Plan for Today

Yesterday has been an introduction to judgment aggregation, covering motivating paradoxes, the formal framework, concrete aggregators, and a basic impossibility result.

Today we will discuss more advanced topics in JA:

- **Agenda characterisation**: types of agendas on which paradoxical outcomes can be avoided. This includes:
  - **Possibility**: existence of acceptable rules on certain agendas
  - **Safety**: guaranteed consistency of outcomes for all relevant rules on certain agendas (also: complexity of deciding safety)
- **Strategic behaviour** (briefly)
- **Applications** (very briefly)

Reminder: Formal Framework

Notation: Let $\sim \varphi := \varphi'$ if $\varphi = \neg \varphi'$ and let $\sim \varphi := \neg \varphi$ otherwise.

An agenda $\Phi$ is a finite nonempty set of propositional formulas (w/o double negation) closed under complementation: $\varphi \in \Phi \Rightarrow \sim \varphi \in \Phi$.

A judgment set $J$ on an agenda $\Phi$ is a subset of $\Phi$. We call $J$:

- **complete** if $\varphi \in J$ or $\sim \varphi \in J$ for all $\varphi \in \Phi$
- **complement-free** if $\varphi \notin J$ or $\sim \varphi \notin J$ for all $\varphi \in \Phi$
- **consistent** if there exists an assignment satisfying all $\varphi \in J$

Let $\mathcal{J}(\Phi)$ be the set of all complete and consistent subsets of $\Phi$.

Now a finite set of individuals $N = \{1, \ldots, n\}$, with $n \geq 2$, express judgments on the formulas in $\Phi$, producing a profile $J = (J_1, \ldots, J_n)$.

An aggregation procedure for an agenda $\Phi$ and a set of $n$ individuals is a function mapping a profile of complete and consistent individual judgment sets to a single collective judgment set: $F : \mathcal{J}(\Phi)^n \rightarrow 2^\Phi$.

Agenda Characterisations

Recall yesterday’s impossibility theorem: no consistent aggregator is independent, neutral, and anonymous for agendas $\Phi \supseteq \{p, q, p \land q\}$.

More interesting question:

- For which class of agendas is consistent aggregation (im)possible?

We will give several answers to this generic question . . .

Remark: Note that the characterisation results we have seen yesterday (e.g., axiomatisation of the majority rule) are rather different. They don’t involve consistency (i.e., they don’t involve any logic).
Consistent Aggregation under the Majority Rule

Yesterday we saw that the majority rule can produce an inconsistent outcome for some (not all) profiles based on agendas $\Phi \supseteq \{p, q, p \land q\}$. How can we characterise the class of agendas with this problem?

A set of formulas $\Phi$ satisfies the median property if every inconsistent subset of $\Phi$ does itself have an inconsistent subset of size $\leq 2$.

Lemma 1 (Nehring and Puppe, 2007) Let $n \geq 3$. The majority rule is consistent for a given agenda $\Phi$ iff $\Phi$ has the median property.

Remark: Note how $\{p, \neg p, q, \neg q, p \land q, \neg (p \land q)\}$ violates the MP.

Proof

Claim: $\Phi$ is safe [$F_{\text{maj}}(J)$ is consistent] $\iff$ $\Phi$ has the median property

($\Leftarrow$) Let $\Phi$ be an agenda with the median property. Now assume that there exists an admissible profile $J$ such that $F_{\text{maj}}(J)$ is not consistent.

$\neg \Rightarrow$ There exists an inconsistent set $\{\varphi, \psi\} \subseteq F_{\text{maj}}(J)$.

$\neg \Rightarrow$ Each of $\varphi$ and $\psi$ must have been accepted by a strict majority.

$\neg \Rightarrow$ One individual must have accepted both $\varphi$ and $\psi$.

$\neg \Rightarrow$ Contradiction (individual judgment sets must be consistent).

($\Rightarrow$) Let $\Phi$ be an agenda that violates the median property, i.e., there exists a minimally inconsistent set $X = \{\varphi_1, \ldots, \varphi_k\} \subseteq \Phi$ with $k > 2$.

Consider the profile $J$, in which individual $i$ accepts all formulas in $X$ except for $\varphi_{1+(i \mod 3)}$. Note that $J$ is consistent. But the majority rule will accept all formulas in $X$, i.e., $F_{\text{maj}}(J)$ is inconsistent.
Preparation: Winning Coalitions

First, a helpful reinterpretation of the axioms:

\( F \) is independent iff for every \( \varphi \in \Phi \) there’s a set of winning coalitions \( W_\varphi \subseteq 2^N \) such that \( \varphi \in F(J) \iff N^J_\varphi \in W_\varphi \) for all \( J \in J(\Phi)_n \).

If \( F \) is furthermore neutral, then it is determined by a single \( W \subseteq 2^N \):

\[ \varphi \in F(J) \iff N^J_\varphi \in W \text{ for all } J \in J(\Phi)_n. \]

Aside: What does anonymity correspond to? And unanimity?

Remark: Winning coalitions correspond to what we had called weakly decisive coalitions in preference aggregation.

Proof Plan: Possibility Theorem

Note that the impossibility direction of our theorem is equivalent to:

Claim: If a neutral, independent, and monotonic aggregator \( F \) is complete and consistent for an agenda \( \Phi \) violating the median property, then \( F \) must be a dictatorship.

So suppose \( \Phi \) violates the MP and \( F \) has the properties on the left.

By independence and neutrality, there exists a (single) family of winning coalitions \( W \subseteq 2^N \) determining \( F \): \( \varphi \in F(J) \iff N^J_\varphi \in W \).

We will show that \( W \) is an ultrafilter on \( N \), which means:

(i) The empty coalition is not winning: \( \emptyset \notin W \).

(ii) Closure under intersection: \( C, C' \in W \implies C \cap C' \in W \).

(iii) Maximality: \( C \in W \) or \( \overline{C} := N \setminus C \in W \).

Appealing to the finiteness of \( N \), this will allow us to show that \( W = \{ C \subseteq N \mid i^\star \in C \} \) for some \( i^\star \in N \), i.e., that \( F \) is dictatorial.

Proof: Noninclusion of the Empty Set

Claim: \( \emptyset \notin W \).

We will use monotonicity and complement-freeness:

For the sake of contradiction, assume \( \emptyset \in W \).

From monotonicity (i.e., closure under supersets): \( N \in W \) as \( \emptyset \subseteq N \).

But now consider some profile \( J \) with \( p \in J_i \) for all individuals \( i \in N \).

\( \sim \) we get \( N^J_p = N \) and \( N^J_{\neg p} = \emptyset \).

\( \sim \) that is, \( p \in F(J) \) and \( \neg p \in F(J) \), as both \( N \in W \) and \( \emptyset \in W \).

\( \sim \) contradiction with complement-freeness \( \checkmark \).

Proof: Maximality

Claim: \( C \in W \) or \( \overline{C} := N \setminus C \in W \) for all \( C \subseteq N \).

We will use the fact that \( F \) is supposed to be complete:

- take any coalition \( C \subseteq N \) and any formula \( \varphi \in \Phi \)
- construct a profile \( J \) with \( N^J_\varphi = C \)
- from completeness: \( \varphi \in F(J) \) or \( \sim \varphi \in F(J) \)
- from \( W \)-determination of \( F \): \( N^J_\varphi \in W \) or \( N^J_{\sim \varphi} \in W \)
- from \( J \) being complete and complement-free: \( N^J_{\sim \varphi} = \overline{N^J_\varphi} \)
- putting everything together: \( C \in W \) or \( \overline{C} \in W \).
Proof: Closure under Taking Intersections

Claim: \( C, C' \in \mathcal{W} \Rightarrow C \cap C' \in \mathcal{W} \) for all \( C, C' \subseteq \mathcal{N} \).

We'll use MP-violation, monotonicity, consistency, and completeness.

MP-violation means: there's a \textit{mi-subset} \( X = \{ \varphi_1, \ldots, \varphi_k \} \subseteq \Phi \) with \( k \geq 3 \).

We can construct a complete and consistent profile \( J \) with these properties:

\begin{itemize}
  \item \( N^J_{\varphi_1} = C \)
  \item \( N^J_{\varphi_2} = C' \cup (\mathcal{N} \setminus C) \)
  \item \( N^J_{\varphi_3} = \mathcal{N} \setminus (C \cap C') \)
  \item \( N^J_{\psi} = \mathcal{N} \) for all \( \psi \in X \setminus \{ \varphi_1, \varphi_2, \varphi_3 \} \)
\end{itemize}

Thus: everyone accepts \( k-1 \) of the propositions in \( X \). And \( N^J_{\varphi_3} = C \cap C' \).

\begin{itemize}
  \item \( C \in \mathcal{W} \Rightarrow \varphi_1 \in F(J) \)
  \item From monotonicity: \( C' \in \mathcal{W} \Rightarrow (C' \cup (\mathcal{N} \setminus C)) \in \mathcal{W} \Rightarrow \varphi_2 \in F(J) \)
  \item From maximality: \( \emptyset \notin \mathcal{W} \Rightarrow \varphi_3 \notin F(J) \)
\end{itemize}

Thus: for consistency we need \( \varphi_3 \notin F(J) \), i.e., for completeness \( \sim \varphi_3 \in F(J) \).

In other words: \( N^J_{\varphi_3} = C \cap C' \in \mathcal{W} \).

Proof: Dictatorship

We have shown that the family of winning coalitions \( \mathcal{W} \) is an \textit{ultrafilter} on the (\textit{finite!}) set of individuals \( \mathcal{N} \):

(i) The empty coalition is not winning: \( \emptyset \notin \mathcal{W} \)

(ii) Closure under intersection: \( C, C' \in \mathcal{W} \Rightarrow C \cap C' \in \mathcal{W} \)

(iii) Maximality: \( C \in \mathcal{W} \) or \( C : = \mathcal{N} \setminus C \in \mathcal{W} \)

From (i) and completeness: \( \mathcal{N} \notin \mathcal{W} \) (btw: this is unanimity).

\textbf{Contraction Lemma:} if \( C \in \mathcal{W} \) and \( |C| \geq 2 \), then \( C' \in \mathcal{W} \) for some \( C' \subset C \).

Proof: Let \( C_1 \cup C_2 = C \). If \( C_1 \notin \mathcal{W} \), then \( C_1 \notin \mathcal{W} \) by maximality.

But then \( C \cap C_1 = C_2 \notin \mathcal{W} \) by closure under intersection. \( \checkmark \)

By induction: \( \{ i^* \} \in \mathcal{W} \) for one \( i^* \in \mathcal{N} \), i.e., \( \mathcal{W} = \{ C \subseteq \mathcal{N} | i^* \in C \} \).

That is, \( i^* \) is a \textit{dictator}. \( \checkmark \)

Remark: The above just spells out the well-known fact that every ultrafilter on a finite set must be \textit{principal}, i.e., of the form \( \mathcal{W} = \{ C \subseteq \mathcal{N} | i^* \in C \} \).

Example for a Safety Theorem

Suppose we know that the group will use some aggregation procedure meeting certain requirements, but we do not know which procedure exactly. Can we guarantee that the outcome will be consistent?

A typical result (for the majority rule axioms, minus monotonicity):

\textbf{Theorem 3 (Endriss et al., 2012)} An agenda \( \Phi \) is safe for any anonymous, neutral, independent, complete and complement-free aggregation procedure if \( \Phi \) has the \textit{simplified median property}.

An agenda \( \Phi \) has the simplified median property if every inconsistent subset of \( \Phi \) has itself an inconsistent subset \( \{ \varphi, \psi \} \) with \( \varphi \leftrightarrow \neg \psi \).

\textbf{Note:} The SMP is more restrictive than the MP (see: \( \{ \neg p, p \land q \} \)).
**Proof**

**Claim:** \( \Phi \) is safe for any ANI/complete/comp-free rule \( F \iff \Phi \) has SMP

\((\iff) \) Suppose \( \Phi \) has the SMP. For the sake of contradiction, assume \( F(J) \) is inconsistent. Then \( \{\varphi, \psi\} \subseteq F(J) \) with \( \models \varphi \iff \neg \psi \). Now:

\[ \neg \varphi \in J_i \iff \neg \psi \in J_i \] for each individual \( i \) (from \( \models \varphi \iff \neg \psi \) together with consistency and completeness of individual judgment sets).

\[ \neg \varphi \in F(J) \iff \neg \psi \in F(J) \] (from neutrality)

\[ \text{both } \psi \text{ and } \neg \psi \text{ in } F(J) \iff \text{contradiction (with complement-freeness)} \]

\((\Rightarrow) \) Suppose \( \Phi \) violates the SMP. Take any minimally inconsistent \( X \subseteq \Phi \).

If \( |X| > 2 \), then also the MP is violated and we already know that the majority rule is not consistent. \( \checkmark \) So we can assume \( X = \{\varphi, \psi\} \).

W.l.o.g., must have \( \varphi \models \neg \psi \) but \( \neg \psi \not\models \varphi \) (otherwise SMP holds).

But now we can find a rule \( F \) that is not safe: accept a formula if at most one individual does and take a profile with \( J_1 = \{\neg \varphi, \neg \psi, \ldots\} \), \( J_2 = \{\varphi, \psi, \ldots\} \), and \( J_3 = \{\varphi, \neg \psi, \ldots\} \). Then \( F(J) = \{\varphi, \psi, \ldots\} \). \( \checkmark \)

---

**Comparing Possibility and Safety Results**

Possibility theorems and safety theorems are closely related:

- **Possibility:** some aggregator in the class determined by the given axioms will produce consistent outcomes if/iff the agenda has a given property
- **Safety:** all aggregators in the class determined by the given axioms will produce consistent outcomes if/iff the agenda has a given property

In what situations do we need these results?

- **Possibility:** a mechanism designer wants to know whether she can design an aggregation rule meeting a given list of requirements
- **Safety:** a system might know certain properties of the aggregator users will employ (but not all properties) and we want to be sure there won’t be any problem (we might want to check this again and again)

For safety problems in particular we might want to develop **algorithms**, i.e., **complexity** plays a role.

---

**Example: Strategic Manipulation**

Suppose we use the **premise-based procedure**:

<table>
<thead>
<tr>
<th>Agent 1</th>
<th>Agent 2</th>
<th>Agent 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>q</td>
<td>p \lor q</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PBP:</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

If agent 3 only cares about the conclusion \( p \lor q \), she could **manipulate** the aggregation by claiming to believe that \( p \) is true.
Strategic Behaviour

Note that in *pure* JA, we cannot talk about strategic behaviour, as there is no notion of preference. We need to add one! *How?*

This is still underexplored territory. Main definition in use so far:

- Your true judgment set is your most preferred outcome.
- The closer an outcome to your true judgment set, in terms of the *Hamming distance*, the more you prefer that outcome.

Remarks:

- good news: manipulation for the PBP is *NP-hard*
- other forms of strategic behaviour: bribery and control


Applications

Some recent work has suggested possible directions for using judgment aggregation techniques in applications. *Examples:*

- Collective decision making in multiagent systems
- Ontology merging on the Semantic Web
- Aggregating crowdsourced data (e.g., for computational linguistics)


Summary

We have discussed the core themes in research on JA, where the views to be amalgamated are modelled as formulas of propositional logic:

- *specific aggregators*: quota rules, premise-based aggregation, (conclusion-based aggregation), distance-based aggregation
- *axioms*: independence, neutrality, anonymity, monotonicity, …
- *characterisation of rules* in terms of axioms (quota rules)
- agenda characterisation results:
  - possibility: agenda property ⇔ ∃ consistent rule in class*
  - safety: agenda property ⇔ ∀ rules in class are consistent
  - both: agenda has median property ⇔ majority rule consistent
  *one direction may be read as an impossibility theorem
- strategic behaviour: manipulation, bribery, control

What next?

The final topic of the course will be *fair division*.

This means returning to *preferences* as our objects of aggregation, but this time mostly cardinal preferences (*utilities*).

We will see *axiomatic* results, analyse concrete division *procedures*, and discuss *algorithmic* considerations.