Group Manipulation in Judgment Aggregation

Sirin Botan   Arianna Novaro   Ulle Endriss

ILLC – UNIVERSITY OF AMSTERDAM

Judgment aggregation is a formal framework for integrating the views of several agents into a single collective view. This is the first study of strategic behaviour by groups of agents in judgment aggregation. We introduce the concept of **group manipulation** – where a coalition of agents can cooperate to manipulate together – and characterise the family of **aggregation rules** for which group manipulation can be avoided.

Judgment Aggregation

Agenda $\quad \Phi := \Phi^+ \cup \{\neg\varphi \mid \varphi \in \Phi^+\}$

- finite set of formulas of propositional logic
- only non-negated formulas in the *pre-agenda* Φ^+
- **atomic** if Φ^+ only contains atomic propositions

Judgment Set for Φ $\quad J \subseteq \Phi$

- *complete* if $\varphi \in J$ or $\neg\varphi \in J$ for all $\varphi \in \Phi^+$
- *consistent* if it is logically consistent
- $\mathcal{J}(\Phi)$ is the set of complete & consistent judgment sets over Φ

Agents and Profiles

- $\mathcal{N} = \{1, \dots, n\}$ is a finite set of **agents**
- $\boldsymbol{J} = (J_1, \dots, J_n)$ is a **profile**, vector of *individual* judgment sets
- $N_\varphi^{\boldsymbol{J}} = \{i \in \mathcal{N} \mid \varphi \in J_i\}$ is the *coalition of supporters* of φ in $\boldsymbol{J}$
- $(\boldsymbol{J}_{-i}, J_i')$ is a profile like $\boldsymbol{J}$, except that J_i' replaced J_i
- $\boldsymbol{J}$ and $\boldsymbol{J'}$ are **C-variants**, for $C \subseteq \mathcal{N}$, if $J_i = J_i'$ for all $i \in \mathcal{N} \setminus C$

Flipping

- $J^{\rightleftarrows\varphi}$ means replacing φ by $\neg\varphi$ or $\neg\varphi$ by φ
- $\boldsymbol{J}^{\rightleftarrows S}$ means flipping formulas in S in all judgment sets in $\boldsymbol{J}$

Aggregation Rules $\quad F : \mathcal{J}(\Phi)^n \to 2^\Phi$

- **uniform quota rules** $F_q(\boldsymbol{J}) = \{\varphi \in \Phi \mid \#N_\varphi^{\boldsymbol{J}} \geqslant q\}$ for quota q
  - **nomination rule** if $q = 1$
  - **weak majority rule** if $q = \lceil \frac{n}{2} \rceil$
  - **unanimity rule** if $q = n$

Axioms for Aggregation Rule F

- **independence** $\quad N_\varphi^{\boldsymbol{J}} = N_\varphi^{\boldsymbol{J'}}$ implies $\varphi \in F(\boldsymbol{J}) \Leftrightarrow \varphi \in F(\boldsymbol{J'})$
- **monotonicity** $\quad \varphi \in J_i' \setminus J_i$ implies $\varphi \in F(\boldsymbol{J}) \Rightarrow \varphi \in F(\boldsymbol{J}_{-i}, J_i')$
- **neutrality** $\quad N_\varphi^{\boldsymbol{J}} = N_\psi^{\boldsymbol{J}}$ implies $\varphi \in F(\boldsymbol{J}) \Leftrightarrow \psi \in F(\boldsymbol{J})$
- **unbiasedness** $\quad F(\boldsymbol{J}^{\rightleftarrows S}) = F(\boldsymbol{J})^{\rightleftarrows S}$ for any $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$ and $S \subseteq \Phi^+$ where $\boldsymbol{J}^{\rightleftarrows S} \in \mathcal{J}(\Phi)^n$

Preferences

- J_i is the most preferred judgment set of agent i
- preference ranking in terms of distance to J_i
- **Hamming Distance** $\quad H(J, J') = |J \setminus J'| + |J' \setminus J|$
- *weak order* on judgment sets $\quad J \succcurlyeq_i^{\boldsymbol{J}} J' \Leftrightarrow H(J, J_i) \leqslant H(J', J_i)$

Example. *If agent 3 only cares about the conclusion $(p \wedge q)$ she can manipulate the outcome in her favour by rejecting q.*

	p	q	$p \wedge q$
Agent 1	✓	✓	✓
Agent 2	✓	×	×
Agent 3	×	✓	×
PB-Rule	✓	✓	✓

Single-Agent Strategyproofness

A rule is strategyproof if no agent has an incentive to manipulate by reporting an untruthful opinion.

Definition 1. *A rule F is* **strategyproof**, *if for all profiles $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, agents $i \in \mathcal{N}$, and judgment sets $J_i' \in \mathcal{J}(\Phi)$ it is the case that $F(\boldsymbol{J}) \succcurlyeq_i^{\boldsymbol{J}} F(\boldsymbol{J}_{-i}, J_i')$.*

Some rules, e.g. uniform quota rules, are strategyproof.

Theorem 1. *A neutral and unbiased aggregation rule F is single-agent strategyproof* **iff** *it is both independent and monotonic.*

Group Strategyproofness

A rule is group-strategyproof if no coalition of manipulators has an incentive to report untruthful judgments.

Definition 2. *A rule F is* **group-strategyproof** *against coalitions of up to k manipulators, if for all profiles $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, coalitions $C \subseteq \mathcal{N}$ with $|C| \leqslant k$, and C-variants $\boldsymbol{J'} \in \mathcal{J}(\Phi)^n$ of $\boldsymbol{J}$ it is the case that $F(\boldsymbol{J}) \succcurlyeq_i^{\boldsymbol{J}} F(\boldsymbol{J'})$ for all agents $i \in C$.*

Example. *If the first three agents form a coalition, they will benefit from flipping their judgments on the indicated formulas.*

	φ_1	φ_2	φ_3	$\neg\varphi_1$	$\neg\varphi_2$	$\neg\varphi_3$
Agent 1	×	✓	✓	✓	×	×
Agent 2	✓	×	✓	×	✓	×
Agent 3	✓	✓	×	×	×	✓
Agent 4	×	×	×	✓	✓	✓
Agent 5	×	×	×	✓	✓	✓
Majority	×	×	×	✓	✓	✓

Almost no rule is group-strategyproof.

Theorem 2. *Suppose the agenda Φ is atomic. Then a neutral and unbiased aggregation rule F is group-strategyproof against coalitions of up to 3 manipulators* **iff** *F is independent and monotonic, and if none of the restrictions of F to 3 agents and 3 pre-agenda formulas is either the nomination rule or the unanimity rule.*

Uniform quota rules are not group-strategyproof.

Corollary 3. *No uniform quota rule F_q with a quota q satisfying $3 \leqslant q \leqslant n$ or $1 \leqslant q \leqslant n-2$ that is defined on an atomic agenda Φ is group-strategyproof.*

Strategyproofness for Fragile Coalitions

A manipulator may decide to unilaterally opt-out of a manipulation.

Definition 3. *A rule F is* **group-strategyproof against fragile coalitions** *of up to k manipulators, if for all profiles $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, coalitions $C \subseteq \mathcal{N}$ with $|C| \leqslant k$, and C-variants $\boldsymbol{J'} \in \mathcal{J}(\Phi)^n$ of $\boldsymbol{J}$ with $F(\boldsymbol{J'}) \succ_i^{\boldsymbol{J}} F(\boldsymbol{J})$ and $F(\boldsymbol{J'}_{-i}, J_i) \neq F(\boldsymbol{J'})$ for all $i \in C$ it is the case that $F(\boldsymbol{J'}_{-i}, J_i) \succ_i^{\boldsymbol{J}} F(\boldsymbol{J'})$ for some $i \in C$.*

If agents can opt-out, strategyproof rules are group-strategyproof.

Theorem 4. *A neutral and unbiased aggregation rule F is group-strategyproof against fragile coalitions of manipulators* **iff** *it is independent and monotonic.*