Aggregating value judgments

Wlodek Rabinowicz

Stephan Hartmann, Soroush Rafiee Rad

My focus

aggregation of *preferences*

vs

aggregation of *judgments*

with the relevant judgments concerning an evaluative question:

What is to be done?

What is the best alternative?

How are alternatives to be ranked from the best to the worst?

Preferring one alternative to another is desiring it more.
It is *not* the same as judging it to be better.

Preferences often are based on value judgments.

But not always: I might prefer *a* to *b*, while lacking a clear view about their relative value.

Indeed, I might even judge *b* to be better than *a* and still prefer *a* to *b*
(perhaps because I believe that *a* is better *for me*,
though I believe *b* to be better overall;
or perhaps because I am simply irrational).

Consequently, aggregation of preferences is not reducible to aggregation of value judgments.

Recent work on judgment aggregation

List & Pettit, Dietrich & List, Nehring & Puppe, Dokov & Holzman, Mongin, Pauly & Van Hees, Gärdenfors, Pigozzi, Bovens & Rabinowicz, …

See Christian List's bibliography: http://personal.lse.ac.uk/list/doctrinalparadox.htm#23

Much of this work is concerned with the issue:
Is there a procedure for judgment aggregation that satisfies reasonable requirements?

Some plausible requirements on *judgment aggregation*

- non-dictatorship,
- unanimity,
- universal domain
- consistency [collective judgments are internally consistent]
- anonymity [equal treatment of voters],
- neutrality [equal treatment of propositions on the agenda],
- independence [collective judgment on each proposition should only depend on the individual judgments on *that* proposition]

…

Impossibility theorems for judgment aggregation.
Analogy:
Arrow's impossibility theorem for preference aggregation.

Similar points of departure for both kinds of theorems.

Cf. *discursive dilemma* (doctrinal paradox):
Majority for *A*, majority for *B*, but majority against *A*&*B*
Possible if the overlap between the majorities for *A* and for *B* isn't too large.

Analogy: *Condorcet's voting paradox*.
Majority for *a* over *b*, majority for *b* over *c*, but majority against *a* over *c*.
Possible if the overlap between the majorities for *a* over *b* and for *b* over *c* isn't too large.

A way to finesse the impossibility theorems

Think of the aggregation procedure as an optimization task.
The output should reflect individual imputs as much as possible.
I.e. the task is to find an output that is *maximally similar* to individual inputs.
(For preference aggregation, Kemeny 1959, Kemeny & Snell 1962.
For judgment aggregation, Pigozzi 2006, Miller & Osherson 2009)

Inputs and output – objects of the same category.

Example: Suppose that individual inputs are *rankings* and that an output should also be a ranking.
- For preference aggregation $\Rightarrow$ preferential rankings
- For judgment aggregation $\Rightarrow$ value rankings

Optimization procedure need not deliver a unique result!

(i) Specify a *distance measure* between rankings

Ex. Kemeny-Snell: Count the number of ordered pairs with respect to which two rankings differ.

A pair (*a*, *b*) belongs to a ranking *x* if and only *x* ranks *a* at least as highly as *b*.

The more pairs two rankings have in common the more similar they are.

Another distance measure can be chosen instead!
Cook & Seiford (1978), Duddy & Piggins (2011)

(ii) **Minimize the output's distance to inputs.**

Ex**:** Kemeny rule: Minimize the average distance.

Another rule can be chosen instead!
Subject distances to inputs to a convex transform (ex. squaring) and take the average. Or: leximin.

Another example of the distance minimization approach

If inputs are sets of judgments from a certain agenda and the output is also a set of judgments from the same agenda,

then we can use as the measure of distance
the number of judgments with respect to which the output differs from an input,
i.e. the number of judgments that belong to the input or to the output, but not to both.

("Hamming distance")

and choose the output that minimizes the average distance.

(Pigozzi 2006)

Distance between rankings can be re-interpreted as distance between sets of judgments.

A value ranking of *a*, *b*, *c* can be represented as a set of judgments:
{*a* is at least as good as *b*, *b* is not at least as good as *a*, *a* is at least as good as *c*, …}

Aggregation of rankings is then reducible to the aggregation of judgment sets.

(Cf. List & Pettit, 2004, Dietrich& List, 2007)

In particular,the KS-distance between rankings reduces to the Hamming distance between the corresponding judgment sets

$\Rightarrow$ Kemeny rule (=minimization of the average KS-distance) is reducible to the minimization of average Hamming distance.

Kemeny & Snell have shown that their distance measure between rankings is the only one that satisfies their axioms, one of which is *the axiom of betweenness*:

If x lies between y and z , then $d(x, z) = d(x,y) + d(y,z)$.

But this result is based on a stringent definition of betweenness that might be questioned:

x lies *between y* and *z* if and only if

(i) the intersection of *y* and *z* is included in *x,* and

(ii) *x* is included in the union of *y* and *z.*

Other axioms imposed by K&S are rather unproblematic:

Neutrality: d is invariant under permutations of alternatives;

Reduction: if x and y agree on their top (bottom) alternatives, then d(x,y) = the distance between reduced *x* and *y*, with the top (bottom) alternatives removed.

Minimum: Minimal positive distance is 1.

The KS measure is not very favourable to compromises

Example: Suppose the input rankings are as in Condorcet's paradox:

x	*y*	*z*
a	*b*	*c*
b	*c*	*a*
c	*a*	*b*

The natural compromise in this case would be to go for the equal ranking

a,b,c

as output.
But the average distance of that ranking to {*x, y, z*} >
the average distance to {*x, y, z*} from each ranking in {*x, y, z*}.
(3 > 8/3)

A distance measure that avoids this problem (Cook and Seiford 1978)
Assign *CS-numbers* to the alternatives in a ranking, starting with 1 for the top alternative. (Alternatively, one can start from the bottom.)
In a tie, assign the average number to the alternatives.
Ex.: If two alternatives on top, each gets 1,5 [(1+2)/2].

a_i^x - the CS-number assigned to a_i in x.

Distance between rankings:

$$d(x, y) = \Sigma_i \left| a_i^x - a_i^y \right|$$

Variant of Borda count.
On this proposal, the equal ranking (a, b, c) does minimize the average distance to the rankings in Condorcet's paradox:

x	y	z
a	b	c
b	c	a
c	a	b

The average distance of that ranking to {x, y, z} is 2, while
the average distance to {x, y, z} from each ranking in {x, y, z} is 8/3.

Duddy and Piggins (2011)

The distance between two rankings = the smallest number of steps needed to transform one ranking into the other.
Just one step is needed to move from a ranking *x* to another ranking *y* iff one can reach *y* from *x* by raising or lowering the position of some alternative *a* with respect to some set *S* of alternatives that are equal-ranked in *x* and making no other changes in the relative positions of the alternatives.
This is possible only in two cases:

(i) *a* is equal-ranked in *x* with the alternatives in *S,*
(ii) *a* in *x* is immediately above or below *S*.

If (i), then one can move *a* to a position immediately above or below *S*.
If (ii), then one can move *a* to *S*'s level.

Example of a 1-step change:

<u>*x*</u>	<u>*y*</u>
a	
b c	a b c

Question: Are there any differences in formal requirements on the aggregation procedure depending on *what* is being aggregated?

Many standard requirements seem to be the same:
Non-dictatorship, unanimity, anonymity, universal domain, etc.
One formal difference, though, *when the inputs are rankings*: **Pareto**.

<u>The Pareto Condition</u>:
If some individuals rank *a* above *b* while everyone else ranks *a* and *b* equally, then *a* is ranked above *b* in the output.

Intuitively plausible for preference aggregation,
but *not* for judgment aggregation.

Observation 1:

Minimization of average distance violates Pareto,
if the individuals who rank *a* above *b* are in minority (while the majority ranks *a* and *b* equally), and *a*, *b* are the only alternatives.

This result obtains for all possible distance measures.

It also holds when minimization of average distance is replaced by, say, minimization of average squared distance, etc, or by leximin.
Tringle inequality isn’ used in the proof. I.e. the proof can be generalized to all similarity measures.

But if there are other alternatives apart from *a* and *b*, and some of them come in between *a* and *b* in some input rankings,
things might be different.

A related condition:

Indifference: The collective ranking of the alternatives doesn't change if voters who rank all the alternatives equally are removed from consideration (as long as some voters still remain to be considered).

Indifference + Unanimity + Independence of Irrelevant Alternatives $\Rightarrow$ Pareto.

Again, Indifference is fine for preference aggregation, but not for aggregation of value rankings.

Observation 2:
Minimization of average distance violates Indifference, if the individuals who rank *a* above *b* are in minority (while the majority ranks *a* and *b* equally), and *a* and *b* are the only alternatives.

The equal ranking is the collective output here (cf. Observation 1), but removing voters who rank *a* and *b* equally changes the collective ranking to the one in which *a* is placed above *b*.

Again, this result holds for all possible distance measures (and more generally for all similarity measures) and it also holds for other distance-based aggregation rules.

Main lesson

(i) Pareto and Indifference represent an important dividing line between preference aggregation and aggregation of value rankings.

(ii) The distance-based approach to aggregation violates Pareto and Indifference.

(iii) Therefore this approach doesn't fit preference aggregation (insofar as the latter should respect these conditions).

(iv) But it does seem ok for judgment aggregation (insofar as Pareto and Indifference are implausible for the aggregation of value rankings).

Other formal differences between preference aggregation and aggregation of judgments, apart from Pareto?

- room for abstaining (for individuals, for collectives)?
- logical structure?
- anonymity condition?

Condorcet jury theorem:
A majority judgment is more epistemically reliable than that of a single voter, if voters are independent and relatively competent. The reliability of the majority judgment converges to 1 if the number of voters increases.

Can this epistemic approach be used for the aggregation of value judgments?

Yes,

if such judgments are independently true or false (or correct/incorrect)

and

if epistemic competence with respect to value judgments is possible

If the aggregation procedure consists in similarity maximization,
then how does such a procedure fare from the epistemic standpoint?
How good is it as a truth-tracker?

Can one prove something like Condorcet's theorem for this kind of procedure?

And what about its truth-tracking performance in comparison with other aggregation procedures?

How do different distance measures perform on this score?

We can try to compute, for an aggregation outcome, not only the *probability of truth*,
i.e. the probability of it being the true ranking,
but also the *expected verisimilitude,*
i.e. its expected distance to the true ranking.

A reasonable question about a distance-based procedure:
How good is it, as compared with the average voter,
not only when it comes to the probability of truth,
but also when it comes to the expected verisimilitude?

Summing up

- My focus: Contrast between aggregation of preferences and aggregation of judgments (and especially value rankings).
- Distance-based methods seem to be an attractive way of finessing impossibility results.
- But these methods are appropriate only for judgment aggregation and not for preference aggregation,
due to the role of the Pareto condition.
- Distance-based methods for the aggregation of judgments, and in particular of value rankings, are interesting from the epistemic perspective.
- The epistemic question: How good are these methods
in increasing the probability of truth
and in increasing the expected verisimilitude?

The Epistemic Perspective

- We have seen that there are several different distance-based aggregation procedures for value judgments, including the Kemeny-Snell measure, the Cook and Seiford measure and the Duddy and Piggins measure.
- Let us now assume that there is a fact of the matter which ordering is the right one and ask the following questions:
  1. Which aggregation procedure is the best truth-tracker?
  2. Which procedure minimizes the distance to the truth?
- To do so, we focus on the special case of three alternatives and make a number of modeling assumptions, guided by an analogy to the Condorcet Jury Theorem.

Possible Orderings for Three Alternatives

$$O_1 := \begin{vmatrix} a, b, c \end{vmatrix}$$

$$O_2 := \begin{vmatrix} a, b \\ c \end{vmatrix}, \quad O_3 := \begin{vmatrix} a, c \\ b \end{vmatrix}, \quad O_4 := \begin{vmatrix} b, c \\ a \end{vmatrix}$$

$$O_5 := \begin{vmatrix} a \\ b, c \end{vmatrix}, \quad O_6 := \begin{vmatrix} b \\ a, c \end{vmatrix}, \quad O_7 := \begin{vmatrix} c \\ a, b \end{vmatrix}$$

$$O_8 := \begin{vmatrix} a \\ b \\ c \end{vmatrix}, \quad O_9 := \begin{vmatrix} a \\ c \\ b \end{vmatrix}, \quad O_{10} := \begin{vmatrix} b \\ a \\ c \end{vmatrix}$$

$$O_{11} := \begin{vmatrix} b \\ c \\ a \end{vmatrix}, \quad O_{12} := \begin{vmatrix} c \\ a \\ b \end{vmatrix}, \quad O_{13} := \begin{vmatrix} c \\ b \\ a \end{vmatrix}$$

The Model

- Consider a group of n members.
- Each group member i has a certain competence r_i to identify the true value ordering $O^* \in \{O_1, \ldots, O_{13}\}$.
- In a *homogeneous group*, all group members have the same competence, in an *non-homogeneous group* not.
- Each group member i submits a value ordering o_i.
- The individual orderings are then aggregated to obtain to a social ordering using one of the three similarity measures.
- The outcome is then evaluated according to the two mentioned criteria: (i) which procedure is the best truth tracker? (ii) which procedure minimizes the distance to the truth?

Details I: Reliabilities and Identifying an Ordering

Model I: We assume that the group member i picks the right ordering with probability r_i and one of the other orderings with probability $(1 - r_i)/12$.

Model II: We assume that the group member i picks the right ordering with probability r_i and one of the other orderings with a probability which is related to the distance of the ordering to the true ordering.

Details II: Evaluation (1)

We evaluate the result of the aggregation procedure according to two criteria.

1. Truth Tracking. We calculate the probability that the aggregation procedure D tracks the truth. Let $\mathrm{F_D}$ be the proposition "The aggregation procedure identifies the true ordering", i.e. it picks out O_1 if O_1 is true, etc. Then

$$P(\mathrm{F_D}) = \sum_{i=1}^{13} P(\mathrm{F_D}|O_i)\, P(O_i)$$

N.B. The conditional probabilities $P(\mathrm{F_D}|O_i)$ can be obtained from a computer simulation.

Details II: Evaluation (2)

2. Verisimilitude. First, assume that O_i is the true ordering. Then D picks out one of the O_i (with $i = 1, \ldots, 13$) with a certain probability. Let $\mathrm{S}_{\mathrm{D}}^{(\mathrm{j})}$ be the proposition "The aggregation procedure picks out ordering O_j". The average distance of the picked ordering from the truth is then given by

$$V_i^{(D)} = \sum_{j=1}^{13} P(\mathrm{S}_{\mathrm{D}}^{(\mathrm{j})}|O_i)\, d_D(O_i, O_j).$$

Second, the *expected distance from the truth* $V^{(D)}$ is

$$V^{(D)} :=< V_i^{(D)} >= \sum_{i=1}^{13} V_i^{(D)}\, P(O_i)\ .$$

Again, the $P(\mathrm{S}_{\mathrm{D}}^{(\mathrm{j})}|O_i)$ can be obtained from a computer simulation.

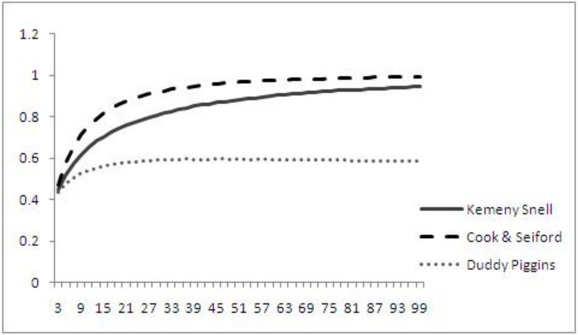
Results for Model I

We plot $P(\mathrm{F_D})$ and $V^{(D)}$ as a function of the number of group members (n) for fixed values of the reliability parameters (r_i) for (i) homogeneous and (ii) non-homogeneous groups for the three measures.

We assume that all orderings have the same prior probability, i.e. that $P(O_i) = 1/13$, for $i = 1, \ldots, 13$.

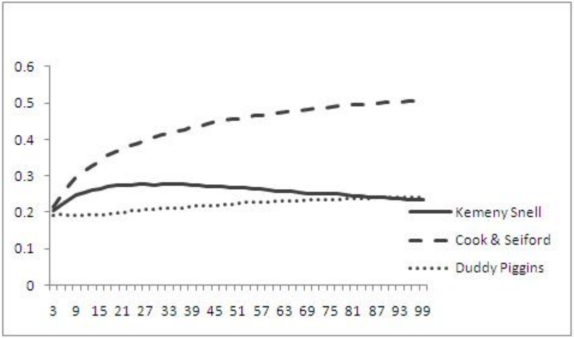
$\mathrm{F_D}$ for Homogeneous Groups

All group members have a reliability of 0.4.



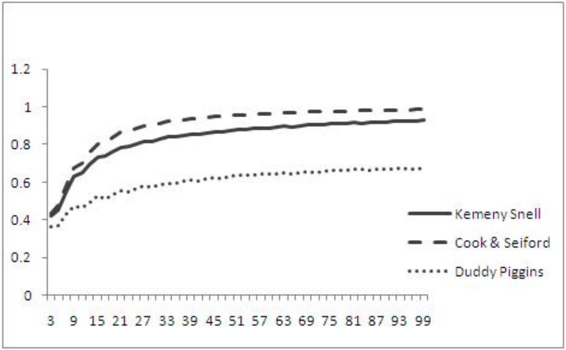
$\mathrm{F_D}$ for Homogeneous Groups Cont'd

All group members have a reliability of 0.2.



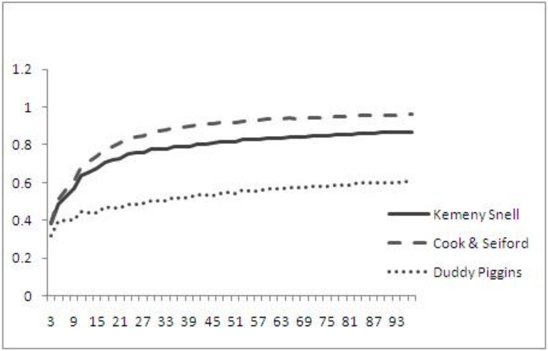
$\mathrm{F_D}$ for Non-Homogeneous Groups

One third of the group has a reliability of 0.5, all other group members have a reliability of 0.3.



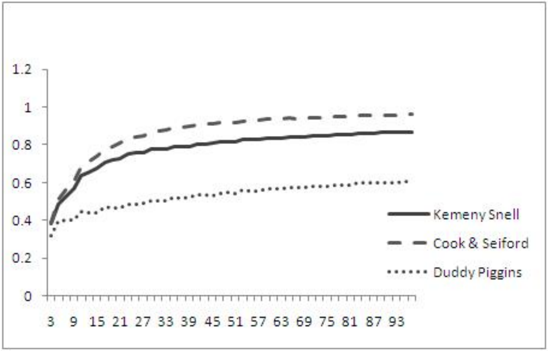
$\mathrm{F_D}$ for Non-Homogeneous Groups

One third of the group has a reliability of 0.2, all other group members have a reliability of 0.4.



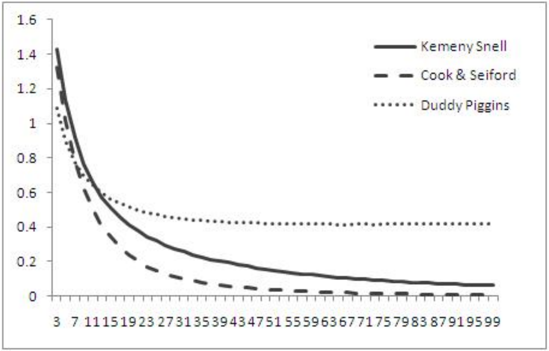
$\mathrm{F_D}$ for Non-Homogeneous Groups Cont'd

One third of the group has a reliability of 0.5, all other group members have a reliability of 0.3.



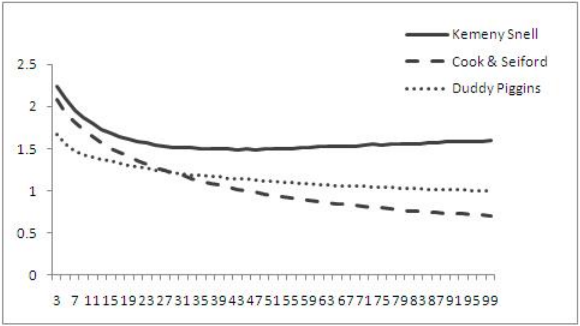
$\mathrm{V_D}$ for Homogeneous Groups

All group members have a reliability of 0.4.



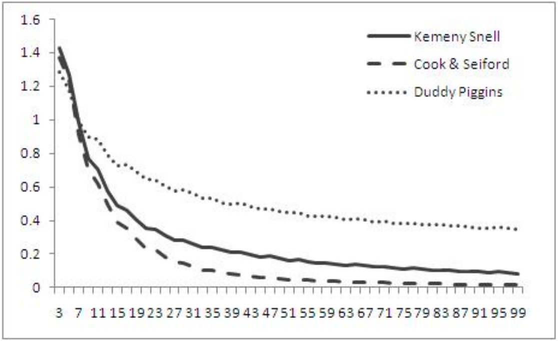
V_D for Homogeneous Groups Cont'd

All group members have a reliability of 0.2.



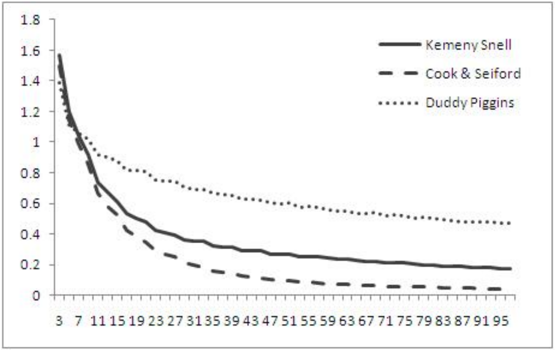
V_D for Non-Homogeneous Groups

One third of the group has a reliability of 0.5, all other group members have a reliability of 0.3.



$\mathrm{V_D}$ for Non-Homogeneous Groups Cont'd

One third of the group has a reliability of 0.2, all other group members have a reliability of 0.4.



Results for Model II

- The upshot of our simulations is that what matters is which distance measure is used when the group members pick their ordering.
- Once this is done, all three measures perform equally well with regard to the aggregation (for truth tracking as well as for the verisimilitude).

Outlook

In future work, we want to address the following problems:

1. Explain why the Cook & Seiford measure does so well.
2. Explore which distance measure real people use (joint work with Ulrike Hahn).

Thanks for your attention!