Group Manipulation
in Judgment Aggregation

Sirin Botan, Arianna Novaro and Ulle Endriss
ILLC - Universiteit van Amsterdam

November 20, 2015

Motivating Example

Judgment Aggregation: Combine agents' opinions about some issues into a collective decision on them.

	p	q	**p ∧ q**
Agent 1	✓	✓	✓
Agent 2	✓	×	×
Agent 3	×	✓	×
PB-Rule	✓	✓	✓

We will talk about:

$\Rightarrow$ Different type of Rules

$\Rightarrow$ More general type of Preferences

Outline of the Talk

1. JA Framework & Quota Rules
2. Single-agent manipulation
3. Group manipulation
4. Conclusions

Notation and Formal Framework

$\mathcal{N} = \{1, \ldots, n\}$ is the set of **agents**.

Φ is the **agenda** (finite non-empty set of propositional formulas and their negations).

$J_i \subseteq \Phi$ is the **individual judgment set** for agent *i*.

$\boldsymbol{J} = (J_1, \ldots, J_n)$ is the **profile** on agenda Φ.

$\mathcal{J}(\Phi)$ is the set of all *complete* & *consistent* subsets of Φ.

An **aggregation rule** for an agenda Φ and a set of *n* agents is a function from profiles to (collective) judgment sets:

$$F : \mathcal{J}(\Phi)^n \to 2^{\Phi}.$$

Uniform Quota Rules

A **uniform quota rule** is defined by $q \in \{0, 1, \dots, n+1\}$:

$$F_q(\boldsymbol{J}) = \{\varphi \in \Phi \mid \#\{i \in \mathcal{N} \mid \varphi \in J_i\} \geq q\}.$$

	r	s	t	¬r	¬s	¬t
J_1	×	✓	✓	✓	×	×
J_2	✓	×	✓	×	✓	×
J_3	✓	✓	×	×	×	✓
J_4	×	×	×	✓	✓	✓
J_5	×	×	×	✓	✓	✓
$F_3(\boldsymbol{J})$	×	×	×	✓	✓	✓

In this example, F_3 is the Majority Rule.

Individual Preferences

The *Hamming Distance* is defined as

$$H(J, J') := |J \setminus J'| + |J' \setminus J|.$$

The Hamming Preferences of agent i are such that

$$J \succeq_i J' \;\Leftrightarrow\; H(J, J_i) \leq H(J', J_i).$$

We will assume Hamming Preferences for our theorems.

Single-Agent Strategy-Proofness

Agent i manipulates whenever she does not report her *truthful* judgment set J_i.

Agent i has an incentive to manipulate if for some $J_i' \in \mathcal{J}(\Phi)$:

$$F(\boldsymbol{J}_{-i}, J_i') \succ_i F(\boldsymbol{J}).$$

A rule F is single-agent strategy-proof, if for no truthful profile $\boldsymbol{J}$ there is an agent with an incentive to manipulate.

Theorem. Quota Rules are single-agent strategy-proof.

Dietrich & List. Strategy-Proof Judgment Aggregation. *Economics & Philosophy*, 2007.

Group Strategy-Proofness

A coalition C of agents is a subset of $\mathcal{N}$.

$\boldsymbol{J'}$ is a C-variant of $\boldsymbol{J}$ if $J_i = J'_i$ for all agents i not in C.

F is group strategy-proof against coalitions of size $\leq k$, if for all truthful profiles $\boldsymbol{J}$, for all coalitions C of size $\leq k$, and for all C-variants $\boldsymbol{J'}$ of $\boldsymbol{J}$ we have $F(\boldsymbol{J}) \succeq_i F(\boldsymbol{J'})$ for all agents $i \in C$.

Manipulation by Two Agents

Theorem. Uniform Quota Rules are strategy-proof against coalitions of manipulators of at most 2 agents.

Proof. We can distinguish two cases:
1 agent Follows from previous theorem. ✓
2 agents Formulas on which the agents *agree*: already both rejecting or both accepting them.
$\Rightarrow$ Changes useless or counterproductive.
Formulas on which the agents *disagree*: if agent 1 changes her opinion on some φs, she goes against her interest to possibly help agent 2 (by changing the outcome).
$\Rightarrow$ Agent 1 needs "in return" strictly more formulas from agent 2 (Hamming Distance preferences).
$\Rightarrow$ The reasoning is symmetric for both agents. ✓

Manipulation by Three Agents (or more)

Theorem. If the (atomic) agenda Φ includes at least 3 (non-negated) formulas, then every Uniform Quota Rule F_q such that $3 \leq q \leq n$ (or $1 \leq q \leq n - 2$) is not group strategy-proof against coalitions of size ≤ 3.

Proof. We show, for any such $3 \leq q \leq n$ (other case similar), a general method for constructing a profile manipulable by three agents. By checking the Hamming Distances we see that they have an incentive to manipulate *together*.

Proof

Consider the truthful profile $\boldsymbol{J}$ . . .

	φ_1	φ_2	φ_3	...	$\neg\varphi_1$	$\neg\varphi_2$	$\neg\varphi_3$	...
J_1	$\times$	$\checkmark$	$\checkmark$	...	$\checkmark$	$\times$	$\times$	...
J_2	$\checkmark$	$\times$	$\checkmark$	...	$\times$	$\checkmark$	$\times$	...
J_3	$\checkmark$	$\checkmark$	$\times$	...	$\times$	$\times$	$\checkmark$	...
J_4	$\checkmark$	$\checkmark$	$\checkmark$	...	$\times$	$\times$	$\times$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	...
J_q	$\checkmark$	$\checkmark$	$\checkmark$	...	$\times$	$\times$	$\times$	...
J_{q+1}	$\times$	$\times$	$\times$	...	$\checkmark$	$\checkmark$	$\checkmark$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	...
J_n	$\times$	$\times$	$\times$	...	$\checkmark$	$\checkmark$	$\checkmark$	...
$\mathrm{F_q}(\boldsymbol{J})$	$\times$	$\times$	$\times$	...	?	?	?	...

Proof

... and the manipulated profile $\boldsymbol{J'}$.

	φ_1	φ_2	φ_3	...	$\neg\varphi_1$	$\neg\varphi_2$	$\neg\varphi_3$	...
J'_1	✓	✓	✓	...	×	×	×	...
J'_2	✓	✓	✓	...	×	×	×	...
J'_3	✓	✓	✓	...	×	×	×	...
J_4	✓	✓	✓	...	×	×	×	...
⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	...
J_q	✓	✓	✓	...	×	×	×	...
J_{q+1}	×	×	×	...	✓	✓	✓	...
⋮	⋮	⋮	⋮	...	⋮	⋮	⋮	...
J_n	×	×	×	...	✓	✓	✓	...
$\mathrm{F_q}(\boldsymbol{J'})$	✓	✓	✓	...	?	?	?	...

Strategy-Proofness with Opting Out

⇒ What happens if agents in our construction are allowed to opt out of the jointly agreed plan?
⇒ What happens if agents are risk-averse (to the possibility of the rest of the coalition opting out)?

	φ_1	φ_2	φ_3	$\neg\varphi_1$	$\neg\varphi_2$	$\neg\varphi_3$
J_1	×	✓	✓	✓	×	×
J_2	✓	✓	✓	×	×	×
J_3	✓	✓	✓	×	×	×
J_4	×	×	×	✓	✓	✓
J_5	×	×	×	✓	✓	✓
$F_3(\boldsymbol{J})$	×	✓	✓	✓	×	×

Theorem. If agents are risk-averse and may opt out, then Uniform Quota Rules are group strategy-proof.

Conclusion & Future Work

We introduced the notion of group manipulation in JA.

For Uniform Quota Rules we get the following results:

- ✓ Strategy-proof against single agent (D. & L., 2007).
- ✓ Strategy-proof against two manipulators.
- × Manipulable by three (or more) agents.
- ✓ Strategy-proof against unstable groups.

Similar results for more general rules (Independent and Monotonic).