Scoring rules for judgment aggregation

Franz Dietrich
CNRS (Paris, France) & Uni. East Anglia (Norwich, U.K.)
www.franzdietrich.net

PET 13

Lisbon, Portugal

July 2013

Background

- The JA problem: How can/should we merge many individuals' yes/no judgments on some interconnected propositions?
- Very general problem!

Example 1: Preference aggregation

- Propositions: of the form 'option x is better than option y'.
- Interconnections: given by transitivity, etc.
- Propositionwise majority voting ($=$ pairwise majority voting) generates inconsistent collective judgment sets

Example 2: Jury example

- Propositions:
  - p : the defendant has broken the contract;
  - q : the contract is legally valid;
  - r : the defendant is liable.
- Interconnections: Following legal doctrine, r (the 'conclusion') is true if and only if both p and q (the 'premises') are true

Example 2 (cont.)

- Propositionwise majority rule may again generate inconsistent collective judgments:

	premise p	premise q	conclusion r ($\Leftrightarrow p \wedge q$)
Juror 1	Yes	Yes	Yes
Juror 2	Yes	No	No
Juror 3	No	Yes	No
Majority	Yes	Yes	No

Current stage of theory

- After all these impossibility theorems, time to construct concrete JA rules!
  - An experimental, playful, 'fun' phase.
  - Much seems permitted: we can try out rules.

Paradigms on the market

- **Premise- and conclusion-based rules** (e.g., Kornhauser and Sager 1986, Pettit 2001, List & Pettit 2002, Dietrich 2006, Dietrich and Mongin 2010)
- **Sequential priority rules** (e.g., List 2004, Dietrich and List 2007)
- **Quota rules with 'well-calibrated' thresholds/quota** (e.g., Dietrich and List 2007)
- **Distance-based rules** (e.g., Konieczny & Pino-Perez 2002, Pigozzi 2005, Miller & Osherson 2008, Eckert & Klamler 2009, Hartmann , Pigozzi & Sprenger 2010, Lang, Pigozzi, Slavkovik & van der Torre 2011, Duddy and Piggins 2011)
- **'Condorcet admissible' aggregation** (Nehring, Pivato and Puppe 2011)
- An **(incomplete) Borda-type proposal** (Zwicker 2011)

New proposal: scoring rules

- Idea: the set of collective judgments should have highest total 'score'.
- Inspired from classical scoring rules in preference aggregation theory, such as Borda rule (e.g., Smith 1973, Young 1975, Myerson 1995, Zwicker 2008, Pivato 2011))

Strength of judgment?

Goals for today

- Define scoring rules.
- Explore various ways to define scores
  - some lead to ('rationalize') existing aggregation rules,
  - others to new rules
- ... such as a Borda rule for JA!
  - 'Generalizing Borda to JA': a long-lasting open problem.
  - Bill Zwicker (2011), and Conal Duddy and Ashley Piggins (2013) have proposals.

Plan

Part 1: The JA framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Plan

Part 1: The JA framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

The agenda

- Set of n (≥ 2) individuals, denoted $N = \{1, ..., n\}$.
- **Agenda** of propositions on which judgments are needed. Formally, the agenda is any finite set X (of 'propositions') endowed with
  - a partition into binary 'issues' $\{p, p'\}$ (whose members p and p' are the 'negations' of each other, written $\neg p = q$ and $q = \neg p$),
  - interconnections, i.e., a specification of which judgment sets $J \subseteq X$ are rational, or formally, a set $\mathcal{J}$ of ('rational') sets $J \subseteq X$, each containing exactly one proposition from each issue.

Notation

- A judgment set is often abbreviated by concatenating its members:
  –> $p\neg q\neg r$ is short for $\{p, \neg q, \neg r\}$

Example 1: the ‘doctrinal paradox’ agenda

- This agenda is

$$X = \{p, \neg p, q, \neg q, r, \neg r\},$$

- where logical interconnections are defined relative to the external constraint $r \leftrightarrow (p \wedge q)$. So, there are 4 rational judgment sets:

$$\mathcal{J} = \{pqr, p\neg q\neg r, \neg pq\neg r, \neg p\neg q\neg r\}.$$

Example 2: the preference agenda

- For an arbitrary, finite set of alternatives A, the *preference agenda* is defined as

$$X = X_A = \{xPy : x, y \in A, x \neq y\},$$

- where the negation of xPy is of course $\neg xPy = yPx$,
- and where interconnections are defined relative to the usual conditions of transitivity, asymmetry and connectedness, which define a *strict linear order*.
- Formally, to each binary relation $\succ$ over A uniquely corresponds a judgment set, denoted $J_\succ = \{xPy \in X : x \succ y\}$, and the set of all rational judgment sets is

$$\mathcal{J} = \{J_\succ : \ \succ \text{ is a strict linear order over } A\}.$$

Aggregation rules

- A **(multi-valued) aggregation rule** is a correspondence F which to every profile of 'individual' judgment sets $(J_1, ..., J_n)$ (from some domain, usually $\mathcal{J}^n$) assigns a set $F(J_1, ..., J_n)$ of 'collective' judgment sets.
- Typically, the output $F(J_1, ..., J_n)$ is a singleton set $\{C\}$, in which case we identify this set with C and write $F(J_1, ..., J_n) = C$.

Aggregation rules

- A standard (single-valued) aggregation rule is **majority rule**, given by

$$F(J_1, ..., J_n) = \{p \in X : |\{i : p \in J_i\}| > n/2\}.$$

- It generates inconsistent collective judgment sets for many agendas and profiles.
- If both individual and collective judgment sets are rational (i.e., in $\mathcal{J}$), the aggregation rule defines a correspondences $\mathcal{J}^n \rightrightarrows \mathcal{J}$, and in the case of single-valuedness a function $\mathcal{J}^n \to \mathcal{J}$.

Plan

Part 1: The JA framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.a: Definition

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Definition

- Scoring rules are aggregation rules defined on the basis of a *scoring function* (or '*scoring*').
- A **scoring** is a function $s : X \times \mathcal{J} \to \mathbb{R}$ which to each proposition p and rational judgment set J assigns a number $s_J(p)$, called the *score* of p given J and measuring how p performs from the perspective of judgment set J.
- E.g., **simple scoring** is given by:

$$s_J(p) = \begin{cases} 1 & \text{if } p \in J \\ 0 & \text{if } p \notin J, \end{cases} \tag{1}$$

Definition (cont.)

- For any scoring s, the **scoring rule w.r.t.** s is the aggregation rule $F_s : \mathcal{J}^n \rightrightarrows \mathcal{J}$ given by

$$F_s(J_1, ..., J_n) = \text{judgment set(s) in } \mathcal{J} \text{ with highest total score}$$
$$= \text{argmax}_{C \in \mathcal{J}} \sum_{p \in C, i \in N} s_{J_i}(p).$$

- By a 'scoring rule' simpliciter we of course mean an aggregation rule which is a scoring rule w.r.t. *some* scoring.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.b: Simple scoring and the Kemeny rule

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Simple scoring illustrated

- For *simple* scoring (1), the scoring rule works as follows in the face of the 'doctrinal paradox' agenda and profile:

	Score of...									
Individual	p	$\neg p$	q	$\neg q$	r	$\neg r$	pqr	$p\neg q\neg r$	$\neg pq\neg r$	$\neg p\neg q\neg r$
1 (pqr)	1	0	1	0	1	0	3	1	1	0
2 ($p\neg q\neg r$)	1	0	0	1	0	1	1	3	1	2
3 ($\neg pq\neg r$)	0	1	1	0	0	1	1	1	3	2
Group	2	1	2	1	1	2	5*	5*	5*	4

- So: a tie between the premise-based outcome pqr and the conclusion-based outcomes $p\neg q\neg r$ and $\neg pq\neg r$. Formally:

$$F(J_1, J_2, J_3) = \{pqr, p\neg q\neg r, \neg pq\neg r\}.$$

Distance-based rules

- Consider any **distance function** ('metric') d over $\mathcal{J}$.[1]
- Most common example: **Kemeny distance** $d = d_{\text{Kemeny}}$, given by:

$$d_{\text{Kemeny}}(J, K) = \text{number of judgment reversals}$$
$$\text{needed to transform } J \text{ into } K \qquad (2)$$
$$= |J \backslash K| = |K \backslash J| = \frac{1}{2} |J \triangle K| .$$

E.g., the Kemeny-distance between pqr and $p\neg q\neg r$ (for our doctrinal paradox agenda) is 2.

[1] A *distance function* or *metric* over $\mathcal{J}$ is a function $d : \mathcal{J} \times \mathcal{J} \to [0, \infty)$ satisfying three conditions: for all $J, K, L \in \mathcal{J}$, (i) $d(J, K) = 0 \Rightarrow J = K$, (ii) $d(J, K) = d(K, J)$ ('symmetry'), and (iii) $d(J, L) \leq d(J, K) + d(K, L)$ ('triangle inequality').

Distance-based rules (cont.)

- The **distance-based rule** w.r.t. a distance d is the aggregation rule F_d which for any profile $(J_1, ..., J_n) \in \mathcal{J}^n$ returns:

$$F_d(J_1, ..., J_n) = \text{judgment set(s) in } \mathcal{J} \text{ with minimal sum-distance to the profile}$$

$$= \text{argmin}_{C \in \mathcal{J}} \sum_{i \in N} d(C, J_i).$$

Distance-based rules

- The most popular example, *Kemeny rule* $F_{d_{\mathsf{Kemeny}}}$, can be characterized as a scoring rule:

Proposition 1 *The simple scoring rule is the Kemeny rule.*

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.c: Classical scoring rules for preference aggregation

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Classical scoring

- Consider the preference agenda X for a given set of alternatives A of finite size k.
- Classical scoring rules (such as Borda rule) are defined by assigning scores to alternatives in A, not to propositions xPy in X.
- Given a strict linear order $\succ$ over A – equivalently, a rational judgment set $J \in \mathcal{J}$, each alternative $x \in A$ is assigned a score $SCO_J(x) \in \mathbb{R}$.
- *Borda scoring*: the highest ranked alternative in A scores k, the second-highest scores $k - 1$, ...

Classical scoring rules

- Given a profile $(J_1, ..., J_n)$ of rational judgment sets (equivalently, strict linear orders), the collective ranks the alternatives $x \in X$ according to their sum-total score $\sum_{i \in N} SCO_{J_i}(x)$.
- Formally, a *classical scoring* is a function $SCO : A \times \mathcal{J} \to \mathbb{R}$.
- The *classical scoring rule* w.r.t. SCO is the JA rule $F \equiv F_{SCO}$ for the preference agenda which for every profile $(J_1, ..., J_n) \in \mathcal{J}^n$ returns:

$$F(J_1, ..., J_n) = \{C \in \mathcal{J} : C \text{ contains all } xPy \in X \text{ s.t. } \sum_{i \in N} SCO_{J_i}(x) > \sum_{i \in N} SCO_{J_i}(y)\}.$$

Classical scoring and 'our' scoring

- Any given classical (alternative-based) scoring SCO induces a scoring s in our (proposition-based) sense.
- In fact, in two plausible (and as we'll see, equivalent!) ways, namely either by

$$s_J(xPy) = SCO_J(x) - SCO_J(y), \tag{3}$$

or by

$$s_J(xPy) = \max\{SCO_J(x) - SCO_J(y), 0\} \tag{4}$$

Proposition 2 *In the case of the preference agenda, every classical scoring rule is a scoring rule, namely one with respect to a scoring s derived from the classical scoring SCO via (3) or via (4).*

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.d: Reversal scoring and a Borda rule for judgment aggregation

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Reversal scoring

- Define so-called *reversal scoring* by:

$$s_J(p) = \text{nb. of judgment reversals needed to reject } p \quad (5)$$

$$= \min_{J' \in \mathcal{J} : p \notin J'} d_{\text{Kemeny}}(J, J'). \quad (6)$$

- E.g., $s_J(p) = 0$ if $p \notin J$.

Reversal scoring

- Let's try out reversal scoring for our doctrinal paradox agenda and profile:

	Score of...									
Individual	p	$\neg p$	q	$\neg q$	r	$\neg r$	pqr	$p\neg q\neg r$	$\neg pq\neg r$	$\neg p\neg q\neg r$
1 (pqr)	2	0	2	0	2	0	6	2	2	0
2 ($p\neg q\neg r$)	1	0	0	2	0	2	1	5	2	4
3 ($\neg pq\neg r$)	0	2	1	0	0	2	1	2	5	4
Group	3	2	3	2	2	4	8	9*	9*	8

- E.g., individual 1's judgment set pqr leads to a score of 2 for p, since rejecting p requires negating not just p (as $\neg pqr$ is inconsistent), but also r (where $\neg pq\neg r$ is consistent).
- Notice: a tie between the conclusion-based judgment sets $p\neg q\neg r$ and $\neg pq\neg r$!

Reversal scoring and classical Borda scoring

- The remarkable feature of reversal scoring is its link to classical Borda scoring for the preference agenda:

Remark 1 *In the case of the preference agenda (for any finite set of alternatives), reversal scoring* s *is given by*

$$s_J(xPy) = \mathsf{max}\{SCO_J(x) - SCO_J(y), 0\}$$

where SCO *is classical Borda scoring.*

See why?

Reversal scoring *rule* and classical Borda *rule*

Remark 1 and Proposition 2 imply:

Proposition 3 *The reversal scoring rule generalizes Borda rule, i.e., matches it in the case of the preference agenda (for any finite set of alternatives).*

Excursion: Zwicker's and Duddy-Piggin's ways to generalize Borda rule

Zwicker's approach

- Zwicker (2011) takes an interesting, very different strategy to extending Borda rule.
- The motivation derives from a geometric characterization of Borda preference aggregation obtained by Zwicker (1991).
- Write the agenda as $X = \{p_1, \neg p_1, p_2, \neg p_2, ..., p_m, \neg p_m\}$.
- Each profile gives rise to a vector $\mathbf{v} \equiv (v_1, ..., v_m)$ in $\mathbb{R}^m$ whose j^{th} entry v_j is the *net support for* p_j.

Excursion (cont.)

- Zwicker writes the vector $\mathbf{v}$ as an orthogonal sum $\mathbf{v}_{\text{consistent}} + \mathbf{v}_{\text{inconsistent}}$.
- Intuitively, '$\mathbf{v}_{\text{consistent}}$' contains the profile's 'consistent component'.
- Zwicker's Borda-type rule accepts all p_j for which $\mathbf{v}_{\text{consistent},j} > 0$.
- Problem: the decomposition $\mathbf{v}_{\text{consistent}} + \mathbf{v}_{\text{inconsistent}}$ so far 'works' only for special agendas.

Excursion (cont.)

In summary, there seem to exist two quite different approaches to generalizing Borda:

- Zwicker's approach is geometric and seeks to filter out the profile's 'inconsistent component'.
- My approach
  - retains the principle of score-maximization inherent in Borda aggregation (with scoring now defined at the level of propositions, not alternatives)
  - uses information about someone's *strength* of accepting a proposition (as measured by the score), just as classical Borda rule uses information about *strength* of preference (as measured by classical scores of alternatives).

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.e: A generalization of reversal scoring

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

A generalization of reversal scoring

- For *any* given distance function d over $\mathcal{J}$ (not necessarily Kemeny distance!), one might consider the scoring s defined by

$$s_J(p) = \text{distance by which one must} \tag{7}$$
$$\text{depart from } J \text{ to reject } p \tag{8}$$
$$= \min_{J' \in \mathcal{J} : p \notin J'} d(J, J').$$

- This yields a whole class of scoring rules, all of which are variants of our judgment-theoretic Borda rule. In the special case of the preference agenda, we thus obtain new variants of classical Borda rule.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.f: Scoring based on logical entrenchment

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Score as ‘logical entrenchment’

- We now consider scoring rules which explicitly exploit the logical structure of the agenda.
- Think of the score of a proposition p ($\in X$) given the judgment set J ($\in \mathcal{J}$) as the degree to which p is logically entrenched in the belief system J, i.e., as the ‘strength’ with which J entails p.
- We measure this strength by the number of ways in which p is entailed by J, where each ‘way’ is given by a particular judgment *subset* $S \subseteq J$ which entails p, i.e., for which $S \cup \{\neg p\}$ is inconsistent.
- There are different ways to formalise this idea!

First (naive) attempt

- Let's count *each* judgment subset which entails p as a separate, full-fledged 'way' in which p is entailed.
- This leads to so-called *entailment scoring*, defined by:

$$\begin{aligned} s_J(p) &= \text{number of judgment subsets entailing } p \\ &= \left|\{S \subseteq J : S \text{ entails } p\}\right| . \end{aligned} \tag{9}$$

- Objection: lots of redundancies, i.e., 'multiple counting'.

Second attempt

- To respond to the redundancy objection, let's count two entailments of p as different only if they have no premise in common.
- Formally, define *disjoint-entailment scoring* by:

$$\begin{aligned} s_J(p) &= \text{nb. of } \textit{disjoint} \text{ judgment subsets entailing } p \\ &= \max\{m : J \text{ has } m \text{ disjoint subsets each entailing } p\}. \end{aligned} \tag{10}$$

Example

- For our doctrinal paradox profile, we get the following disjoint-entailment scores

Individual	Score of...									
	p	$\neg p$	q	$\neg q$	r	$\neg r$	pqr	$p\neg q\neg r$	$\neg pq\neg r$	$\neg p\neg q\neg r$
1 (pqr)	2	0	2	0	2	0	6	2	2	0
2 ($p\neg q\neg r$)	1	0	0	2	0	2	1	5	2	4
3 ($\neg pq\neg r$)	0	2	1	0	0	2	1	2	5	4
Group	3	2	3	2	2	4	8	9*	9*	8

- E.g., for individual 2 proposition $\neg r$ scores 2 because $\neg r$ is entailed by $\{\neg r\}$ and by $\{p, \neg q\}$.

Borda again

- Applied to the preference agenda, disjoint entailment scoring matches reversal scoring.
  (but the two come apart for other agendas)
- So we've another, *different*, Borda extension!

Another option

- Counting only *minimal* entailments:

$$
\begin{aligned}
s_J(p) &= \text{nb. of judgment subsets } \textit{minimally} \text{ entailing } p \\
&= |\{S \subseteq J : S \text{ minimally entails } p\}| \, .
\end{aligned}
$$

Yet another option

- Counting only *irreducible* entailments:

$$s_J(p) = \text{nb. of judgment subsets } \textit{irreducibly} \text{ entailing } p$$
$$= |\{S \subseteq J : S \text{ irreducibly entails } p\}| \,.$$

- This again generalizes Borda!

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

2.g: More that can be done

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Premise-based rule as as as scoring rule

- Use a scoring which assigns far higher scores to accepted premises than to accepted conclusions!

Conclusion-based rule as as as scoring rule

- Use a scoring which assigns far higher scores to accepted conclusions than to accepted premises!

Scoring rules to ‘repair’ quota rules

- A quota rule: accepts each proposition $p \in X$ iff at least some number m_p of individuals accept p.
- Such a rule can generate irrational (e.g., inconsistent, or incomplete, or not deductively closed) outputs!
- A suitable scoring rule can ‘repair’ the quota rule:
  – this scoring rule matches the quota rule whenever the quota rule has a rational output, while rendering the output rational otherwise.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time allows)

Part 4: Concluding remarks

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time)

3.a: Definition

Part 4: Concluding remarks

Set scoring

- A *set scoring function* – or simply *set scoring* – is a function $\sigma : \mathcal{J} \times \mathcal{J} \longrightarrow \mathbb{R}$ which to every pair of rational judgment sets C and J assigns a real number $\sigma_J(C)$, the *score* of C given J.
- Elementary example ('*naive*' set scoring):

$$\sigma_J(C) = \begin{cases} 1 & \text{if } C = J \\ 0 & \text{if } C \neq J. \end{cases} \tag{11}$$

Set scoring rules

- Given a set scoring σ, the **set scoring rule** (or **generalized scoring rule**) *w.r.t.* σ is the aggregation rule $F_\sigma : \mathcal{J}^n \rightrightarrows \mathcal{J}$ given by:

$$F_\sigma(J_1, ..., J_n) = \text{argmax}_{C \in \mathcal{J}} \sum_{i \in N} \sigma_{J_i}(C).$$

- An aggregation rule is a **set scoring rule** simpliciter if it is the set scoring rule w.r.t. to some set scoring σ.

Set scoring rules generalize scoring rules

- To any ordinary scoring s corresponds a set scoring σ, given by

  $$\sigma_J(C) \equiv \sum_{p \in C} s_J(p),$$

  and the ordinary scoring rule w.r.t. s coincides with the set scoring rule w.r.t. σ.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time)

3.b: Some standard rules as set scoring rules

Part 4: Concluding remarks

Plurality rule as a set scoring rule

- **Plurality rule** is the aggregation rule F which for every profile $(J_1, ..., J_n) \in \mathcal{J}^n$ returns:

$$\begin{aligned} F(J_1, ..., J_n) &= \text{most frequently submitted judgment set(s)} \\ &= \mathsf{argmax}_{C \in \mathcal{J}} \left| \{ i : J_i = C \} \right| . \end{aligned}$$

- Normatively questionable!
- As one easily shows:

Remark 2 *The naive set scoring rule is plurality rule.*

Distance-based rule as a set scoring rule

- Given an arbitrary distance function d over $\mathcal{J}$, consider **distance-based** set scoring, defined by

$$\sigma_J(C) = -d(C, J). \tag{12}$$

- This renders sum-score-maximization equivalent to sum-distance-minimization:

Remark 3 *For every given distance function over $\mathcal{J}$, the distance-based set scoring rule is the distance-based rule.*

–> Conversely, not all set scoring rules are distance-based rules.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time)

3.d: ‘Epistemic’ rules as set scoring rules

Part 4: Concluding remarks

Further set scoring rules

- Let's take the *epistemic* or *truth-tracking* approach to JA.
- In a full probabilistic model of votes and the 'unknown truth', one may define:
  - the *maximum-likelihood rule*, which returns collective judgments whose truth would make the profile (the 'data') maximally likely;
  - the *maximum-posterior rule*, which returns the collective judgments whose posterior probability of truth given the profile is maximal.
- See, e.g., work by Pivato (2011) (and Dietrich 2006).
- Under particular conditions, these 'epistemic' rules can be modelled as particular scoring rules.

Plan

Part 1: The judgment aggregation framework

Part 2: Scoring rules

Part 3: Set scoring rules (if time)

Part 4: Concluding remarks

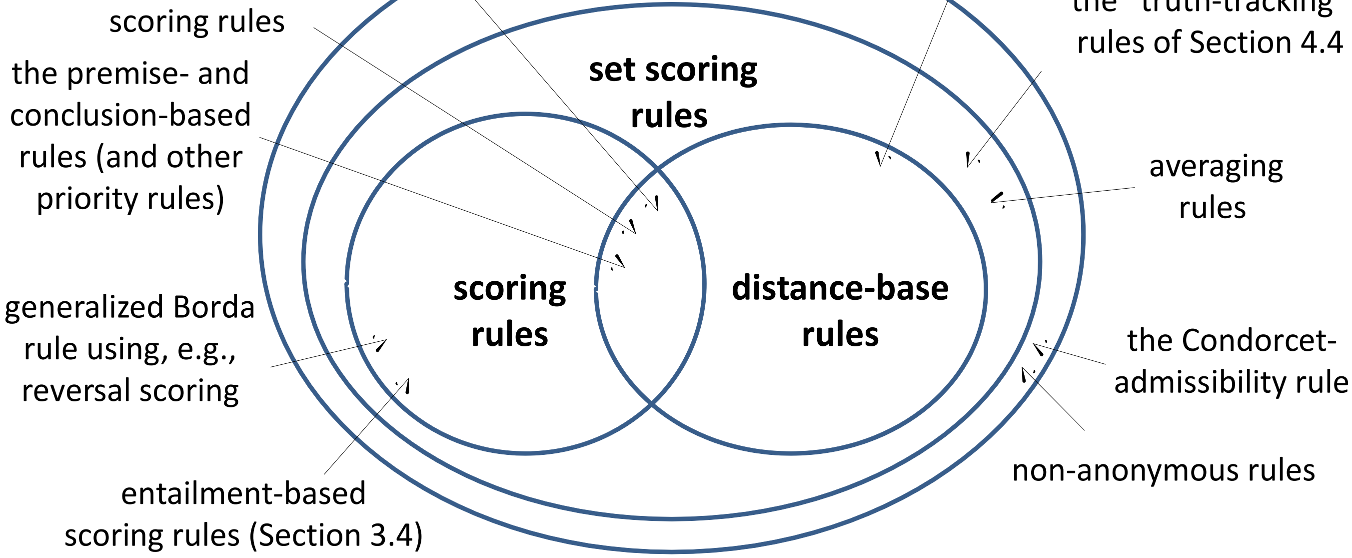

Figure 1: A map of judgment aggregation possibilities

Where do we stand?

Two possible extensions

Two plausible generalizations of (set) scoring rules:

- Allow scoring to depend on the individual i!
  - This leads to non-anonymous rules.
- Maximize total score within a larger set than the set $\mathcal{J}$ of fully rational judgment sets (such as the set of consistent but possibly incomplete judgment sets)!
  - This leads to 'boundedly rational scoring rules'.