Computational Social Choice: Spring 2015

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

Plan for Today

We continue with identifying properties of agendas that facilitate well-behaved judgment aggregation:

- *Existential Agenda Characterisation:*
  Fix a class of aggregation rules by means of fixing some axioms. For which range of agendas is there a consistent rule in that class?

We also discuss how *embedding preference aggregation* into JA allows us to recover classical results, specifically *Arrow's Theorem*.

The results covered today have originally been proved for somewhat different formal frameworks than we use here. For presentations close to what we do here, refer to the expository papers cited below.

C. List and C. Puppe. Judgment Aggregation: A Survey. In P. Anand, P. Pattanaik, and C. Puppe (eds.), *Handbook of Rational and Social Choice*. OUP, 2009.

U. Endriss. Judgment Aggregation. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A.D. Procaccia (eds.), *Handbook of Computational Social Choice*. CUP, 2015.

Agenda Characterisation: Bigger Picture

Consider a *class of rules*, possibly determined by a set of *axioms:*

- *Existential Agenda Characterisation*

  Question: Is there *some* rule meeting certain axioms that is consistent for every agenda with a given property?

- *Universal Agenda Characterisation* ("*Safety of the Agenda*")

  Question: Is *every* rule meeting certain axioms consistent for every agenda with a given property?

Note: In case the class of rules includes just a single rule, the two types of results coincide.

Reminder: Consistent Aggregation under Majority

Recall that an agenda Φ satisfies the *median property* (MP) *iff* all its *minimally inconsistent subsets* have *size at most 2*.

We had proved:

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

If we look at a broader class of rules, can we find one that is consistent for a wider range of agendas?

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.

Axioms

We will use the following axioms (the last two are new) for rules F:

- F is *neutral* if $N_\varphi^{\boldsymbol{J}} = N_\psi^{\boldsymbol{J}}$ implies $\varphi \in F(\boldsymbol{J}) \Leftrightarrow \psi \in F(\boldsymbol{J})$.
- F is *independent* if $N_\varphi^{\boldsymbol{J}} = N_\varphi^{\boldsymbol{J}'}$ implies $\varphi \in F(\boldsymbol{J}) \Leftrightarrow \varphi \in F(\boldsymbol{J}')$.
- F is *monotonic* if $N_\varphi^{\boldsymbol{J}} \subset N_\varphi^{\boldsymbol{J}'}$ implies $\varphi \in F(\boldsymbol{J}) \Rightarrow \varphi \in F(\boldsymbol{J}')$.
- F is (propositionwise) *unanimous* of $N_\varphi^{\boldsymbol{J}} = \mathcal{N}$ implies $\varphi \in F(\boldsymbol{J})$.
- F is a *dictatorship* if there exists an agent $i^\star \in \mathcal{N}$ (the dictator) such that $F(\boldsymbol{J}) = J_{i^\star}$ for every profile $\boldsymbol{J}$.

  Otherwise, F is *nondictaorial*.

Discussion: Note how nondictatoriality is a weakening of anonymity.

An Existential Agenda Characterisation Theorem

Let again $n \geqslant 3$ (number of agents).

Theorem 2 (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 MP.*

Proof: One direction (right-to-left) follows from Theorem 1:

Suppose Φ has the MP.

$\rightsquigarrow$ the majority rule will be consistent and complete (previous result)

$\rightsquigarrow$ there exists a rule 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.

Reminder: Winning Coalitions

F is *independent* *iff* there exists a family of *winning coalitions* of agents $\mathcal{W}_\varphi \subseteq 2^{\mathcal{N}}$, one for each $\varphi \in \Phi$, s.t. $\varphi \in F(\boldsymbol{J}) \Leftrightarrow N_\varphi^{\boldsymbol{J}} \in \mathcal{W}_\varphi$.

F is independent and *neutral* if furthermore we have $\mathcal{W}_\varphi = \mathcal{W}_\psi$ for all formulas $\varphi, \psi \in \Phi$. So simply write $\mathcal{W}$.

Remark: As discussed before, this is only true for nontrivial agendas (e.g., including at least two atoms). So we assume that.

Now suppose F is independent and neutral, and defined by $\mathcal{W}$. Then:

- F is *monotonic* *iff* $\mathcal{W}$ is upward closed: $C \in \mathcal{W}$ and $C \subseteq C'$ entail $C' \in \mathcal{W}$ for all $C, C' \subseteq \mathcal{N}$.
- F is *complete* *iff* $\mathcal{W}$ is maximal: $C \in \mathcal{W}$ or $\overline{C} \in \mathcal{W}$ for all $C \subseteq \mathcal{N}$.
- F is *complement-free* *iff* $C \notin \mathcal{W}$ or $\overline{C} \notin \mathcal{W}$ for all $C \subseteq \mathcal{N}$.

Exercise: What does $\mathcal{W}$ look like for a dictatorship F?

Proof Plan: Impossibility Direction

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

> Claim: If a *neutral*, *independent*, and *monotonic* rule F is *complete* and *consistent* for an agenda Φ *violating the MP*, 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 $i \in \mathcal{N}$.

$\rightsquigarrow$ we get $N_p^{\boldsymbol{J}} = \mathcal{N}$ and $N_{\neg p}^{\boldsymbol{J}} = \emptyset$
$\rightsquigarrow$ that is, $p \in F(\boldsymbol{J})$ and $\neg p \in F(\boldsymbol{J})$, as both $\mathcal{N} \in \mathcal{W}$ and $\emptyset \in \mathcal{W}$
$\rightsquigarrow$ 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: $\emptyset \not\in \mathcal{W} \;\Rightarrow\; \mathcal{N} \in \mathcal{W} \;\Rightarrow\; X \setminus \{\varphi_1, \varphi_2, \varphi_3\} \subseteq F(\boldsymbol{J})$

Thus: for consistency we need $\varphi_3 \not\in F(\boldsymbol{J})$, 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: $\mathcal{N} \in \mathcal{W}$ (b.t.w.: 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 Characterisation Result

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

Theorem 3 (Dokow and Holzman, 2010) *There exists a unanimous, independent, and nondictatorial rule that is complete and consistent for a agenda Φ iff Φ is well-behaved.*

Proof and exact definition of agenda properties: Omitted.

Remark: An added challenge here is to prove neutrality, so we can work with (one family of) winning coalitions. Rest of the proof is similar.

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

Relevance to Preference Aggregation

What makes the impossibility direction of Dokow and Holzman's result particularly interesting is that it may be considered a generalisation of the most famous theorem in social choice theory:

- *Arrow's Theorem* for *preference aggregation* (*next slide*)

We have already seen how to embed preference aggregation into JA (actually: we have seen it only for binary aggregation, but it's similar). For a proof of using this embedding, see Dietrich and List (2007).

For a direct proof (directly in preference aggregation) using the same ultrafilter technique we have seen, refer to my paper cite below.

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

U. Endriss. Logic and Social Choice Theory. In A. Gupta and J. van Benthem (eds.), *Logic and Philosophy Today*. College Publications, 2011.

Arrow's Theorm

Let $\mathcal{X}$ be a finite set of *alternatives*, $\mathcal{L}(\mathcal{X})$ the set of *linear orders* on $\mathcal{X}$, and $\mathcal{N} = \{1, \ldots, n\}$ a set of *agents*.

A *preference aggregator* $F : \mathcal{L}(\mathcal{X})^n \to \mathcal{L}(\mathcal{X})$ might be:

- *Pareto:* if every agent prefers x over y, then so should the collective preference order returned by F
- *IIA* (independent of irrelevant alternatives): the relative ranking of x and y in the output only depends on how agents rank x and y

Theorem 4 (Arrow, 1951) *For $|\mathcal{X}| \geqslant 3$, every preference aggregator that satisfies the Pareto and IIA conditions must be a dictatorship.*

Here as well: *dictatorship* $=$ output always copied from dictator in $\mathcal{N}$

K.J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963. First edition published in 1951.

Summary

We have seen two (existential) agenda characterisation theorems of the following form:

> There exists a nondictatorial complete and consistent *rule* meeting certain *axioms* $\Leftrightarrow$ the *agenda* has a certain *property*.

Both directions are of interest:

($\Leftarrow$) *Possibility direction:* If the agenda property holds for your problem, then "reasonable" and consistent aggregation is possible.

($\Rightarrow$) *Impossibility direction:* For structurally rich domains, all seemingly "reasonable" rules are in fact dictatorial.

Possibility is proved by providing a concrete rule doing the job.

Impossibility is (sometimes) proved using the *ultrafilter technique*.

What next?

We will see yet another connection between the axioms constraining the rules and the structural properties of the domain of aggregation those rules may or may not respect.

But we will switch from *semantic* properties (e.g., the median property) to *syntactic* properties (e.g., not using disjunction).

To do so, we return to *binary aggregation with integrity constraints*.