Judgment Aggregation June Project: Safety of the Agenda

Zoi Terzopoulou

Institute for Logic, Language, and Computation
University of Amsterdam

6/6/2018

(based on the slides of Ulle Endriss)

Goals

Consider a given aggregation rule F (such as the majority rule).

- Which agendas Φ guarantee *consistent outcomes* for F?
- In other words... which agendas are *safe*?

The *discursive dilemma* shows that not all agendas are safe. We:

- characterise safe agendas for the majority rule
- and safe agendas for all rules meeting certain axioms

This is the problem of the *safety of the agenda*.

From now on... be ready for the stars ⭐ !

Remember the Impossibility Theorem...

Theorem (List and Pettit, 2002)

No judgment aggregation rule for an agenda Φ with $\{p, q, p \wedge q\} \subseteq \Phi$ satisfies all of the axioms of anonymity, neutrality, independence, completeness, and consistency.

But for which agendas exactly is this the case? To find out, we need a *characterisation of the class of agendas* for which satisfying A, N, I, together with consistency is (im)possible.

C. List and P. Pettit. Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy*, 18(1):89–110, 2002.

Agenda Properties

Two useful properties of agendas Φ (i.e., of sets of formulas):

- *Median Property (MP)*: Φ has the MP iff every minimally inconsistent (mi) subset of Φ has size $\leqslant 2$.
- *Simplified Median Property (SMP)*: iff every non-singleton mi subset of Φ is of the form $\{\varphi, \psi\}$ with $\Vdash \varphi \leftrightarrow \neg\psi$.

Obviously, if Φ has the SMP, then it also has the MP.

If Φ has the MP, it does not necessarily have the SMP.

Example for the latter: $\{p, \neg p, p \wedge q, \neg(p \wedge q)\}$

Consistent Aggregation for the Majority

We now see a *safety result* for a *specific aggregation rule*.

Theorem (Nehring and Puppe, 2007)

Let the number of agents $n \geqslant 3$. The (strict) majority rule is consistent for a given agenda Φ iff Φ has the MP.

Note how the agenda in L&P's impossibility violates the MP:

$$\{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$$

(However, deciding if an agenda has the MP is Π_2^P-*complete*!)

K. Nehring and C. Puppe. The Structure of Strategy-proof Social Choice. Part I: General Characterization and Possibility Results on Median Space. *Journal of Economic Theory*, 135(1):269–305, 2007.

Proof

($\Leftarrow$) Take Φ with the MP. If there exists an admissible profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$ such that $F_{maj}(\boldsymbol{J})$ is *not consistent*, then:

- There exists an inconsistent set $\{\varphi, \psi\} \subseteq F_{maj}(J)$.
- So both φ, ψ must have been accepted by a strict majority.
- Thus one individual must have accepted both φ and ψ.
- *Contradiction* (individual judgment sets are consistent). ✓

($\Rightarrow$) Take Φ that violates the MP. Then, there exists a minimally inconsistent set $\Delta = \{\varphi_1, \ldots, \varphi_k\} \subseteq \Phi$ with $k > 2$. Consider the profile $\boldsymbol{J}$, in which individual i accepts all formulas in Δ except for $\varphi_{1+(i \mod 3)}$. Note that $\boldsymbol{J}$ is consistent. But the majority rule will accept all formulas in Δ, i.e., $F_{maj}(\boldsymbol{J})$ is *not consistent*. ✓

Agenda Safety for Classes of Rules

Instead of a single rule, suppose we are interested in a *class of rules*, for instance those corresponding to a set of *axioms*.

- *Existential Agenda Characterisation*: Some rule meeting the axioms is consistent for every appropriate agenda.
- *Universal Agenda Characterisation*: All rules meeting the axioms are consistent for every appropriate agenda.

Any real-life scenarios where these are important questions? ⭐

Now we will look into the latter, also called *safety results*.

Why the "majority safety theorem" above is a safety result? ⭐

Example for a Safety Theorem

For the remainder of today, assume that Φ contains no tautologies, and thus no contradictions (simplifies the result).

A theorem for the majority-rule axioms, minus monotonicity:

Theorem (Endriss et al., 2012)

Φ is safe for all anonymous, neutral, independent, complete and complement-free aggregation rules iff Φ has the SMP.

Recall that SMP = all inconsistencies due to some $\{\varphi, \psi\}$ with $\Vdash \varphi \leftrightarrow \neg\psi$. We now give a proof for the case where n is *odd*.

U. Endriss, U. Grandi, and D. Porello. Complexity of Judgment Aggregation. *Journal of Artificial Intelligence Research (JAIR)*, 45:481–514, 2012.

Proof

($\Leftarrow$) Take Φ with the SMP and assume some $F(\boldsymbol{J})$ is inconsistent. Then: $\{\varphi, \psi\} \subseteq F(\boldsymbol{J})$ with $\Vdash \varphi \leftrightarrow \neg\psi$. Then:

- $\varphi \in J_i \Leftrightarrow {\sim}\psi \in J_i$ for each individual i. (why? ⭐)
- So $\varphi \in F(\boldsymbol{J}) \Leftrightarrow {\sim}\psi \in F(\boldsymbol{J})$ (from neutrality).
- Thus both ψ and ${\sim}\psi$ in $F(\boldsymbol{J})$ (*contradiction* with complement-freeness). ✓

($\Rightarrow$) Take Φ that violates the SMP, and some mi $X \subseteq \Phi$. If $|X| > 2$, then also the MP is violated and we already know that the majority rule is not consistent. ✓

So we can assume $X = \{\varphi, \psi\}$. W.l.o.g., we must have that $\varphi \Vdash \neg\psi$, but $\neg\psi \nVdash \varphi$ (otherwise SMP holds).

But now there is a non-safe rule: the parity rule F_{par} accepts a formula iff an odd number of agents does. Consider a profile $\boldsymbol{J}$ with $J_1 \supseteq \{{\sim}\varphi, {\sim}\psi\}, J_2 \supseteq \{{\sim}\varphi, \psi\}, J_3 \supseteq \{\varphi, {\sim}\psi\}$. Then, $F_{par}(\boldsymbol{J}) \supseteq \{\varphi, \psi\}$. ✓

Summary of Part A

We discussed the problem of the *safety of the agenda*.

Are we convinced this is an actual problem? ⭐ We know that:

- The majority rule always produces consistent outcomes iff every possible inconsistency in the agenda can be explained in terms of just two formulas (*median property*).
- For the consistency of all rules that share the properties of the majority except for monotonicity, we simplify agendas even further and only permit inconsistencies arising from logical complements (*simplified median property*).
- Similar results hold for other combinations for axioms.