Computational Social Choice: Autumn 2013

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

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* Φ 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 Φ is a subset of Φ. 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 Φ.

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

An *aggregation procedure* for an agenda Φ 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: $F : \mathcal{J}(\Phi)^n \to 2^\Phi$.

Agenda Characterisations

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

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 $\Phi \supseteq \{p, q, p \wedge q\}$. How can we *characterise* the class of agendas with this problem?

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

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

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

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

<u>Claim:</u> Φ is safe [$F_{\mathrm{maj}}(\boldsymbol{J})$ is consistent] $\Leftrightarrow$ Φ has the median property

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

- $\leadsto$ There exists an inconsistent set $\{\varphi, \psi\} \subseteq F_{\mathrm{maj}}(\boldsymbol{J})$.
- $\leadsto$ Each of φ and ψ must have been accepted by a strict majority.
- $\leadsto$ One individual must have accepted both φ and ψ.
- $\leadsto$ Contradiction (individual judgment sets must be consistent). ✓

($\Rightarrow$) Let Φ 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 $\boldsymbol{J}$, in which individual i accepts all formulas in X except for $\varphi_{1+(i \bmod 3)}$. Note that $\boldsymbol{J}$ is consistent. But the majority rule will accept all formulas in X, i.e., $F_{\mathrm{maj}}(\boldsymbol{J})$ is inconsistent. ✓

Agenda Characterisation for Classes of Rules

Now instead of a single aggregator, suppose we are interested in a *class of aggregators*, possibly determined by a set of *axioms*.

We might ask:

- *Possibility:* Does there exist an aggregator meeting certain axioms that will be consistent for any agenda with a given property?
- *Safety:* Will every aggregator meeting certain axioms be consistent for any agenda with a given property?

Discussion: In what situations are these relevant questions?

Example for a Possibility Theorem

Let again $n \geqslant 3$.

Theorem 1 (Nehring and Puppe, 2007) *There exists a neutral, independent, monotonic, and nondictatorial aggregator that is complete and consistent for agenda Φ iff Φ has the median property.*

Proof: One direction (right-to-left) follows from our lemma:

Suppose Φ has the median property.

$\rightsquigarrow$ the majority rule will be consistent and complete (by the lemma)

$\rightsquigarrow$ there exists an aggregator with all the required properties (namely the majority rule) ✓

Next we will prove the *impossibility direction* (left-to-right).

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.

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* $\mathcal{W}_\varphi \subseteq 2^{\mathcal{N}}$ such that $\varphi \in F(\boldsymbol{J}) \Leftrightarrow N_\varphi^{\boldsymbol{J}} \in \mathcal{W}_\varphi$ for all $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$.

If F is furthermore *neutral*, then it is determined by a *single* $\mathcal{W} \subseteq 2^{\mathcal{N}}$: $\varphi \in F(\boldsymbol{J}) \Leftrightarrow N_\varphi^{\boldsymbol{J}} \in \mathcal{W}$ for all $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$ and all $\varphi \in \Phi$.

If on top of those two axioms, F is *monotonic*, then $\mathcal{W}$ is closed under taking supersets: $C \in \mathcal{W} \Rightarrow C' \in \mathcal{W}$ for all $C, C' \subseteq \mathcal{N}$ with $C \subseteq C'$.

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 Φ *violating the median property*, then F must be a *dictatorship*.

So suppose Φ violates the MP and F has the properties on the left.

By *independence* and *neutrality*, there exists a (single) family of winning coalitions $\mathcal{W} \subseteq 2^{\mathcal{N}}$ determining F: $\varphi \in F(\boldsymbol{J}) \Leftrightarrow N_{\varphi}^{\boldsymbol{J}} \in \mathcal{W}$.

We will show that $\mathcal{W}$ is an *ultrafilter* on $\mathcal{N}$, which means:

(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 $\overline{C} := \mathcal{N} \setminus C \in \mathcal{W}$

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

Proof: Noninclusion of the Empty Set

Claim: $\emptyset \notin \mathcal{W}$.

We will use *monotonicity* and *complement-freeness:*

For the sake of contradiction, assume $\emptyset \in \mathcal{W}$.

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

But now consider some profile $\boldsymbol{J}$ with $p \in J_i$ for all individuals $i \in \mathcal{N}$.

$\leadsto$ we get $N_p^{\boldsymbol{J}} = \mathcal{N}$ and $N_{\neg p}^{\boldsymbol{J}} = \emptyset$
$\leadsto$ that is, $p \in F(\boldsymbol{J})$ and $\neg p \in F(\boldsymbol{J})$, as both $\mathcal{N} \in \mathcal{W}$ and $\emptyset \in \mathcal{W}$
$\leadsto$ contradiction with complement-freeness ✓

Proof: Maximality

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

We will use the fact that F is supposed to be *complete:*

- take any coalition $C \subseteq \mathcal{N}$ and any formula $\varphi \in \Phi$
- construct a profile $\boldsymbol{J}$ with $N_\varphi^{\boldsymbol{J}} = C$
- from completeness: $\varphi \in F(\boldsymbol{J})$ or $\sim\!\varphi \in F(\boldsymbol{J})$
- from $\mathcal{W}$-determination of F: $N_\varphi^{\boldsymbol{J}} \in \mathcal{W}$ or $N_{\sim\varphi}^{\boldsymbol{J}} \in \mathcal{W}$
- from $\boldsymbol{J}$ being complete and complement-free: $N_{\sim\varphi}^{\boldsymbol{J}} = \overline{N_\varphi^{\boldsymbol{J}}}$
- putting everything together: $C \in \mathcal{W}$ or $\overline{C} \in \mathcal{W}$ ✓

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 *mi-subset* $X = \{\varphi_1, \ldots, \varphi_k\} \subseteq \Phi$ with $k \geqslant 3$.

We can construct a complete and consistent profile $\boldsymbol{J}$ with these properties:

- $N^{\boldsymbol{J}}_{\varphi_1} = C$
- $N^{\boldsymbol{J}}_{\varphi_2} = C' \cup (\mathcal{N} \setminus C)$
- $N^{\boldsymbol{J}}_{\varphi_3} = \mathcal{N} \setminus (C \cap C')$
- $N^{\boldsymbol{J}}_{\psi} = \mathcal{N}$ for all $\psi \in X \setminus \{\varphi_1, \varphi_2, \varphi_3\}$

Thus: everyone accepts $k{-}1$ of the propositions in X. And $N^{\boldsymbol{J}}_{\sim\varphi_3} = C \cap C'$.

- $C \in \mathcal{W} \;\Rightarrow\; \varphi_1 \in F(\boldsymbol{J})$
- From monotonicity: $C' \in \mathcal{W} \;\Rightarrow\; C' \cup (\mathcal{N} \setminus C) \in \mathcal{W} \;\Rightarrow\; \varphi_2 \in F(\boldsymbol{J})$
- From maximality: $\emptyset \notin \mathcal{W} \;\Rightarrow\; \mathcal{N} \in \mathcal{W} \;\Rightarrow\; X \setminus \{\varphi_1, \varphi_2, \varphi_3\} \subseteq F(\boldsymbol{J})$

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

In other words: $N^{\boldsymbol{J}}_{\sim\varphi_3} = C \cap C' \in \mathcal{W}$ ✓

Proof: Dictatorship

We have shown that the family of winning coalitions $\mathcal{W}$ is an *ultrafilter* on the (*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 $\overline{C} := \mathcal{N} \setminus C \in \mathcal{W}$

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

Contraction Lemma: if $C \in \mathcal{W}$ and $|C| \geqslant 2$, then $C' \in \mathcal{W}$ for *some* $C' \subset C$.

> Proof: Let $C_1 \uplus C_2 = C$. If $C_1 \notin \mathcal{W}$, then $\overline{C_1} \in \mathcal{W}$ by maximality.
> But then $C \cap \overline{C_1} = C_2 \in \mathcal{W}$ by closure under intersection. ✓

By induction: $\{i^\star\} \in \mathcal{W}$ for *one* $i^\star \in \mathcal{N}$, i.e., $\mathcal{W} = \{C \subseteq \mathcal{N} \mid i^\star \in C\}$.

That is, $i^\star$ is a *dictator*. ✓

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

Second Example for a Possibility Theorem

Call an agenda Φ *well-behaved* [on this slide only] if it is not both *totally blocked* and *even-number-negatable*.

Theorem 2 (Dokow and Holzman, 2010) *There exists a unanimous, independent, and nondictatorial aggregator that is complete and consistent for a given agenda Φ iff Φ is well-behaved.*

Proof and exact definition of agenda properties: Omitted.

We get *Arrow's Theorem* as a corollary, because the *preference agenda* (discussed yesterday) is not well-behaved.

E. Dokow and R. Holzman. Aggregation of Binary Evaluations. *Journal of Economic Theory*, 145(2):495–511, 2010.

F. Dietrich and C. List. Arrow's Theorem in Judgment Aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007.

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):

Theorem 3 (Endriss et al., 2012) *An agenda Φ is safe for any anonymous, neutral, independent, complete and complement-free aggregation procedure iff Φ has the simplified median property .*

An agenda Φ has the *simplified median property* if every inconsistent subset of Φ has itself an inconsistent subset $\{\varphi, \psi\}$ with $\models \varphi \leftrightarrow \neg\psi$.

Note: The SMP is more restrictive than the MP (see: $\{\neg p, p \wedge q\}$).

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

Proof

Claim: Φ is safe for any ANI/complete/comp-free rule $F \Leftrightarrow \Phi$ has SMP

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

- $\rightsquigarrow$ $\varphi \in J_i \Leftrightarrow {\sim}\psi \in J_i$ for each individual i (from $\models \varphi \leftrightarrow \neg\psi$ together with consistency and completeness of individual judgment sets)
- $\rightsquigarrow$ $\varphi \in F(\boldsymbol{J}) \Leftrightarrow {\sim}\psi \in F(\boldsymbol{J})$ (from neutrality)
- $\rightsquigarrow$ both ψ and ${\sim}\psi$ in $F(\boldsymbol{J}) \rightsquigarrow$ contradiction (with complement-freeness) ✓

($\Rightarrow$) Suppose Φ 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. ✓ 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 $J_1 = \{{\sim}\varphi, {\sim}\psi, \ldots\}$, $J_2 = \{{\sim}\varphi, \psi, \ldots\}$, and $J_3 = \{\varphi, {\sim}\psi, \ldots\}$. Then $F(\boldsymbol{J}) = \{\varphi, \psi, \ldots\}$. ✓

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 *iff* the agenda has a given property
- Safety: *all* aggregators in the class determined by the given axioms will produce consistent outcomes *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.

Complexity of Safety of the Agenda

Deciding whether a given agenda is safe for the majority rule (as well as several classes of rules we get by relaxing the axioms defining the majority rule) is located at the second level of the polynomial hierarchy.

Proving those results involves the following lemma (and variations):

Lemma 2 (Endriss et al., 2012) *Deciding whether a given agenda has the median property is Π_2^p-complete.*

Proof: Omitted.

$\Pi_2^p = \mathrm{coNP}^{\mathrm{NP}}$ is the class of problems for which we can verify a certificate for a negative answer in polynomial time if we have access to an NP oracle. A typical problem in the class is deciding truth of formulas of the form $\forall \boldsymbol{x} \exists \boldsymbol{y} \varphi(\boldsymbol{x}, \boldsymbol{y})$. So: very hard.

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

Example: Strategic Manipulation

Suppose we use the *premise-based procedure:*

	p	q	$p \vee q$
Agent 1:	No	No	No
Agent 2:	Yes	No	Yes
Agent 3:	No	Yes	**Yes**
PBP:	No	No	No

If agent 3 only cares about the conclusion $p \vee q$, 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*

F. Dietrich and C. List. Strategy-proof Judgment Aggregation. *Economics and Philosophy*, 23(3):269–300, 2007.

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

D. Baumeister, G. Erdélyi, O.J. Erdélyi, and J. Rothe. Bribery and Control in Judgment Aggregation. Proc. COMSOC-2012.

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)

M. Slavkovik. Judgment Aggregation for Multiagent Systems. PhD thesis, University of Luxembourg, 2012.

D. Porello and U. Endriss. Ontology Merging as Social Choice: Judgment Aggregation under the Open World Assumption. *J. Logic and Computation*. In press.

U. Endriss and R. Fernández. Collective Annotation of Linguistic Resources: Basic Principles and a Formal Model. Proc. ACL-2013.

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 $\Leftrightarrow$ $\exists$ consistent rule in class*
  - *safety:* agenda property $\Leftrightarrow$ $\forall$ rules in class are consistent
  - both: agenda has *median property* $\Leftrightarrow$ *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.