However, $\mathcal{L}_{CO}$ only decides on an argument if at least one agent decides on this argument and agrees with $\mathcal{L}_{CO}$ on it. Then, this agent also agrees on A with $\mathcal{L}_{SCO}$. Since $\mathcal{L}_X$, by assumption, Pareto dominates $\mathcal{L}_{SCO}$, $\mathcal{L}_X$ also needs to agree with this agent on A. This is the case for every argument $A \in \mathtt{dec}(\mathcal{L}_{CO})$. Hence, $\forall A \in \mathtt{dec}(\mathcal{L}_{CO}) : \mathcal{L}_{CO}(A) = \mathcal{L}_X(A)$. Then, $\mathcal{L}_{CO} \sqsubseteq \mathcal{L}_X$. By definition, $\mathcal{L}_{SCO}$ is the smallest element (w.r.t $\sqsubseteq$) of the set of all complete labelings that are bigger or equally committed than $\mathcal{L}_{CO}$. Then, $\mathcal{L}_{CO} \sqsubseteq \mathcal{L}_{SCO} \sqsubseteq \mathcal{L}_X$.

$\mathcal{L}_X$ should be different from $\mathcal{L}_{SCO}$ to dominate it. Then, $\exists A \in \mathtt{undec}(\mathcal{L}_{SCO}) \cap \mathtt{dec}(\mathcal{L}_X)$. We will show that $\forall A \in \mathtt{undec}(\mathcal{L}_{SCO}) \cap \mathtt{dec}(\mathcal{L}_X)$ then $\forall i \in Ag : \mathcal{L}_i(A) = \mathtt{undec}$. This is enough to reach a contradiction because it shows that all agents agree on at least one argument with $\mathcal{L}_{SCO}$ while disagree with $\mathcal{L}_X$ on that argument.

Suppose, for contradiction, that there exists an agent j such that $\mathcal{L}_j(A) = \mathcal{L}_X(A) \in \{\mathtt{in}, \mathtt{out}\}$. Since $A \in \mathtt{undec}(\mathcal{L}_{SCO})$, then $A \in \mathtt{undec}(\mathcal{L}_{CO})$. However, $\mathcal{L}_X$ is a complete labeling which means that it is also an admissible labeling, and from Theorem 7, $\mathcal{L}_{CO}$ is Pareto optimal in the set of all admissible labelings that are compatible ($\approx$) with each of the participants' labelings. Then:

$$\forall B \in \mathcal{A}, \neg\exists i \in Ag \text{ s.t. } \mathcal{L}_{CO}(B) \neq \mathcal{L}_i(B) \wedge \mathcal{L}_X(B) = \mathcal{L}_i(B) \tag{32}$$

Contradiction. Then, all agents need to agree with $\mathcal{L}_{CO}$ and $\mathcal{L}_{SCO}$ on every A s.t. $A \in \mathtt{undec}(\mathcal{L}_{SCO}) \cap \mathtt{dec}(\mathcal{L}_X)$ (and disagree with $\mathcal{L}_X$ on A). ■

Theorem 9. *If individual preferences are Issue-wise distance based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal (w.r.t $\sqsubseteq$) to* *each of the participants' labelings.*

Proof. Let $\mathcal{I}$ be the set of all issues in $\mathcal{AF}$, P be a profile of labelings, and $\mathcal{L}_{SO} = so_{\mathcal{AF}}(P)$. Suppose, towards a contradiction, that there exists an admissible labeling $\mathcal{L}_X$ s.t. $\mathcal{L}_X \sqsubseteq \mathcal{L}_i\ \forall i \in Ag$, and $\mathcal{L}_X$ dominates $\mathcal{L}_{SO}$ (w.r.t $\succeq_{i,|\ominus_{\mathcal{W}}|}$). Then, there needs to be at least one issue on which $\mathcal{L}_X$ agrees with some labeling $\mathcal{L}_j$ (by an agent j), while $\mathcal{L}_{SO}$ disagrees with $\mathcal{L}_j$ on that issue:

$$\exists j \in Ag, \exists \mathcal{B} \in \mathcal{I} \text{ s.t. } (\mathcal{L}_X(A) = \mathcal{L}_j(A)) \wedge (\mathcal{L}_{SO}(A) \neq \mathcal{L}_j(A)) \\ \text{for some (equiv. all) } A \in \mathcal{B} \tag{33}$$

However, from Theorem 1, $\mathcal{L}_{SO}$ is the biggest labeling (w.r.t $\sqsubseteq$) in $\mathcal{X}$. Then $\mathcal{L}_X \sqsubseteq \mathcal{L}_{SO} \sqsubseteq \mathcal{L}_i$ $\forall i \in Ag$. $\mathcal{L}_X$ should be different from $\mathcal{L}_{SO}$ to dominate it. Then, $\exists A \in \mathtt{undec}(\mathcal{L}_X) \cap \mathtt{dec}(\mathcal{L}_{SO})$ (where A belongs to some issue $\mathcal{B} \in \mathcal{I}$). However, $\mathcal{L}_{SO}$ only decides on an argument if all agents decide on this argument and agree on it with $\mathcal{L}_{SO}$. Accordingly, all agents disagree with $\mathcal{L}_X$ on A. Note that this holds for all $A \in \mathtt{undec}(\mathcal{L}_X) \cap \mathtt{dec}(\mathcal{L}_{SO})$. Additionally, $\forall B \notin \mathtt{undec}(\mathcal{L}_X) \cap \mathtt{dec}(\mathcal{L}_{SO}) : \mathcal{L}_{SO}(B) = \mathcal{L}_X(B)$. Hence:

$$\neg \exists \mathcal{L}_X \in \mathcal{X} \text{ s.t. } \exists j \in Ag : (\mathcal{L}_X(A) = \mathcal{L}_j(A)) \wedge (\mathcal{L}_{SO}(A) \neq \mathcal{L}_j(A)) \text{ for any } A \in \mathcal{A} \tag{34}$$

Contradiction. ■

Observation 12. *If individual preferences are Issue-wise distance based, then the credulous (resp. the super credulous) aggregation operator is not Pareto optimal in the set of admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings. In Figure 10, $\mathcal{L}_{CO}$ represents the outcome of the credulous (or the super credulous) aggregation operator. Note that, both labelings of $\mathcal{L}_{CO}$ and $\mathcal{L}_X$ are compatible with both $\mathcal{L}_1$ and $\mathcal{L}_2$, but $\mathcal{L}_X$ is closer when applying Issue-wise distance. $\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_{CO} = \mathcal{L}_2 \ominus_{\mathcal{W}} \mathcal{L}_{CO} = \{\{C,D\},\{E\},\{F\}\}$, so Issue-wise distance is 3, whereas $\mathcal{L}_1 \ominus_{\mathcal{W}} \mathcal{L}_X = \mathcal{L}_2 \ominus_{\mathcal{W}} \mathcal{L}_X = \{\{A,B\},\{C,D\}\}$, so Issue-wise distance is 2.*

Proposition 1. *If individual preferences are IUO Hamming sets (resp. distance) based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings.*

Proof. Let $\mathcal{S}$ be the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings. Then, $\mathcal{S} \subseteq \mathcal{X}$, where $\mathcal{X}$ is defined in Theorem 3. From Theorem 5 and Theorem 3 the skeptical aggregation operator is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Hamming sets based. From Theorem 6 and Theorem 3 the skeptical aggregation operator is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Hamming distance based. ■

Proposition 2. *If individual preferences are IUO Hamming sets based, then the credulous aggregation operator is Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Let $\mathcal{S}$ be the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings. Then, $\mathcal{S} \subseteq \mathcal{X}$, where $\mathcal{X}$ is defined in Theorem 3 (actually $\mathcal{S} = \mathcal{X}$ here). From Theorem 7 and Theorem 3 the credulous aggregation operator is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Hamming sets based. ■

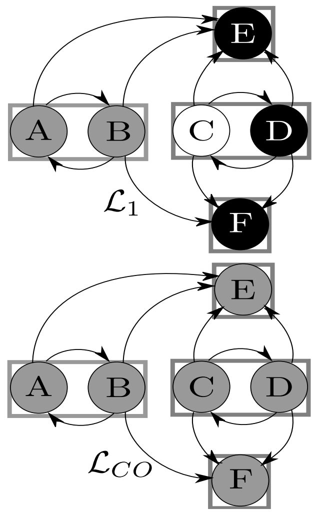

Figure 10: If individuals' preferences are Issue-wise distance based, the (super) credulous aggregation operator is not Pareto optimal in the set of admissible (resp. complete) labelings that are compatible ($\approx$) with each of the participants' labelings. The set of arguments located in one box form an issue.

Proposition 3. *If individual preferences are IUO Hamming distance based, then the credulous aggregation operator is not Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Similar to the previous proposition, from Observation 11 and Theorem 3 the credulous aggregation operator is not Pareto optimal in $\mathcal{S}$ ($\mathcal{S}$ is defined in the previous proposition) when individual preferences are IUO Hamming distance based. ∎

Proposition 4. *If individual preferences are IUO Hamming sets based, then the super credulous aggregation operator is Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Let $\mathcal{S}$ be the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings. Then, $\mathcal{S} \subseteq \mathcal{X}$, where $\mathcal{X}$ is defined in Theorem 3. From Theorem 8 and Theorem 3 the super credulous aggregation operator is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Hamming sets based. ∎

Proposition 5. *If individual preferences are IUO Hamming distance based, then the super credulous aggregation operator is not Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Similar to the previous proposition, from Observation 11 and Theorem 3 the super credulous aggregation operator is not Pareto optimal in $\mathcal{S}$ ($\mathcal{S}$ is defined in the previous proposition) when individual preferences are IUO Hamming distance based. ∎

Proposition 6. *If individual preferences are IUO Issue-wise distance based, then the skeptical aggregation operator is Pareto optimal in the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings.*

Proof. Let $\mathcal{S}$ be the set of admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings. Then, $\mathcal{S} \subseteq \mathcal{X}$, where $\mathcal{X}$ is defined in Theorem 4. From Theorem 9 and Theorem 4 the skeptical aggregation operator is Pareto optimal in $\mathcal{S}$ when individual preferences are IUO Issue-wise distance based. ■

Proposition 7. *If individual preferences are IUO Issue-wise distance based, then the credulous aggregation operator is not Pareto optimal in the set of admissible labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Similar to the previous proposition, from Observation 12 and Theorem 4 the credulous aggregation operator is not Pareto optimal in $\mathcal{S}$ ($\mathcal{S}$ is defined in the previous proposition) when individual preferences are IUO Issue-wise distance based. ■

Proposition 8. *If individual preferences are IUO Issue-wise distance based, then the super credulous aggregation operator is not Pareto optimal in the set of complete labelings that are compatible ($\approx$) with each of the participants' labelings.*

Proof. Similar to the previous proposition, from Observation 12 and Theorem 4 the super credulous aggregation operator is not Pareto optimal in $\mathcal{S}$ ($\mathcal{S}$ is defined in the previous proposition) when individual preferences are IUO Issue-wise distance based. ■

The following example shows how Pareto optimality does not generally carry over from homogeneous preferences to heterogeneous preferences.

Example 3. *Consider the framework and labelings in Figure 11. Let $Ag = \{1,2\}$, $\mathcal{S} = \{\mathcal{L}, \mathcal{L}_X\}$ and $\mathcal{R} = \{|\ominus|, |\ominus_W|\}$. When Ag have homogeneous preferences from $\mathcal{R}$ (i.e. both agents have Hamming distance based preferences or both have Issue-wise distance based preferences), then $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$:*

$$\mathcal{L}_1 \, |\ominus| \, \mathcal{L} = 4 \text{ , } \mathcal{L}_1 \, |\ominus| \, \mathcal{L}_X = 3 \textit{ (i.e. } \mathcal{L}_X \succ_{1,|\ominus|} \mathcal{L}\textit{)}$$
$$\mathcal{L}_2 \, |\ominus| \, \mathcal{L} = 3 \text{ , } \mathcal{L}_2 \, |\ominus| \, \mathcal{L}_X = 4 \textit{ (i.e. } \mathcal{L} \succ_{2,|\ominus|} \mathcal{L}_X\textit{)}$$
$$\mathcal{L}_1 \, |\ominus_W| \, \mathcal{L} = 1 \text{ , } \mathcal{L}_1 \, |\ominus_W| \, \mathcal{L}_X = 2 \textit{ (i.e. } \mathcal{L} \succ_{1,|\ominus_W|} \mathcal{L}_X\textit{)}$$
$$\mathcal{L}_2 \, |\ominus_W| \, \mathcal{L} = 2 \text{ , } \mathcal{L}_2 \, |\ominus_W| \, \mathcal{L}_X = 1 \textit{ (i.e. } \mathcal{L}_X \succ_{2,|\ominus_W|} \mathcal{L}\textit{)}$$

However, if agent 1 *has Hamming distance based preferences and agent* 2 *has Issue-wise distance based preferences then both agents would strictly prefer $\mathcal{L}_X$ over $\mathcal{L}$ and this means that $\mathcal{L}$ would not be Pareto optimal in $\mathcal{S}$ when Ag have heterogeneous preferences from $\mathcal{R}$.*

Theorem 10. *Let $\mathcal{R} = \{\ominus, \ominus^{\mathcal{M}}\}$ be a set of preference classes, Ag be a set of agents, $\mathcal{S}$ be the set of all admissible labelings that are compatible with each individual's labeling, and $\mathcal{L}$ be a labeling from $\mathcal{S}$. If $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ given that Ag have* homogeneous *preferences from $\mathcal{R}$, then $\mathcal{L}$ is Pareto optimal in $\mathcal{S}$ given that Ag have* heterogeneous *preferences from $\mathcal{R}$.*

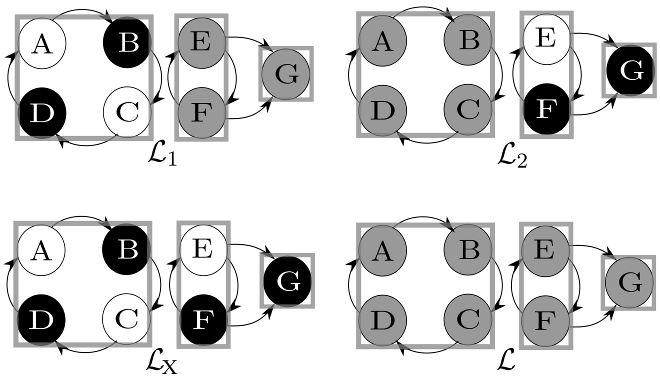

Figure 11: An example showing how a labeling that is Pareto optimal in a set $\mathcal{S}$ given that Ag have *homogeneous* preferences from a set $\mathcal{R}$, might not be Pareto optimal if Ag have *heterogeneous* preferences from $\mathcal{R}$. The set of arguments located in one box form an issue.

Proof. Let Ag be s.t. agents have heterogeneous preferences from $\mathcal{R}$. Suppose towards a contradiction, that $\mathcal{L}$ is not Pareto optimal in $\mathcal{S}$ when each agent i has $c(i) \in \mathcal{R}$ based preferences. Then, there is a labeling $\mathcal{L}_X$ that Pareto dominates $\mathcal{L}$ w.r.t $c(i)$ based preferences, i.e. $\exists \mathcal{L}_X \in \mathcal{S}$ s.t:

$$(\forall i \in Ag : \mathcal{L}_X \succeq_{i,c(i)} \mathcal{L}) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}_X \succ_{j,c(j)} \mathcal{L}) \tag{35}$$

where $c(i) \in \mathcal{R}, \forall i \in Ag$ (i.e. $c(i) = \ominus$ or $c(i) = \ominus^{\mathcal{M}}$). Then, from the definition of set based preference:

$$(\forall i \in Ag : \mathcal{L}_X\ c(i)\ \mathcal{L}_i \subseteq \mathcal{L}\ c(i)\ \mathcal{L}_i) \wedge (\exists j \in Ag \text{ s.t. } \mathcal{L}\ c(j)\ \mathcal{L}_j \nsubseteq \mathcal{L}_X\ c(j)\ \mathcal{L}_j) \tag{36}$$

However, given the compatibility of $\mathcal{L}$ and $\mathcal{L}_X$ with every individuals' labeling (since $\mathcal{L}, \mathcal{L}_X \in \mathcal{S}$) and from Lemma 2 (1), if agents who have Hamming set based preferences switched their classes of preferences to IUO Hamming sets based preferences or vice versa, then their preferences would not change. As a result, the previous equation would hold even when $c(k) = c(l), \forall k, l \in Ag$. This means $\mathcal{L}$ is not Pareto optimal in $\mathcal{S}$ when Ag have homogeneous preferences from $\mathcal{R}$. Contradiction. ∎

Theorem 11. *Let $\mathcal{R} = \{\ominus, \ominus_{\mathcal{W}}, \ominus^{\mathcal{M}}, \ominus^{\mathcal{M}}_{\mathcal{W}}, |\ominus|, |\ominus_{\mathcal{W}}|, |\ominus^{\mathcal{M}}|, |\ominus^{\mathcal{M}}_{\mathcal{W}}|\}$. The skeptical operator is Pareto optimal in the set of all admissible labelings that are smaller or equal ($\sqsubseteq$) to each of the participants' labelings given that individuals have heterogeneous preferences from $\mathcal{R}$.*

Proof. Let $\mathcal{S}$ be the set of all admissible labelings that are smaller or equal ($\sqsubseteq$) to individuals' labelings. Suppose, towards a contradiction that the skeptical operator is not Pareto optimal in $\mathcal{S}$ given that the set of individuals Ag have heterogeneous preferences from $\mathcal{R}$. Then, $\exists \mathcal{L}_X \in \mathcal{S}$ s.t. $\mathcal{L}_X$ Pareto dominates $\mathcal{L}$ (given heterogeneous preferences). Then:

$$\exists j \in Ag \text{ s.t. } \mathcal{L}_X \succ_{j,c(j)} \mathcal{L} \tag{37}$$

However, from Theorem 1, $\mathcal{L}$ is the biggest admissible labeling that is smaller or equal ($\sqsubseteq$) to each individual's labeling. Then, $\forall \mathcal{L}' \in \mathcal{S} : \mathcal{L}' \sqsubseteq \mathcal{L} \sqsubseteq \mathcal{L}_i, \forall i \in Ag$. Hence, for any agent i's labeling $\mathcal{L}_i$, and for any argument (and consequently, any issue) on which $\mathcal{L}$ disagrees with $\mathcal{L}_i$, then $\mathcal{L}_X$ would disagree in exactly the same way. Contradiction. ■

3 Strategy Proofness

Consider an operator Op that only produces labelings that are compatible ($\approx$) with each individual's labeling. The following lemma shows that every strategic lie with the operator Op given IUO Hamming distance based preferences is also a strategic lie given Hamming distance based preferences. This lemma is crucial to show that the benevolence property of lies with the skeptical operator carries over from Hamming distance based preferences to IUO Hamming distance based preferences.

Lemma 6. *Let Op be a* compatible operator. *Let $\mathcal{L}_k$ denote the top preference labeling of agent k. Let P be a profile where each agent submits her most preferred labeling, and let $P' = P_{\mathcal{L}_k/\mathcal{L}'_k}$ be a profile that results from P by changing $\mathcal{L}_k$ to $\mathcal{L}'_k$. Let $\mathcal{L}_{Op} = Op_{\mathcal{AF}}(P)$ be the outcome when agent k does not lie. Let $X^k_{|\ominus^{\mathcal{M}}|}$ (resp. $X^k_{|\ominus|}$) be the set of all labelings $\mathcal{L}'_{Op}$ that satisfy the following two properties:*

1. *There exists some labeling $\mathcal{L}'_k$ s.t. $\mathcal{L}'_{Op} = Op_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$ (i.e. $\mathcal{L}'_{Op}$ is a possible outcome given some lie by agent k), and*
2. *$\mathcal{L}'_{Op} \succ_{k,|\ominus^{\mathcal{M}}|} \mathcal{L}_{Op}$ (resp. $\mathcal{L}'_{Op} \succ_{k,|\ominus|} \mathcal{L}_{Op}$).*

Then $X^k_{|\ominus^{\mathcal{M}}|} \subseteq X^k_{|\ominus|}$.

Proof. $\forall \mathcal{L}'_{Op} \in X^k_{|\ominus^{\mathcal{M}}|}$, we have:

1. There exists some labeling $\mathcal{L}'_k$ s.t. $\mathcal{L}'_{Op} = so_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$, and
2. $\mathcal{L}'_{Op} \succ_{k,|\ominus^{\mathcal{M}}|} \mathcal{L}_{Op}$.

We just need to show that $\mathcal{L}'_{Op} \succ_{k,|\ominus|} \mathcal{L}_{Op}$.

Since $\mathcal{L}'_{Op} \succ_{k,|\ominus^{\mathcal{M}}|} \mathcal{L}_{Op}$, then $\mathcal{L}'_{Op} \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_k < \mathcal{L}_{Op} \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_k$. Then:

$$2 \times |\mathcal{L}'_{Op} \ominus^{io} \mathcal{L}_k| + |\mathcal{L}'_{Op} \ominus^{du} \mathcal{L}_k| < 2 \times |\mathcal{L}_{Op} \ominus^{io} \mathcal{L}_k| + |\mathcal{L}_{Op} \ominus^{du} \mathcal{L}_k| \tag{38}$$

Since $|\mathcal{L}_{Op} \ominus^{io} \mathcal{L}_k| = 0$:

$$2 \times |\mathcal{L}'_{Op} \ominus^{io} \mathcal{L}_k| + |\mathcal{L}'_{Op} \ominus^{du} \mathcal{L}_k| < |\mathcal{L}_{Op} \ominus^{du} \mathcal{L}_k| \tag{39}$$

Which implies:

$$|\mathcal{L}'_{Op} \ominus^{io} \mathcal{L}_k| + |\mathcal{L}'_{Op} \ominus^{du} \mathcal{L}_k| < |\mathcal{L}_{Op} \ominus^{du} \mathcal{L}_k| \tag{40}$$

But $|\mathcal{L}'_{Op} \ominus \mathcal{L}_k| = |\mathcal{L}'_{Op} \ominus^{io} \mathcal{L}_k| + |\mathcal{L}'_{Op} \ominus^{du} \mathcal{L}_k|$ and $|\mathcal{L}_{Op} \ominus \mathcal{L}_k| = |\mathcal{L}_{Op} \ominus^{du} \mathcal{L}_k|$. Then:

$$|\mathcal{L}'_{Op} \ominus \mathcal{L}_k| < |\mathcal{L}_{Op} \ominus \mathcal{L}_k| \tag{41}$$

Which means $\mathcal{L}'_{Op} \succ_{|\ominus|} \mathcal{L}_{Op}$. Hence, $\mathcal{L}'_{Op} \in X^k_{|\ominus|}$. ■

We now show that the benevolence property of lies with an operator carries over from Hamming distance based preferences to IUO Hamming distance based preferences.

Theorem 12. *Consider an operator Op that only produces labelings that are compatible ($\approx$) with each individual's labeling. If all strategic lies are benevolent when agents have Hamming distance based preferences then all strategic lies are benevolent when agents have IUO Hamming distance based preferences.*

Proof. Let Op be a *compatible operator*. Let P be a profile, and $\mathcal{L}'_k$ a strategic lie of agent k. Denote $\mathcal{L}_{Op} = Op_{\mathcal{AF}}(P)$ and $\mathcal{L}'_{Op} = Op_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$. From Lemma 2 (1), since the operator Op only produces labelings that are compatible with all individuals' labelings, then for every agent j s.t. $j \neq k$: ($\mathcal{L}_{Op} \succeq_{j,|\ominus|} \mathcal{L}'_{Op}$ iff $\mathcal{L}_{Op} \succeq_{j,|\ominus^{\mathcal{M}}|} \mathcal{L}'_{Op}$) i.e. Hamming distance based preferences and IUO Hamming distance based preferences are equivalent for all agents other than agent k.

Now given Lemma 6, every strategic lie with the operator Op given IUO Hamming distance based preferences is also a strategic lie given Hamming distance based preferences. However, all those lies are benevolent for every agent $j \neq k$ whether she has Hamming distance based preferences or IUO Hamming distance based preferences. Hence, every lie given IUO Hamming distance based preferences is benevolent. ■

Consider an operator Op that only produces labelings that are compatible ($\approx$) with each individual's labeling. The following lemma shows that every strategic lie with the operator Op given IUO Issue-wise distance based preferences is also a strategic lie given Issue-wise distance based preferences. This lemma is crucial to show that the benevolence property of lies with the operator Op carries over from Issue-wise distance based preferences to IUO Issue-wise distance based preferences.

Lemma 7. *Let Op be a* compatible operator. *Let $\mathcal{L}_k$ denote the top preference labeling of agent k. Let P be a profile where each agent submits her most preferred labeling, and let $P' = P_{\mathcal{L}_k/\mathcal{L}'_k}$ be a profile that results from P by changing $\mathcal{L}_k$ to $\mathcal{L}'_k$. Let $\mathcal{L}_{Op} = Op_{\mathcal{AF}}(P)$ be the outcome when agent k does not lie. Let $X^k_{|\ominus^{\mathcal{M}}_{\mathcal{W}}|}$ (resp. $X^k_{|\ominus_{\mathcal{W}}|}$) be the set of all labelings $\mathcal{L}'_{Op}$ that satisfy the following two properties:*

1. *There exists some labeling $\mathcal{L}'_k$, $\mathcal{L}'_{Op} = Op_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$, and*
2. *$\mathcal{L}'_{Op} \succ_{k,|\ominus^{\mathcal{M}}_{\mathcal{W}}|} \mathcal{L}_{Op}$ (resp. $\mathcal{L}'_{Op} \succ_{k,|\ominus_{\mathcal{W}}|} \mathcal{L}_{Op}$).*

Then $X^k_{|\ominus^{\mathcal{M}}_{\mathcal{W}}|} \subseteq X^k_{|\ominus_{\mathcal{W}}|}$.

Proof. This proof is similar to the one in Lemma 6. ■

Theorem 13. *Consider an operator Op that only produces labelings that are compatible ($\approx$) with each individual's labeling. If all strategic lies are benevolent when agents have Issue-wise distance based preferences then all strategic lies are benevolent when agents have IUO Issue-wise distance based preferences.*

Proof. This proof is similar to the one for Theorem 12, with the use of Lemma 2 (2) and Lemma 7. ■

Observation 13. *The skeptical aggregation operator is not strategy proof for neither Hamming set nor Hamming distance based preferences. Consider the three labelings in Figure 12. Labeling $\mathcal{L}_1$ of agent 1 when aggregated with $\mathcal{L}_2$ gives labeling $\mathcal{L}_3$, which disagrees with $\mathcal{L}_1$ on all three arguments. But, when the agent strategically lies and reports labeling $\mathcal{L}_2$ instead, the result of the aggregation is the same labeling $\mathcal{L}_2$, which differs only on two arguments $\{A,B\}$. The example is valid for both Hamming set and Hamming distance based preferences.*

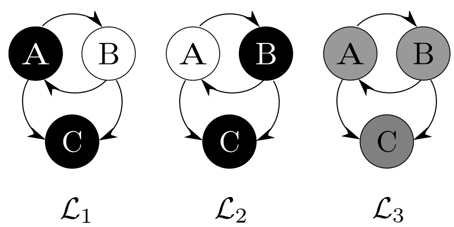

Figure 12: The skeptical operator is not strategy proof.

Observation 14. *The credulous (resp. super credulous) aggregation operator is not strategy proof for neither Hamming set nor Hamming distance based preferences. See the example in Figure 13. Labeling $\mathcal{L}_2$ of agent 2 when aggregated with $\mathcal{L}_1$ gives labeling $\mathcal{L}_{CO}$, which disagrees with $\mathcal{L}_2$ on the two arguments. But, when the agent strategically lies and reports $\mathcal{L}'_2$ instead, the result of the aggregation is $\mathcal{L}'_{CO}$, which matches the labeling $\mathcal{L}_2$. This lie by agent 2 makes the agent with labeling $\mathcal{L}_1$ worse off. The example is valid for both Hamming set and Hamming distance based preferences.*

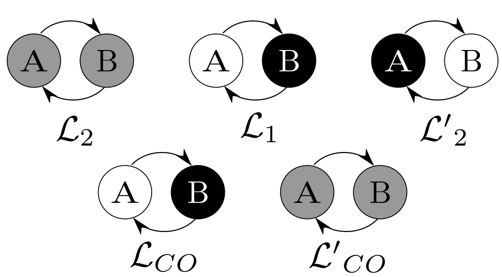

Figure 13: The (super) credulous operator is not strategy proof.

Theorem 14. *Consider the skeptical aggregation operator and Hamming set based preferences. For any agent, her strategic lies are benevolent.*

Proof. Let P be a profile, and $\mathcal{L}'_k$ a strategic lie of agent k. Denote $\mathcal{L}_{SO} = so_{\mathcal{AF}}(P)$ and $\mathcal{L}'_{SO} = so_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$. Agent k's preference is $\mathcal{L}'_{SO} \succ_k \mathcal{L}_{SO}$ (i). We will show that for any agent $i \neq k$, we have $\mathcal{L}'_{SO} \succ_i \mathcal{L}_{SO}$. Since the skeptical aggregation operator produces social outcomes that are less or equally committed to all the individual labelings, we have that $\mathcal{L}'_{SO} \sqsubseteq \mathcal{L}_i$ for all $i \neq k$ (ii). Similarly, we have $\mathcal{L}_{SO} \sqsubseteq \mathcal{L}_k$ (iii). From (i) and (iii), by Lemma 5, we have that $\mathcal{L}_{SO} \sqsubseteq \mathcal{L}'_{SO}$ (iv). From (iv) and (ii) we have $\mathcal{L}_{SO} \sqsubseteq \mathcal{L}'_{SO} \sqsubseteq \mathcal{L}_i$ for all $i \neq k$. Finally, we can apply Lemma 4 to obtain $\mathcal{L}'_{SO} \succeq_i \mathcal{L}_{SO}$ for all $i \neq k$ (v). We showed that a lie cannot be malicious, now we show that it is benevolent.

(iii) implies $\texttt{undec}(\mathcal{L}_k) \subseteq \texttt{undec}(\mathcal{L}_{SO})$ (vi). (i) and (vi) imply $\exists A \in \texttt{dec}(\mathcal{L}_k) : A \in \texttt{undec}(\mathcal{L}_{SO}) \wedge A \in \texttt{dec}(\mathcal{L}'_{SO})$ (vii). From (vii), (ii) and (v) $\mathcal{L}'_{SO} \succ_i \mathcal{L}_{SO}$ for $i \neq k$. ■

Theorem 15. *Consider the skeptical aggregation operator and Hamming distance based preferences. For any agent, her strategic lies are benevolent.*

Proof. Let P be a profile, and $\mathcal{L}'_k$ a strategic lie of agent k whose most preferred labeling is $\mathcal{L}_k$. Denote $\mathcal{L}_{SO} = so_{\mathcal{AF}}(P)$ and $\mathcal{L}'_{SO} = so_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$. We will show that, if $\mathcal{L}'_{SO}$ is strictly preferred to $\mathcal{L}_{SO}$ by agent k, then it is also strictly preferred by any other agent. Without loss of generality we can take agent $j, j \neq k$,whose most preferred labeling is $\mathcal{L}_j$.

Let us partition the arguments into the following disjoint groups:

- $\mathcal{X} = \texttt{dec}(\mathcal{L}_{SO}) \setminus \texttt{dec}(\mathcal{L}'_{SO})$ (decided arguments that became undecided).
- $\mathcal{Y} = \texttt{dec}(\mathcal{L}'_{SO}) \setminus \texttt{dec}(\mathcal{L}_{SO})$ (undecided arguments that became decided).
- $\mathcal{Z} = \texttt{dec}(\mathcal{L}'_{SO}) \cap \texttt{dec}(\mathcal{L}_{SO})$ (arguments decided in both labelings).
- $\mathcal{V} = \texttt{undec}(\mathcal{L}'_{SO}) \cap \texttt{undec}(\mathcal{L}_{SO})$ (arguments undecided in both labelings).

Labelings $\mathcal{L}_{SO}$ and $\mathcal{L}'_{SO}$ agree on the arguments in $\mathcal{V}$ (which are labeled `undec`) and $\mathcal{Z}$ (whose arguments are labeled `in` or `out`). For the arguments in $\mathcal{Z}$ there are no $\texttt{in} - \texttt{out}$ conflicts between $\mathcal{L}_{SO}$ and $\mathcal{L}'_{SO}$ as the skeptical aggregation operator guarantees social outcomes less or equally committed than $\mathcal{L}_j$. Therefore, only arguments from $\mathcal{X}$ and $\mathcal{Y}$ have an impact on the Hamming distance.

Both labelings $\mathcal{L}_k$ and $\mathcal{L}_j$ agree with $\mathcal{L}_{SO}$ on the arguments in $\mathcal{X}$ because $\mathcal{L}_{SO}$ decides on those arguments and is less or equally committed than both labelings. On the other side, $\mathcal{L}'_{SO}$ remains undecided on the arguments in $\mathcal{X}$ so both labelings $\mathcal{L}_k$ and $\mathcal{L}_j$ disagree with $\mathcal{L}'_{SO}$ on $\mathcal{X}$.

$\mathcal{L}'_{SO}$ is less or equally committed than $\mathcal{L}_j$ so, as above, we obtain that on the arguments in $\mathcal{Y}$, $\mathcal{L}_j$ agrees with $\mathcal{L}'_{SO}$ and disagrees with $\mathcal{L}_{SO}$. On the contrary, $\mathcal{L}'_{SO}$ does not have to be less or equally committed than $\mathcal{L}_k$ and so, for agent k, some of the arguments from $\mathcal{Y}$ increase the distance and some of them decrease. If agent k prefers $\mathcal{L}'_{SO}$ to $\mathcal{L}_{SO}$, then the number of the arguments decreasing the distance must be greater than the number of those increasing by more than $|\mathcal{X}|$. But for agent j all the arguments from $\mathcal{Y}$ are decreasing the distance, as $\mathcal{L}_j$ agrees with $\mathcal{L}'_{SO}$ on the whole $\mathcal{Y}$. So, if agent k gains by switching to labeling $\mathcal{L}'_{SO}$, agent j needs to gain at least the same. ■

Observation 15. *The skeptical aggregation operator is not strategy proof for neither Issue-wise set nor Issue-wise distance based preferences. Consider the three labelings in Figure 14.*[11] *Labeling $\mathcal{L}_1$ of agent 1 when aggregated with $\mathcal{L}_2$ gives labeling $\mathcal{L}_3$, which disagrees with $\mathcal{L}_1$ on both of the two issues. But, when the agent strategically lies and reports labeling $\mathcal{L}_2$ instead, the result of the aggregation is the same labeling $\mathcal{L}_2$, which differs only on one issue $\{\{A,B\}\}$. The example is valid for both Issue-wise set and Issue-wise distance based preferences.*

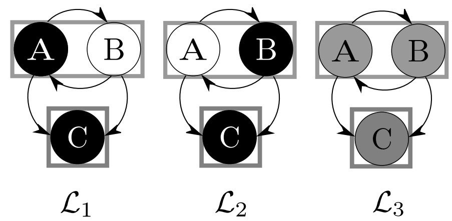

Figure 14: The skeptical operator is not strategy proof. The set of arguments located in one box form an issue.

Observation 16. *The credulous (resp. super credulous) aggregation operator is not strategy proof for neither Issue-wise set nor Issue-wise distance based preferences. In Figure 15,*[12] *labeling $\mathcal{L}_2$ of agent 2 when aggregated with $\mathcal{L}_1$ gives labeling $\mathcal{L}_{CO}$, which disagrees with $\mathcal{L}_2$ on the one and only issue. But, when the agent strategically lies and reports $\mathcal{L}'_2$ instead, the result of the aggregation is $\mathcal{L}'_{CO}$, which matches the labeling $\mathcal{L}_2$. This lie by agent 2 makes the agent with labeling $\mathcal{L}_1$ worse off. The example is valid for both Issue-wise set and Issue-wise distance based preferences.*

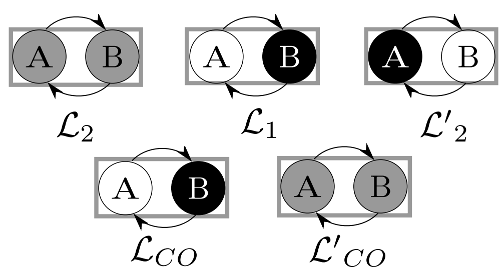

Figure 15: The (super) credulous operator is not strategy proof. The set of arguments located in one box form an issue.

Theorem 16. *Consider the skeptical aggregation operator and Issue-wise distance based preferences. For any agent, her strategic lies are benevolent.*

Proof. Let P be a profile of labelings, and $\mathcal{L}'_k$ be a strategic lie of agent k whose most preferred labeling is $\mathcal{L}_k$. Denote $\mathcal{L}_{SO} = so_{\mathcal{AF}}(P)$ and $\mathcal{L}'_{SO} = so_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$. We show that if $\mathcal{L}'_{SO}$ is strictly

[11]This figure is the same as Fig 12 with issues being evidenced.
[12]This figure is the same as Fig 13 with issues being evidenced.

preferred by an agent k then it is also strictly preferred by any other agent. Without loss of generality, we can take agent j, $j \neq k$, whose most preferred labeling is $\mathcal{L}_j$.

Let $\mathcal{I}_{de}(\mathcal{L})$ (resp. $\mathcal{I}_{un}(\mathcal{L})$) be the set of issues, each of which has arguments that are only decided (resp. undecided) according to $\mathcal{L}$. We call $\mathcal{I}_{de}(\mathcal{L})$ (resp. $\mathcal{I}_{un}(\mathcal{L})$) a decided (resp. undecided) issue (w.r.t $\mathcal{L}$). Let us partition the issues into the following disjoint groups:

- $\mathcal{X} = \mathcal{I}_{de}(\mathcal{L}_{SO}) \setminus \mathcal{I}_{de}(\mathcal{L}'_{SO})$ (decided issues that became undecided).
- $\mathcal{Y} = \mathcal{I}_{de}(\mathcal{L}'_{SO}) \setminus \mathcal{I}_{de}(\mathcal{L}_{SO})$ (undecided issues that became decided).
- $\mathcal{Z} = \mathcal{I}_{de}(\mathcal{L}_{SO}) \cap \mathcal{I}_{de}(\mathcal{L}'_{SO})$ (issues decided in both labelings).
- $\mathcal{V} = \mathcal{I}_{un}(\mathcal{L}_{SO}) \cap \mathcal{I}_{un}(\mathcal{L}'_{SO})$ (issues undecided in both labelings).

The rest is similar to Theorem 15, but using issues instead of arguments. ■

Theorem 17. *The skeptical aggregation operator is strategy proof when individuals have IUO Hamming sets based preferences.*

Proof. Let P be a profile, $\mathcal{L}_k$ be the top preference of agent k, and $\mathcal{L}'_k \neq \mathcal{L}_k$ be any potential lie that agent k might consider. Denote $\mathcal{L}_{SO} = so_{\mathcal{AF}}(P)$ and $\mathcal{L}'_{SO} = so_{\mathcal{AF}}(P_{\mathcal{L}_k/\mathcal{L}'_k})$. We will show that $\neg(\mathcal{L}'_{SO} \succ_{k,\ominus^{\mathcal{M}}} \mathcal{L}_{SO})$. Which means, we need to show:

$$\neg((\mathcal{L}'_{SO} \succeq_{k,\ominus^{\mathcal{M}}} \mathcal{L}_{SO}) \wedge \neg(\mathcal{L}_{SO} \succeq_{k,\ominus^{\mathcal{M}}} \mathcal{L}'_{SO})) \tag{42}$$

$$\neg(\mathcal{L}'_{SO} \succeq_{k,\ominus^{\mathcal{M}}} \mathcal{L}_{SO}) \vee (\mathcal{L}_{SO} \succeq_{k,\ominus^{\mathcal{M}}} \mathcal{L}'_{SO}) \tag{43}$$

In other words:

$$\begin{aligned}\neg((\mathcal{L}'_{SO} \ominus^{io} \mathcal{L}_k \subseteq \mathcal{L}_{SO} \ominus^{io} \mathcal{L}_k) \wedge (\mathcal{L}'_{SO} \ominus^{du} \mathcal{L}_k \subseteq \mathcal{L}_{SO} \ominus^{du} \mathcal{L}_k))\\ \vee(\mathcal{L}_{SO} \succeq_{k,\ominus^{\mathcal{M}}} \mathcal{L}'_{SO})\end{aligned} \tag{44}$$

To reformulate, we only need to show that one of the following holds:

1. $\neg(\mathcal{L}'_{SO} \ominus^{io} \mathcal{L}_k \subseteq \mathcal{L}_{SO} \ominus^{io} \mathcal{L}_k)$, or
2. $\neg(\mathcal{L}'_{SO} \ominus^{du} \mathcal{L}_k \subseteq \mathcal{L}_{SO} \ominus^{du} \mathcal{L}_k)$, or
3.
   (a) $\mathcal{L}_{SO} \ominus^{io} \mathcal{L}_k \subseteq \mathcal{L}'_{SO} \ominus^{io} \mathcal{L}_k$, and
   (b) $\mathcal{L}_{SO} \ominus^{du} \mathcal{L}_k \subseteq \mathcal{L}'_{SO} \ominus^{du} \mathcal{L}_k$.

First, by definition, $\mathcal{L}_{SO}$ is less or equally committed ($\sqsubseteq$) than $\mathcal{L}_k$. So, $\mathcal{L}_{SO} \ominus^{io} \mathcal{L}_k = \emptyset$, However, this is not the case for $\mathcal{L}'_{SO}$ and $\mathcal{L}_k$. So, $\mathcal{L}'_{SO} \ominus^{io} \mathcal{L}_k$ might not be an empty set. Hence, $\mathcal{L}_{SO} \ominus^{io} \mathcal{L}_k \subseteq \mathcal{L}'_{SO} \ominus^{io} \mathcal{L}_k$ i.e. (3)(a) is satisfied. Now we show that either (1),(2) or (3)(b) is satisfied.

Suppose (1) and (2) are violated and we will show that (3)(b) is then satisfied. This shows that (1), (2), and (3)(b) cannot be all violated together.

Since (1) is violated and since $\mathcal{L}_{SO} \ominus^{io} \mathcal{L}_k = \emptyset$ then $\mathcal{L}'_{SO} \ominus^{io} \mathcal{L}_k = \emptyset$ (i). Since (2) is violated then $\forall a : (a \in \mathcal{L}'_{SO} \ominus^{du} \mathcal{L}_k \Rightarrow a \in \mathcal{L}_{SO} \ominus^{du} \mathcal{L}_k)$ (ii). Note that $\forall a : (a \in \mathcal{L}'_{SO} \ominus^{du} \mathcal{L}_k \Rightarrow (a \in \texttt{undec}(\mathcal{L}'_{SO}) \wedge a \in \texttt{dec}(\mathcal{L}_k)))$(iii). Otherwise, we would have $a \in \texttt{dec}(\mathcal{L}'_{SO}) \wedge a \in \texttt{undec}(\mathcal{L}_k)$ and from (ii) we would have $a \in \texttt{dec}(\mathcal{L}_{SO}) \wedge a \in \texttt{undec}(\mathcal{L}_k)$ which contradicts $\mathcal{L}_{SO} \sqsubseteq \mathcal{L}_k$.

From (i) and (iii), $\forall a \in \texttt{in}(\mathcal{L}'_{SO}) \Rightarrow a \in \texttt{in}(\mathcal{L}_k)$ (iv) (from (i), $\mathcal{L}_k(a) \neq \texttt{out}$, and from (iii), $\mathcal{L}_k(a) \neq \texttt{undec}$). Similarly, from (i) and (iii), $\forall a \in \texttt{out}(\mathcal{L}'_{SO}) \Rightarrow a \in \texttt{out}(\mathcal{L}_k)$ (v). From (iv) and (v), $\mathcal{L}'_{SO} \sqsubseteq \mathcal{L}_k$. Since $\forall i \neq k$: $\mathcal{L}'_{SO} \sqsubseteq \mathcal{L}_i$, then $\forall i \in Ag$: $\mathcal{L}'_{SO} \sqsubseteq \mathcal{L}_i$. By Theorem 1, $\mathcal{L}'_{SO} \sqsubseteq \mathcal{L}_{SO}$. Then, $\texttt{undec}(\mathcal{L}_{SO}) \subseteq \texttt{undec}(\mathcal{L}'_{SO})$ (vi).

Now, $\forall a \in \mathcal{L}_{SO} \ominus^{du} \mathcal{L}_k$ then $a \in \texttt{undec}(\mathcal{L}_{SO}) \wedge a \in \texttt{dec}(\mathcal{L}_k)$. From (vi), $a \in \texttt{undec}(\mathcal{L}'_{SO})$. Thus, $a \in \mathcal{L}'_{SO} \ominus^{du} \mathcal{L}_k$. Then, (3)(b) is satisfied. ■

Observation 17. *The skeptical aggregation operator is not strategy proof when individuals have IUO Hamming distance based preferences. Consider the three labelings in Figure 16. Labeling $\mathcal{L}_1$ of agent 1 when aggregated (using skeptical operator) with $\mathcal{L}_2$ gives labeling $\mathcal{L}_3$, which differs from $\mathcal{L}_1$ on all five arguments with respect to* $\texttt{dec} - \texttt{undec}$ *Hamming set. Then, $\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_3 = 2 \times 0 + 1 \times 5 = 5$. But, when the agent strategically lies and reports labeling $\mathcal{L}_2$ instead, the result of the aggregation is the same labeling $\mathcal{L}_2$, which differs only on two arguments $\{A, B\}$ with respect to* $\texttt{in} - \texttt{out}$ *Hamming set. Then, $\mathcal{L}_1 \left|\ominus^{\mathcal{M}}\right| \mathcal{L}_2 = 2 \times 2 + 1 \times 0 = 4$.*

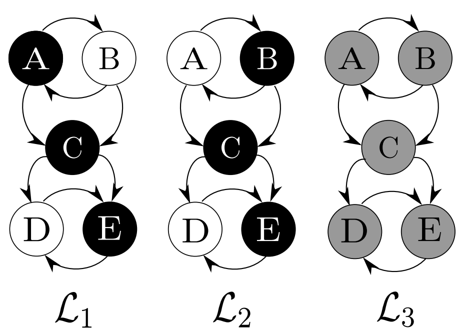

Figure 16: The skeptical operator is not strategy proof when agents have IUO Hamming distance preferences.

Observation 18. *The credulous (resp. super credulous) aggregation operator is not strategy proof for neither IUO Hamming sets nor IUO Hamming distance based preferences. The example in Figure 13 can serve as a counterexample for the case where individuals have IUO Hamming sets (or IUO Hamming distance) based preferences. The agent with labeling $\mathcal{L}_2$ can insincerely report $\mathcal{L}'_2$ to obtain her preferred labeling. This makes an agent with labeling $\mathcal{L}_1$ worse off.*

Proposition 9. *Consider the skeptical aggregation operator and IUO Hamming distance based preferences. For any agent, her strategic lies are benevolent.*

Proof. From Theorem 15 and Theorem 12, the strategic lies are benevolent when individuals have IUO Hamming distance based. ■

Observation 19. *The skeptical aggregation operator is not strategy proof when individuals have IUO Issue-wise distance based preferences. Consider the three labelings in Figure 17.*[13] *Labeling $\mathcal{L}_1$ of agent 1 when aggregated (using skeptical operator) with $\mathcal{L}_2$ gives labeling $\mathcal{L}_3$, which differs from $\mathcal{L}_1$ on all three issues with respect to* $\texttt{dec} - \texttt{undec}$ *Issue-wise set. Then,* $\mathcal{L}_1 \left|\ominus^{\mathcal{M}}_{\mathcal{W}}\right| \mathcal{L}_3 = 2 \times 0 + 1 \times 3 = 3$. *But, when the agent strategically lies and reports labeling $\mathcal{L}_2$ instead, the result of the aggregation is the same labeling $\mathcal{L}_2$, which differs only on one is $\{\{A,B\}\}$ with respect to* $\texttt{in} - \texttt{out}$ *Issue-wise set. Then,* $\mathcal{L}_1 \left|\ominus^{\mathcal{M}}_{\mathcal{W}}\right| \mathcal{L}_2 = 2 \times 1 + 1 \times 0 = 2$.

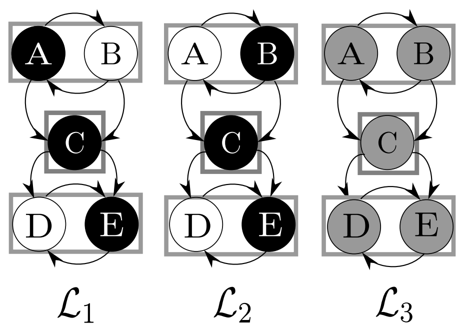

Figure 17: The skeptical operator is not strategy proof when agents have IUO Issue-wise distance preferences. The set of arguments located in one box form an issue.

Observation 20. *The credulous and super credulous aggregation operators are not strategy proof when individuals have IUO Issue-wise sets (resp. distance) based preferences. The example in Figure 15 can serve as a counter example for the case where individuals have IUO Issue-wise sets (resp. distance) based preferences. The agent with labeling $\mathcal{L}_2$ can insincerely report $\mathcal{L}_2'$ to obtain her preferred labeling. This makes an agent with labeling $\mathcal{L}_1$ worse off.*

Proposition 10. *Consider the skeptical aggregation operator and IUO Issue-wise distance based preferences. For any agent, her strategic lies are benevolent.*

Proof. From Theorem 16 and Theorem , the strategic lies are benevolent when individuals have IUO Issue-wise distance based. ■

Theorem 18. *Let $\mathcal{F}$ be the set of all possible classes of preferences, $\mathcal{R}$ be some set s.t. $\mathcal{R} \subseteq \mathcal{F}$, and Ag be the set of agents. If an operator is strategy proof given that Ag have homogeneous preferences from $\mathcal{R}$, then it is strategy proof given that Ag have heterogeneous preferences from $\mathcal{R}$.*

[13]This figure is the same as Fig 16 with issues being evidenced.

Proof. Let Op be an operator that is strategy proof given that Ag have homogeneous preferences from $\mathcal{R}$ (i.e. $\forall i, j \in Ag\ c(i) = c(j) \in \mathcal{R}$). Then, there exists no single agent j that has an incentive to lie about her preferences (given that all agents have the same class of preferences). Note that agent j has no incentive to lie given that the submitted labelings by all agents other than j are fixed. Hence, the classes of preferences that are assumed for any agent $k \neq j$ do not affect the incentive of agent j to lie or otherwise. Thus, the same result would hold whether other agents' preferences are different from $c(j)$ or not. ■