Judgment Aggregation in Abstract Argumentation

Gabriella Pigozzi

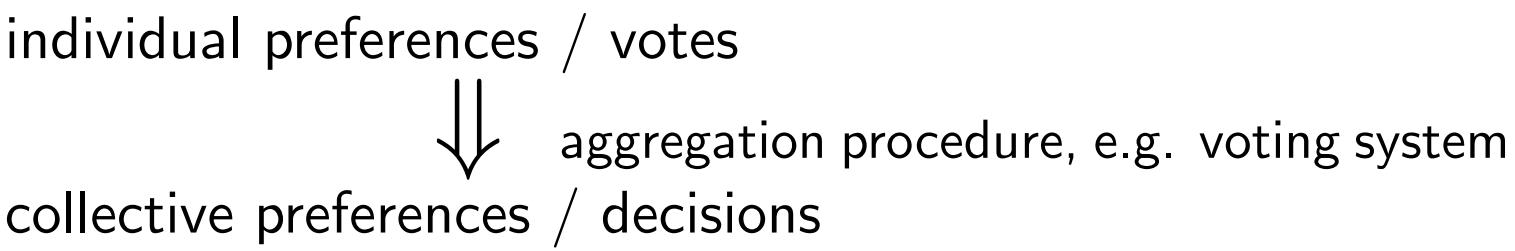
Université Paris-Dauphine
(Joint work with Martin Caminada and Mikolaj Podlaszewski)

Evidence Based Policy Making Workshop
Paris, December 2-3, 2010

- Social choice theory (SCT) addresses collective decision problems.
- SCT focus on the aggregation of individual preferences into collective outcomes. Such models focus primarily on collective choices between alternative outcomes such as candidates, policies or actions.
- However, they do not capture decision problems in which a group has to form collectively endorsed beliefs or judgments on logically interconnected propositions.
- This step has been taken by judgment aggregation.

Social choice theory

Social choice theory models collective decisions as processes of aggregating individual inputs into collective outputs.



Aggregation problems (1)

The *first aggregation problem* (1770): the Marquis de Condorcet proposed a method for the aggregation of preferences which led to the (first) voting paradox:

$$\begin{array}{l} \textit{Person 1: } x > y > z \\ \textit{Person 2: } y > z > x \\ \textit{Person 3: } z > x > y \end{array} \quad \Rightarrow \quad \textit{Group: } x > y > z > x$$

Aggregation problems (2)

Judgment aggregation (JA):

$$(P \wedge Q) \leftrightarrow R$$

	P	Q	R
Individual 1	yes	no	no
Individual 2	no	yes	no
Individual 3	yes	yes	yes
Majority	yes	yes	no

Aggregation problems (3)

The multiple elections paradox [Brams, Kilgour and Zwicker, 1998]:

Voter 1	yes	yes	no
Voter 2	yes	yes	no
Voter 3	yes	no	yes
Voter 4	yes	no	yes
Voter 5	no	yes	yes
Voter 6	no	yes	yes
Voter 7	no	yes	yes
Voter 8	no	yes	yes
Voter 9	yes	no	no
Voter 10	yes	no	no
Majority	yes	yes	yes

Aggregation problems (4)

- *Item-by-item* majority rule may generate inconsistent collective outcomes.
- Bad news: any aggregation procedure that satisfies some desirable properties is condemned to produce sometimes irrational outcomes. ☹

Motivation (1)

- The Condorcet paradox produces a meaningless social outcome.
- The judgment aggregation paradox is meaningless but may also be arbitrary in the sense of the multiple election problem.
- The multiple election paradox produces arbitrary election outcomes.
- Research question: when is a social outcome compatible (*cfr.* legitimate) with the individual positions?
- (Small group) decisions where **any individual has to be able to defend the collective position** $\Rightarrow$ The group outcome is compatible with its members views $\Rightarrow$ It's neither arbitrary nor meaningless, hence the members can defend it and be held *responsible* for it.

Motivation (2)

- On the one hand, sharing information helps making better decisions.
- On the other hand, by pooling private information, agents expose themselves to other people manipulation (*e.g.* individuals with different interests).
- Is there a way to reconcile these two aspects?

Methodology

- Methodology: abstract argumentation. How can individual evaluations of the same argumentation framework be mapped into a collective one?
- Agents have access to the same *evidence* and can interpret it in different ways.
- Two aggregation operators that guarantee a *unique*, *compatible* and *rational* outcome.

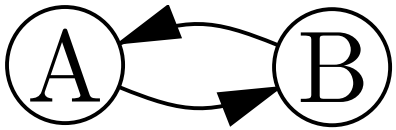
Argumentation framework

- Argumentation framework: a set of arguments and a *defeat* relation among them: $AF = (Ar, def)$.
- Argumentation theory identifies and characterizes the sets of arguments (extensions) that can reasonably survive the conflicts expressed in the argumentation framework.
- An argumentation framework specifies a *directed graph*:

  $$C \rightarrow B \rightarrow A$$

  Which of these arguments should be ultimately accepted?

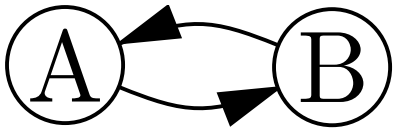
Nixon



A Nixon is a pacifist because be is a quaker.

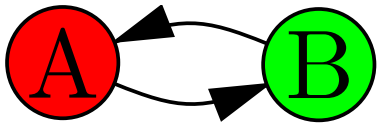
B Nixon is not a pacifist because he is republican.

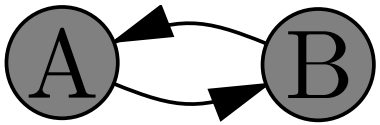
Nixon

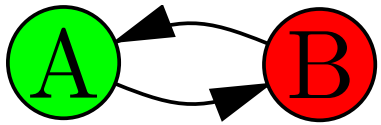


A Nixon is a pacifist because be is a quaker.

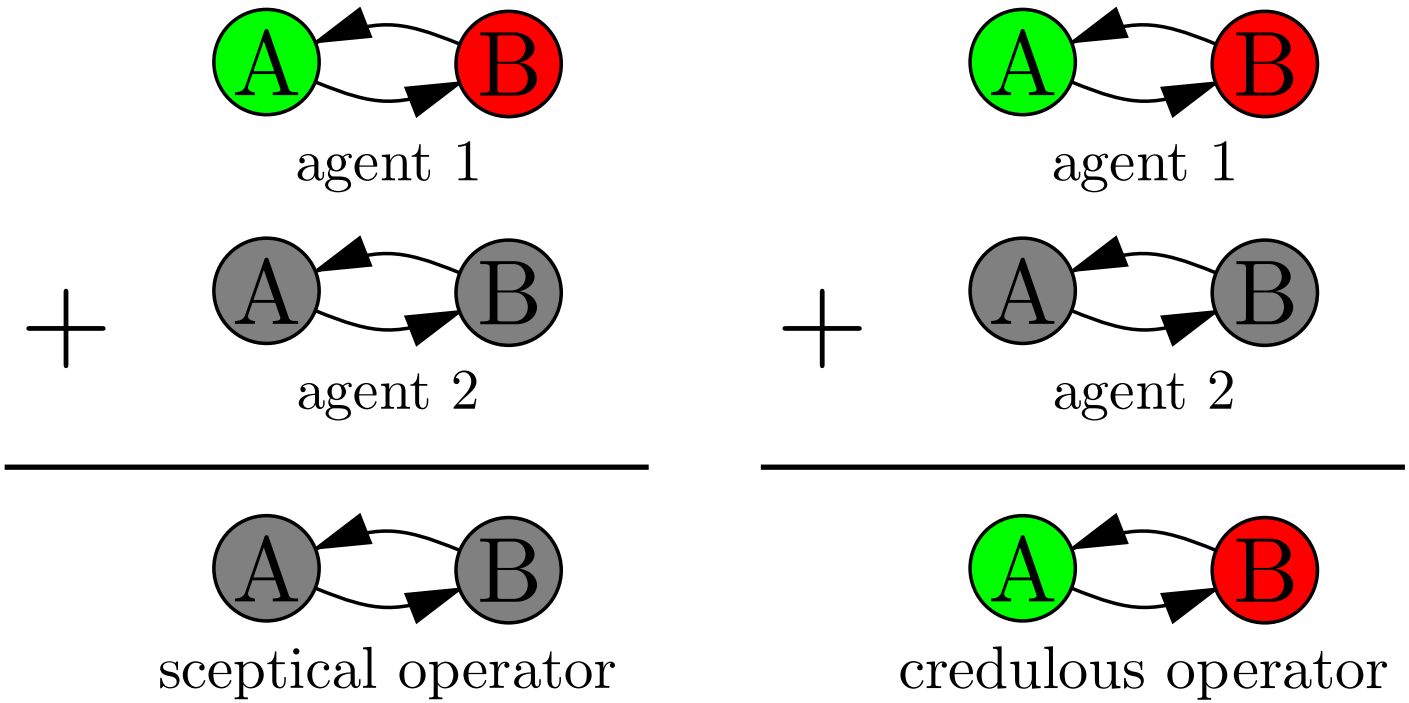
B Nixon is not a pacifist because he is republican.







Sceptical and Credulous Operator

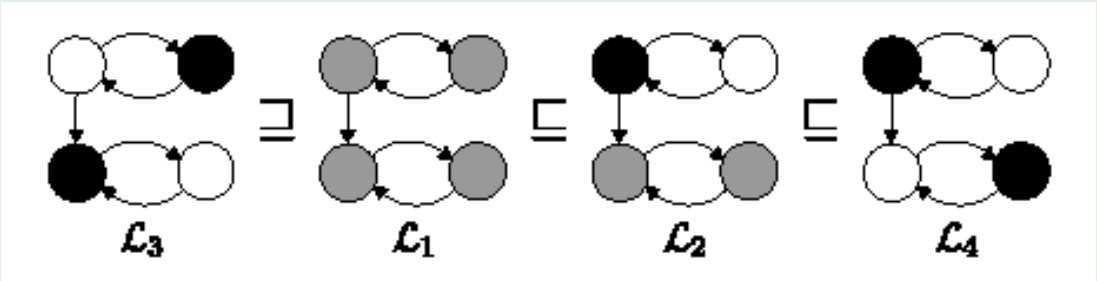


Compatibility

Definition ($\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$)

$\mathcal{L}_1$ is less or equally committed as $\mathcal{L}_2$ ($\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$) iff $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2)$ and $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2)$.

Example



Definition ($\mathcal{L}_1 \approx \mathcal{L}_2$)

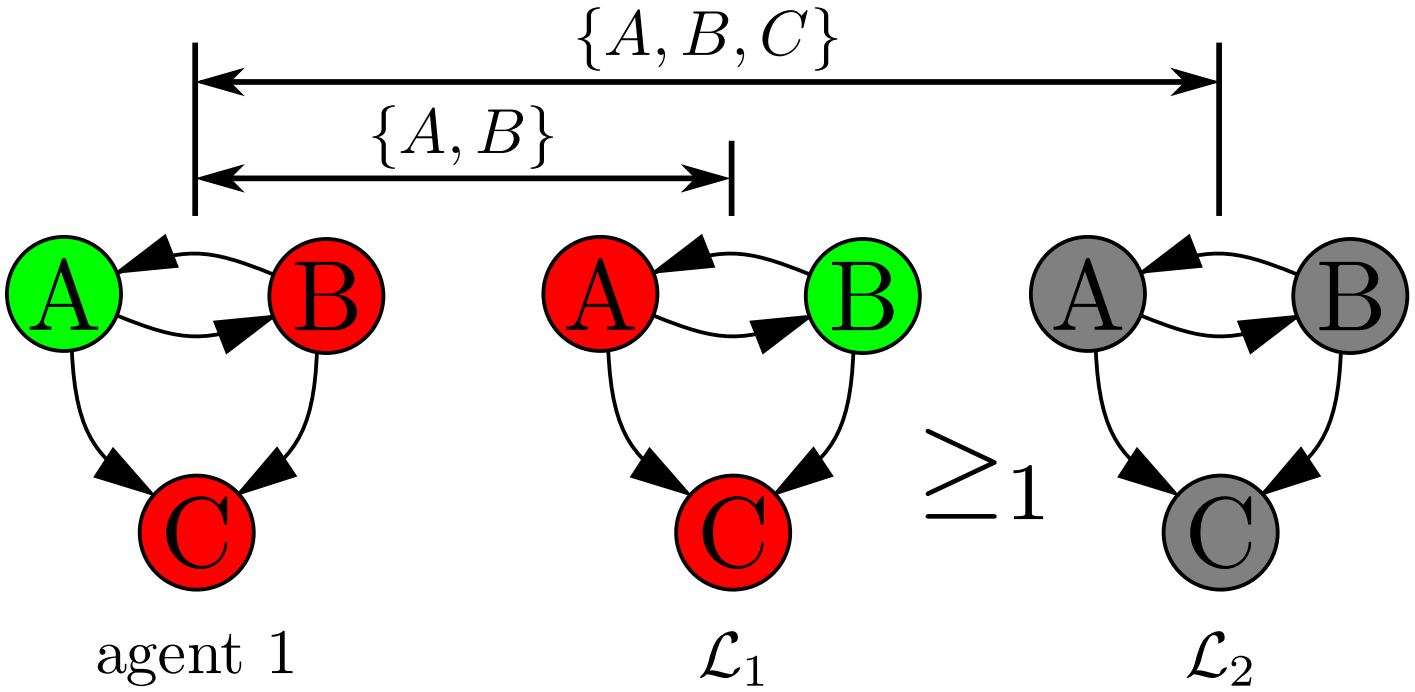
$\mathcal{L}_1$ is compatible with $\mathcal{L}_2$ ($\mathcal{L}_1 \approx \mathcal{L}_2$) iff $\mathtt{in}(\mathcal{L}_1) \cap \mathtt{out}(\mathcal{L}_2) = \emptyset$ and $\mathtt{out}(\mathcal{L}_1) \cap \mathtt{in}(\mathcal{L}_2) = \emptyset$.

Introducing preferences

- Although every social outcome that is compatible with one's own labelling is acceptable, some outcomes are more acceptable than others.
- A collective outcome is more acceptable than another if it is compatible and more similar to one's own position than the other $\Rightarrow$ we introduce the notion of distance among labellings:

1. Are the social outcomes of our aggregation operators Pareto optimal?
2. Do agents have an incentive to misrepresent their own opinion in order to obtain a more favourable outcome? And if so, what are the effects of this from the perspective of social welfare?

Preferences



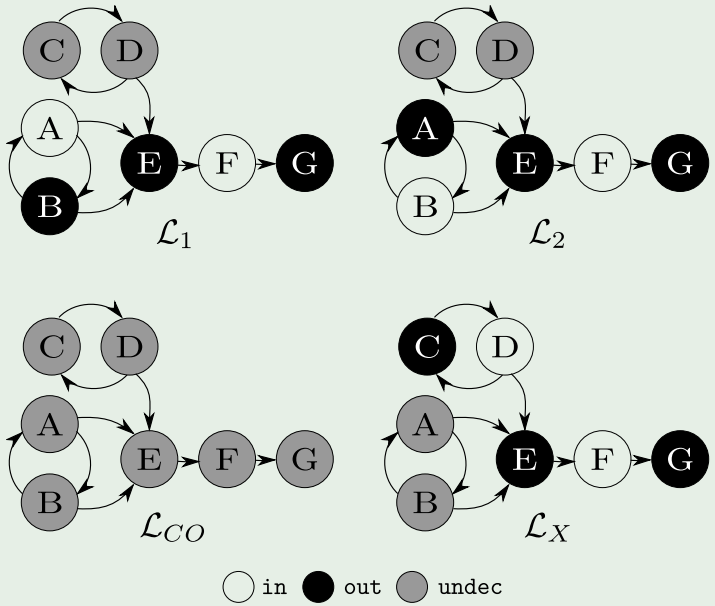
Pareto optimality

Pareto optimality guarantees that it is not possible to improve a social outcome, i.e. it is not possible to make one individual better off without making at least one other person worse off.

Pareto optimality (3)

The credulous aggregation operator is not Pareto optimal when the preferences are *Hamming distance* based. Both $\mathcal{L}_{CO}$ and $\mathcal{L}_X$ are compatible with $\mathcal{L}_1$ and $\mathcal{L}_2$, but $\mathcal{L}_X$ is closer when HD is used. $\mathcal{L}_1 \ominus \mathcal{L}_{CO} = \mathcal{L}_2 \ominus \mathcal{L}_{CO} = \{A, B, E, F, G\}$, so HD is 5, whereas $\mathcal{L}_1 \ominus \mathcal{L}_X = \mathcal{L}_2 \ominus \mathcal{L}_X = \{A, B, C, D\}$, so HD is 4.

Example



Pareto optimality

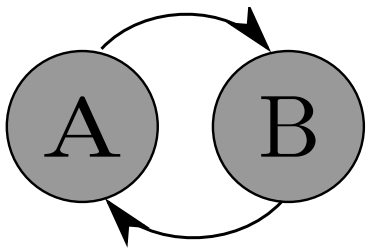
	Sceptical Operator	Credulous Operator
Hamming set	Yes	Yes
Hamming distance	Yes	No

Manipulation (1)

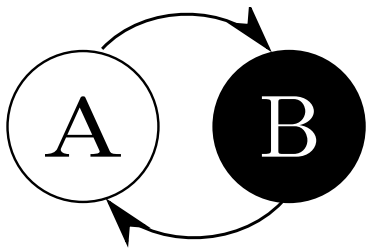
The operator is **strategy-proof** if no individual has an incentive to misrepresent his sincere opinion to obtain a collective outcome that is preferable in his individual perspective. In other words, the best strategy is to be honest.

Manipulation (2)

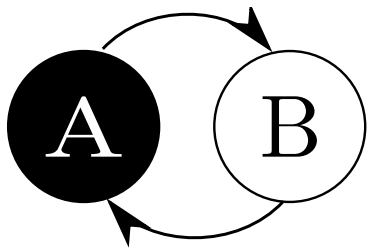
The credulous aggregation operator is not strategy-proof. Agent $\mathcal{L}_2$ can insincerely report $\mathcal{L}'_2$ to obtain his preferred labelling. This makes agent with labelling $\mathcal{L}_1$ worse off (valid for both Hamming set and Hamming distance based preferences).



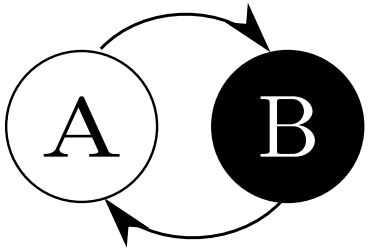
$\mathcal{L}_2$



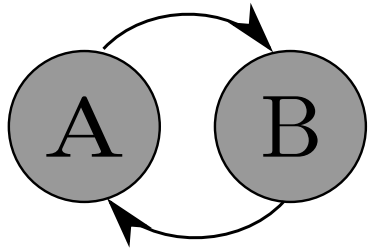
$\mathcal{L}_1$



$\mathcal{L}'_2$



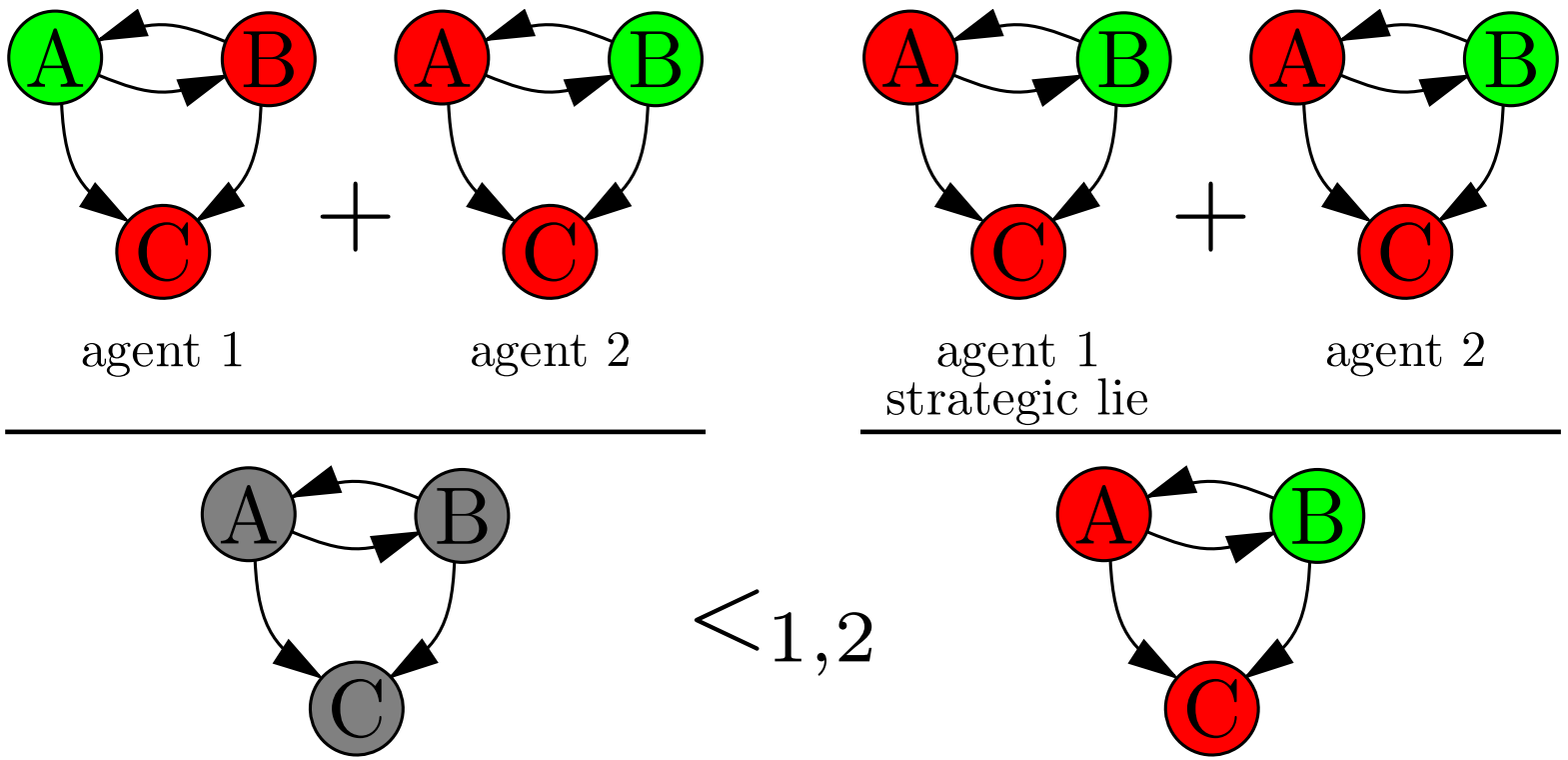
$\mathcal{L}_{CO}$



$\mathcal{L}'_{CO}$

Benevolent lie

The sceptical aggregation operator is not strategy-proof but its lies are benevolent.

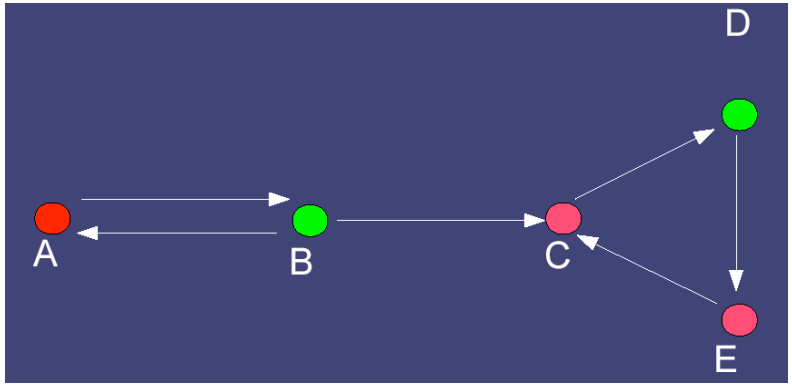


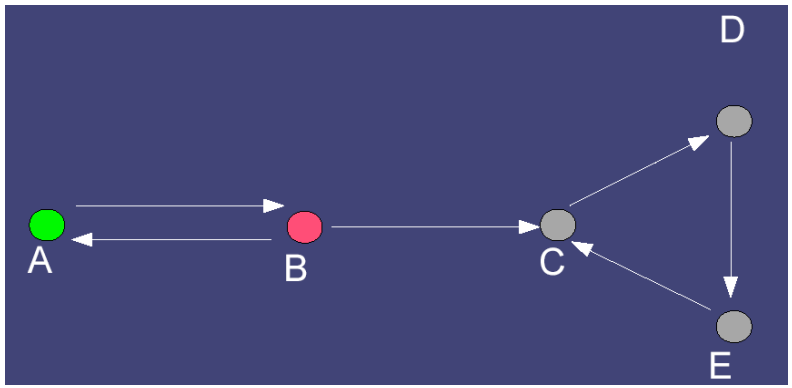
Strategy-proofness

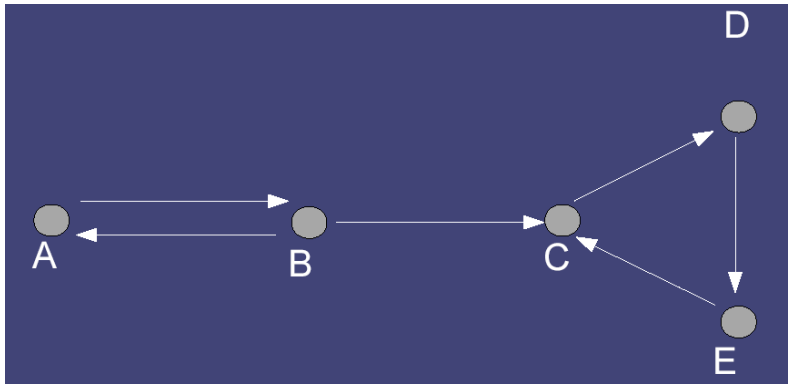
	Sceptical Operator	Credulous Operator
Hamming set	No but benevolent	No and not benevolent
Hamming distance	No but benevolent	No and not benevolent

Conclusion

- One lesson of judgment aggregation is that the aggregation on the evidence and on the recommendation may contradict each other (even when there is unanimity on the recommendation!).
- We introduced a notion of a social outcome that is neither arbitrary nor meaningless (compatibility).
- We defined aggregation operators that guarantee compatible outcomes $\Rightarrow$ 'consensus' aggregation operators.
- Sharing information may trigger strategic manipulation from agents who have different interests, but we have showed a benevolent type of lie.


D
A
B
C
E


D
A
B
C
E


D
A
B
C
E

Labelling based semantics

Definition

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

Definition

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

Definition

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

Why only conflict-free, admissible and complete labellings?

- Some semantics (preferred, stable or semi-stable) would give more than one collective outcome.
- On the other hand, a unique status semantics (like grounded) would be too restrictive as there would be only one reasonable possible position $\Rightarrow$ if disagreement is not possible, why do we need aggregation?
- Since each stable, semi-stable, preferred, or grounded labellings is also a complete (and therefore admissible and conflict-free) labelling, our framework is not too restrictive.

Conditions on labelling aggregation

F_{AF} is a labellings aggregation operator that assigns a collective labelling $\mathcal{L}_{Coll}$ to each profile $\{\mathcal{L}_1, \ldots, \mathcal{L}_n\}$.

Conditions (UD, CR, anonymity and independence) for F_{AF}:

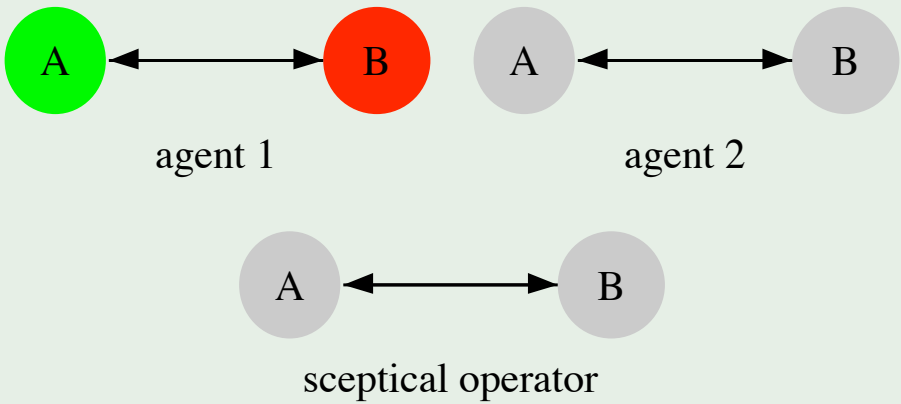
- **Universal domain**: The domain of F_{AF} is the set of all profiles of individual labellings belonging to semantics $\mathcal{T}_{conflict-free}$, $\mathcal{T}_{admissible}$ or $\mathcal{T}_{complete}$.
- **Collective rationality**: $F_{AF}(\{\mathcal{L}_1, \ldots, \mathcal{L}_n\})$ is a labelling belonging to semantics $\mathcal{T}_{conflict-free}$, $\mathcal{T}_{admissible}$ or $\mathcal{T}_{complete}$.

The sceptical aggregation (1)

First phase: the sceptical initial labelling ($\mathcal{L}_{sio}$):

- A is labelled *in* if everyone agrees A is *in*.
- A is labelled *out* if everyone agrees A is *out*.
- A is labelled *undec* in all other cases.

Example

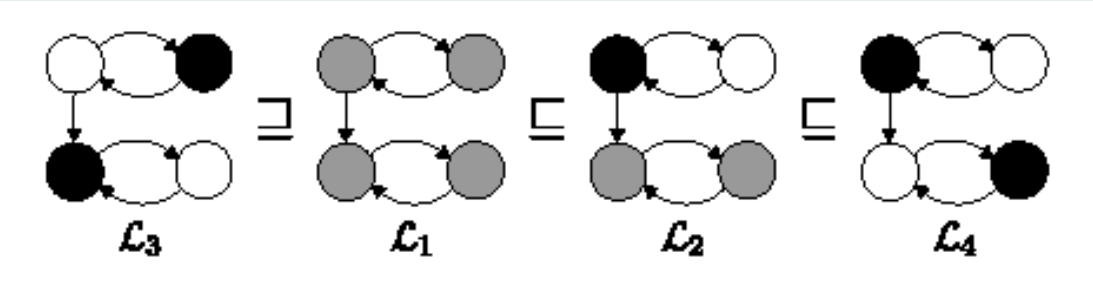


The sceptical aggregation (2)

Definition ($\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$)

$\mathcal{L}_1$ is less or equally committed as $\mathcal{L}_2$ ($\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$) iff $\mathtt{in}(\mathcal{L}_1) \subseteq \mathtt{in}(\mathcal{L}_2)$ and $\mathtt{out}(\mathcal{L}_1) \subseteq \mathtt{out}(\mathcal{L}_2)$.

Example



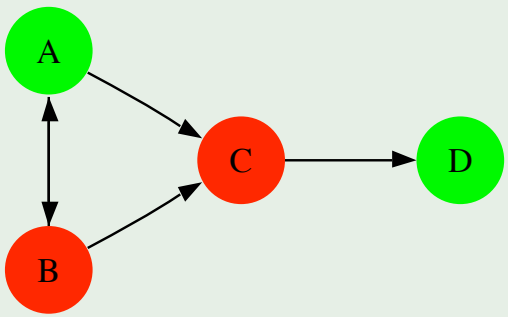
Lemma

$\mathcal{L}_{sio} \sqsubseteq \mathcal{L}_i$

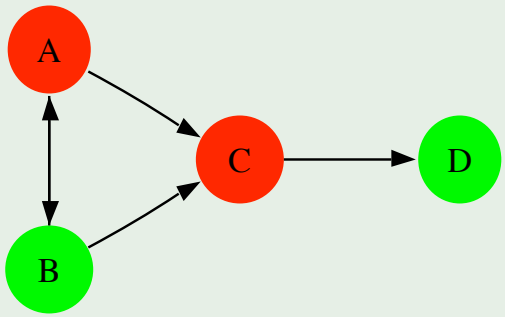
The sceptical aggregation (3)

Problem: $\mathcal{L}_{sio}$ violates collective rationality under any constraint stronger than conflict-freeness.

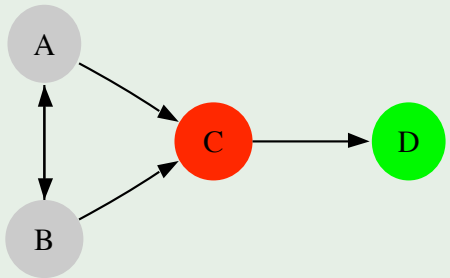
Example



agent 1



agent 2



sceptical initial violates collective rationality under admissibility

The sceptical aggregation (4)

Second phase (iteration): at the end the sceptical labelling ($\mathcal{L}_{so}$):

- **Contraction function** relabels an argument from *in* or *out* to *undec* $\Rightarrow$ contraction sequence of labellings until $\mathcal{L}_{so}$.
- An argument that is accepted without every defeater being rejected can no longer be accepted.
- An argument that is rejected without a defeater that is accepted can no longer be rejected.
- In each of these two cases, the group has to abstain (*undec*) on that argument.

The sceptical aggregation (5)

Theorem

$\mathcal{L}_{so}$ is the (unique) most committed admissible labelling that is less or equally committed than each input-labelling (each argument that is accepted/rejected by the group is also accepted/ rejected by each individual participant) : $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_i$.

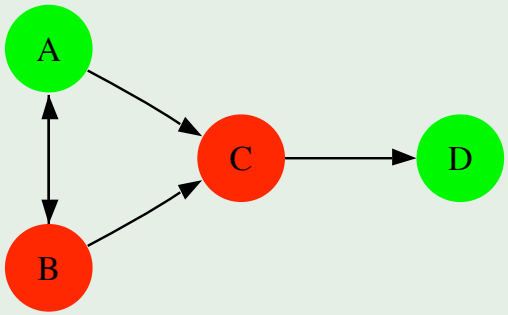
The group outcome is self-justifying.

$\mathcal{L}_{so}$ satisfies collective rationality under conflict-freeness, admissibility and completeness.

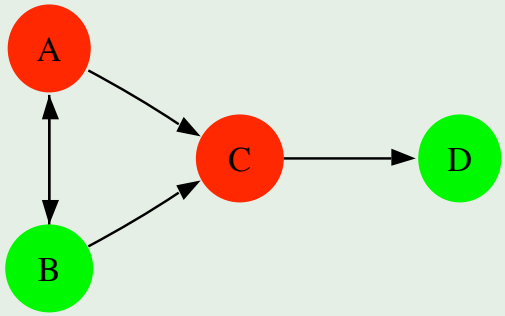
Unanimity (1)

Problem (?): sometimes $\mathcal{L}_{so}$ ignores unanimity.

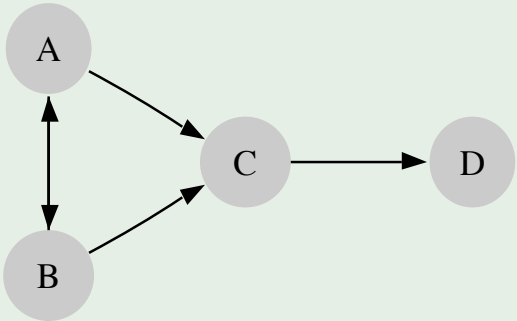
Example



agent 1



agent 2



sceptical outcome violates unanimity

Unanimity (2)

Cfr. floating conclusions: statements that are supported in each extension but by different arguments. In **default logic**, the sceptical approach states that a conclusion should be endorsed only if it is contained in every extension. But **Horty** questions the sceptical policy:

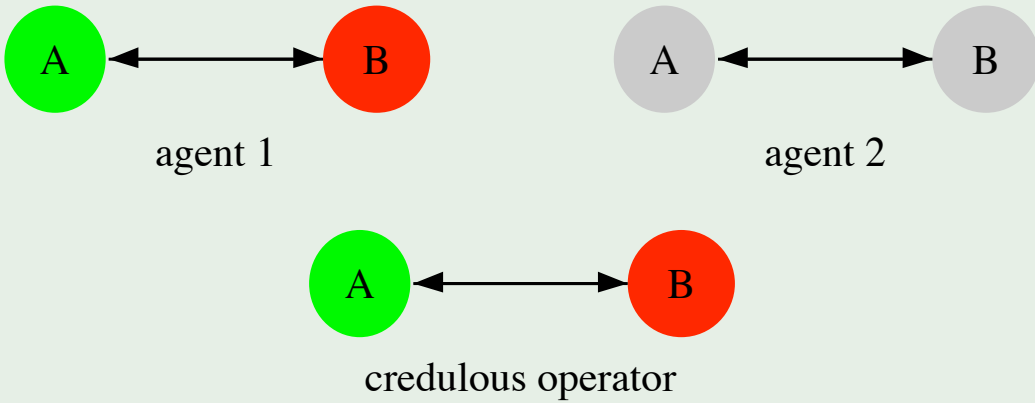
> *The point is not that floating conclusions might be wrong; any conclusion drawn through defeasible reasoning might be wrong. The point is that a statement supported only as floating conclusion seems to be less secure than the same statement when it is uniformly supported by a common argument.*

The credulous aggregation (1)

First phase: the credulous initial labelling ($\mathcal{L}_{cio}$):

- A is labelled *in* if someone thinks A is *in* and nobody thinks A is *out*.
- A is labelled *out* if someone thinks A is *out* and nobody thinks is *in*.
- A is labelled *undec* in all other cases.

Example



The credulous aggregation (2)

Definition ($\mathcal{L}_1 \approx \mathcal{L}_2$)

$\mathcal{L}_1$ is compatible with $\mathcal{L}_2$ ($\mathcal{L}_1 \approx \mathcal{L}_2$) iff $\mathtt{in}(\mathcal{L}_1) \cap \mathtt{out}(\mathcal{L}_2) = \emptyset$ and $\mathtt{out}(\mathcal{L}_1) \cap \mathtt{in}(\mathcal{L}_2) = \emptyset$.

Theorem

$\mathcal{L}_{cio}$ is **compatible** *with each input-labelling.*

$\sqsubseteq$ is stronger than $\approx$: if $\mathcal{L}_1 \sqsubseteq \mathcal{L}_2$, then $\mathcal{L}_1 \approx \mathcal{L}_2$.

Problem: $\mathcal{L}_{cio}$ violates collective rationality even under conflict-freeness (let alone under admissibility and completeness)!

The credulous aggregation (3)

Second phase (iteration): at the end the credulous labelling ($\mathcal{L}_{co}$):

> *Each argument that is accepted or rejected without a justification can no longer be accepted or rejected, so the group has to abstain on it.*

$\mathcal{L}_{co}$ is the most committed position that is less or equally committed than $\mathcal{L}_{cio}$: $\mathcal{L}_{co} \sqsubseteq \mathcal{L}_{cio}$.

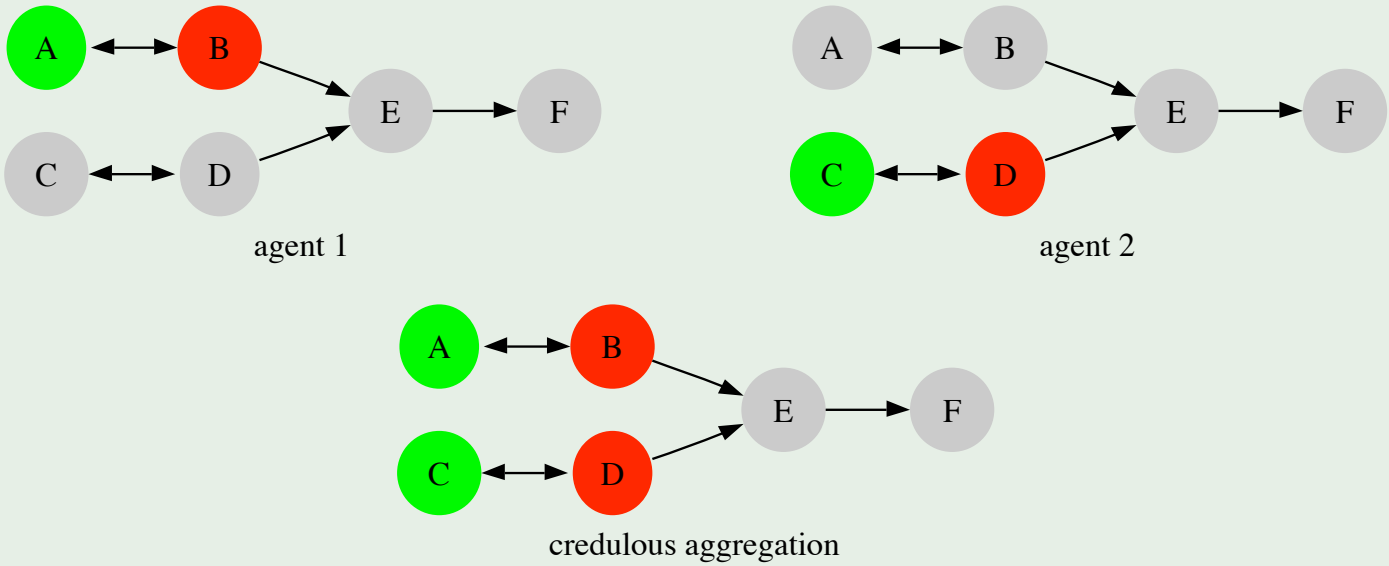
Theorem

$\mathcal{L}_{co}$ *is* **compatible** *with each input-labelling* $\mathcal{L}_i$.

The credulous aggregation (4)

$\mathcal{L}_{co}$ satisfies collective rationality under conflict-freeness and admissibility (but **not** under completeness).

Example



$\mathcal{L}_{co}$ can ignore unanimity.

Relevance of the participants' inputs

1. The credulous outcome labelling is bigger or equal to the sceptical outcome labelling: $\mathcal{L}_{so} \sqsubseteq \mathcal{L}_{co}$.
2. Suppose there is a meeting and suppose that Martin has a more cautious position than Gabriella, i.e. Martin's position is less committed than Gabriella's:
   - If the meeting applies the sceptical aggregation procedure, then Gabriella might as well stay at home.
   - If the meeting applies the credulous aggregation procedure, then Martin might as well stay at home.

Introducing preferences (2)

- $\mathcal{L} \geq_i \mathcal{L}'$ denotes that agent i *prefers* labelling $\mathcal{L}$ to $\mathcal{L}'$.
- Each agent submits his most preferred labelling.
- The order over the other possible labellings is generated according to the distance from the most preferred one.

Definition (Hamming set $\ominus$)

Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two labellings of argumentation framework (Ar, def). We define the Hamming set between these labellings as $\mathcal{L}_1 \ominus \mathcal{L}_2 = \{A \mid \mathcal{L}_1(A) \neq \mathcal{L}_2(A)\}$.

Definition (Hamming distance $|\ominus|$)

We define the Hamming distance between these labellings as $\mathcal{L}_1 \,|\ominus|\, \mathcal{L}_2 = |\mathcal{L}_1 \ominus \mathcal{L}_2|$.

Introducing preferences (3)

Definition (Hamming set based preference)

Agent i's preference is Hamming set based (written as $\geq_{i,\ominus}$) iff $\forall \mathcal{L}, \mathcal{L}' \in \mathcal{L}abellings, \mathcal{L} \geq_i \mathcal{L}' \Leftrightarrow \mathcal{L} \ominus \mathcal{L}_i \subseteq \mathcal{L}' \ominus \mathcal{L}_i$ where $\mathcal{L}_i$ is the agent's most preferred labelling.

Definition (Hamming distance based pref.)

Agent i's preference is Hamming distance based (written as $\geq_{i,|\ominus|}$) iff $\forall \mathcal{L}, \mathcal{L}' \in \mathcal{L}abellings, \mathcal{L} \geq_i \mathcal{L}' \Leftrightarrow \mathcal{L} \,|\ominus|\, \mathcal{L}_i \leq \mathcal{L}' \,|\ominus|\, \mathcal{L}_i$ where $\mathcal{L}_i$ is the agent's most preferred labelling.